Signal Detection in High Dimension: The Multispiked Case

Alexei Onatski, Marcelo J. Moreira and Marc Hallin*

October 23, 2012

Abstract

This paper deals with the local asymptotic structure, in the sense of Le Cam's asymptotic theory of statistical experiments, of the signal detection problem in high dimension. More precisely, 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 asymptotic power envelope and the asymptotic powers of previously proposed tests are derived. These asymptotic powers are shown to lie very substantially below the envelope, at least for relatively small values of the number of symmetry-breaking directions under the alternative. In contrast, the asymptotic power of the likelihood ratio test based on the eigenvalues of the sample covariance matrix is shown to be close to that envelope. These results extend to the case of multispiked alternatives the findings of an earlier study (Onatski, Moreira and Hallin, 2011) of the single-spiked case. The

*Alexei Onatski: University of Cambridge, ao319@cam.ac.uk. Marcelo J. Moreira: Fundação Getulio Vargas, mjmoreira@fgv.br. Marc Hallin: Université libre de Bruxelles and Princeton University, mhallin@ulb.ac.be.

methods we are using here, however, are entirely new, as the Laplace approximations considered in the single-spiked context do not extend to the multispiked case.

Key words: sphericity tests, large dimensionality, asymptotic power, spiked covariance, contiguity, power envelope.

1 Introduction

In a recent paper, Onatski, Moreira and Hallin (2011) (hereafter OMH) analyze the asymptotic power of statistical tests in the detection of a signal in spherical real-valued Gaussian data as the dimensionality of the data and the number of observations diverge to infinity at the same rate. This paper generalizes OMH's alternative of a single symmetry-breaking direction (*single-spiked* alternative) to the alternative of multiple symmetry-breaking directions (*multispiked* alternative), which is more relevant for applied work.

Contemporary tests of sphericity in a high-dimensional environment (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 portion 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 multi-spiked alternatives considered here, however, is all but 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 Maida (2005).

Let us describe the setting and main results in more detail. Suppose that the data consist of n independent observations X_t, $t = 1, ..., n$ of a p-dimensional Gaussian vector with mean zero and positive definite covariance matrix Σ. Let $\Sigma = \sigma^2 (I_p + VHV')$, where I_p is the p-dimensional identity matrix, σ is a scalar, 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 $H_0 : h_1 = ... = h_r = 0$ against the alternative $H_1 : h_j > 0$ for some $j = 1, ..., r$, based on the eigenvalues of the sample covariance matrix of the data when $n, p \to \infty$ so that $p/n \to c$ with $0 < c < \infty$, an asymptotic regime which we abbreviate into $n, p \to_c \infty$. 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 specified and unspecified σ^2. For the sake of simplicity, in the rest of this introduction, we only discuss the case of specified $\sigma^2 = 1$, although the case of unspecified σ^2 is more realistic. Denoting by λ_j the j-th largest sample covariance eigenvalue, let $\lambda = (\lambda_1, ..., \lambda_m)$, where $m = \min(n, p)$. We begin our analysis with a study of the asymptotic properties of the likelihood ratio process $\left\{L\left(h;\lambda\right);h\in\left[0,\bar{h}\right]^{r}\right\}$, where $h=\left(h_1,...,h_r\right)$, $\bar{h}\in\left[0,\sqrt{c}\right)$ and $L\left(h;\lambda\right)$ is defined as the ratio of the density of λ under H_1 to that under H_0, considered as a λ-measurable random variable. Note that $L\left(h;\lambda\right)$ depends on n and p, while λ is $m=\min\left\{n,p\right\}$ -dimensional. An exact formula for $L\left(h;\lambda\right)$ involves the integral $\int_{\mathcal{O}(p)}e^{\mathrm{tr}(AQBQ')}\left(dQ\right)$ over the orthogonal group $\mathcal{O}\left(p\right)$, 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 (see Wang (2010) and Mo (2011) for independent different derivations of this contour integral representation for this particular $r=1$ case). Then, the Laplace approximation to the contour integral is used to derive the asymptotic behavior of $L\left(h;\lambda\right)$.

Onatski (2012) generalizes the contour integral representation to the multi-spiked case ($r>1$). For complex-valued data, such a generalization allows him to extend OMH's results to the multi-spiked context. Unfortunately, for real-valued data, which we are concerned with in this paper, this generalization is not straightforwardly amenable to the Laplace approximation method. Therefore, in this paper, we consider a totally different approach. For the $r=1$ case, Guionnet and Maida (2005) (hereafter GM) use large deviation methods to derive a second-order asymptotic expansion of $\int_{\mathcal{O}(p)}e^{\mathrm{tr}(AQBQ')}\left(dQ\right)$ 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 $L\left(h;\lambda\right)$.

More precisely, we show that, for any $\bar{h}$ such that $0<\bar{h}<\sqrt{c}$, the sequence of log-likelihood processes $\{\ln L\left(h;\lambda\right);h\in[0,\bar{h}]^{r}\}$ converges weakly to a Gaussian process $\left\{\mathcal{L}_{\lambda}(h);h\in\left[0,\bar{h}\right]^{r}\right\}$ under the null hypothesis H_0 as $n,p\to_c\infty$. The index λ in the notation $\mathcal{L}_{\lambda}(h)$ is used to distinguish the limiting λ-log-likelihood process in the case of specified $\sigma^2=1$, from that of the μ-log-likelihood process considered in the case of unspecified σ^2, which we denote by $\mathcal{L}_{\mu}(h)$ (see Section 2). The limiting process has mean $\mathrm{E}\left[\mathcal{L}_\lambda(h)\right] = \frac{1}{4}\sum\limits_{i,j=1}^{r} \ln\left(1 - h_i h_j / c\right)$ and autocovariance function $\mathrm{Cov}\left(\mathcal{L}_\lambda\left(h\right), \mathcal{L}_\lambda(\tilde{h})\right) = -\frac{1}{2}\sum_{i,j=1}^{r} \ln\left(1 - h_i \tilde{h}_j / c\right)$. That convergence entails the weak convergence, in the Le Cam sense, of the h-indexed statistical experiments $\mathcal{E}_\lambda^m$ under which the eigenvalues $\lambda_1, ..., \lambda_m$ are observed, i.e. the statistical experiments with log-likelihood process $\left\{\ln L_\lambda\left(h\right); h \in \left[0, \bar{h}\right]^r\right\}$ (see van der Vaart (1998), page 126). Although this limiting process is Gaussian, it is not a log-likelihood process of the Gaussian shift type, so that the statistical experiments $\mathcal{E}_\lambda^m$ under study are not locally asymptotically normal (LAN) ones. The weak convergence of $\mathcal{E}_\lambda^m$ implies, however, via Le Cam's first lemma (see van der Vaart 1998, p.88), that the joint distributions of the normalized sample covariance eigenvalues under the null and under alternatives associated with $h \in [0, \sqrt{c})$ are mutually contiguous.

An asymptotic power envelope for λ-based tests of H_0 against H_1 can be constructed using the Neyman-Pearson lemma and Le Cam's third lemma. We show that, for tests of size α, the maximum achievable asymptotic power against a point alternative $h = (h_1, ..., h_r)$ equals $1 - \Phi\left[\Phi^{-1}\left(1 - \alpha\right) - \sqrt{W(h)}\right]$, where Φ is the standard normal distribution function and $W(h) = -\frac{1}{2}\sum_{i,j=1}^{r} \ln\left(1 - h_i h_j / c\right)$. As we explain in the paper, this asymptotic power envelope is valid not only for the λ-based tests, but also for all tests that are invariant under left orthogonal transformations of the data X_t, $t = 1, ..., n$.

Next, we consider previously proposed tests of sphericity and of the equality of the population covariance matrix to a given matrix . We focus on the tests studied in Ledoit and Wolf (2002), Bai et al (2009), and Cai and Ma (2012). We find that, in general, the asymptotic powers of those tests are substantially lower than the corresponding asymptotic power envelope value. In contrast, our computations for the case $r = 2$ show that the asymptotic powers of the λ- and μ-based likelihood ratio tests are close to the power envelope.

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 provides an analysis of 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

Let X be a $p\times n_p$ matrix with independent Gaussian $N\left(0,\sigma^2\left(I_p+VHV'\right)\right)$ columns. Let $\lambda_{p1}\geq ...\geq\lambda_{pp}$ be the ordered eigenvalues of $\frac{1}{n_p}XX'$ and write $\lambda_p=(\lambda_{p1},...,\lambda_{pm})$, where $m=\min\{p,n_p\}$. Similarly, let $\mu_{pi}=\lambda_{pi}/(\lambda_{p1}+...+\lambda_{pp})$, $i=1,...,m$ and $\mu_p=\left(\mu_{p1},...,\mu_{p,m-1}\right)$.

As explained in the introduction, our goal is to study the asymptotic power, as $n_p,p\rightarrow_c\infty$, of the eigenvalue-based tests of $H_0:h_1=...=h_r=0$ against $H_1:h_j>0$ for some $i=1,...,r$, where h_j are the diagonal elements of the diagonal matrix H. If σ^2 is specified, the model is invariant with respect to left and right orthogonal transformations; sufficiency and invariance arguments (see Appendix 5.4 for details) lead to considering tests based on λ_p only. If σ^2 is unspecified, the model is invariant with respect to left and right orthogonal transformations and multiplications by non-zero scalars; sufficiency and invariance arguments (see Appendix 5.4) lead to considering tests based on μ_p only. Note that 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, we henceforth assume without loss of generality 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 have

$$f_{\lambda p}\left(\tilde{x};h\right)=\tilde{\gamma}\frac{\prod_{i=1}^{m}x_{i}^{\frac{|p-n_{p}|-1}{2}}\prod_{i<j}^{m}\left(x_{i}-x_{j}\right)}{\prod_{j=1}^{r}\left(1+h_{j}\right)^{n_{p}/2}}\int\limits_{\mathcal{O}(p)}e^{-\frac{n_{p}}{2}\operatorname{tr}(\Pi Q'\mathcal{X}Q)}\left(\mathrm{d}Q\right), \tag{1}$$

where $\tilde{\gamma}$ depends only on n_p and p; $\Pi=\operatorname{diag}\left(\left(1+h_1\right)^{-1},...,\left(1+h_r\right)^{-1},1,...,1\right)$; $\mathcal{X}=\operatorname{diag}\left(x_1,...,x_m,0,...,0\right)$ is a $(p\times p)$ diagonal matrix; $\mathcal{O}\left(p\right)$ is the set of all $p\times p$ orthogonal matrices; and $(\mathrm{d}Q)$ is the invariant measure on the orthogonal group $\mathcal{O}\left(p\right)$, normalized to make the total measure unity. Formula (1) 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 let $y_i=x_i/x$. 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 (1) and integrating x out, we obtain

$$f_{\mu p}\left(\tilde{y};h\right)=\tilde{\gamma}\frac{\prod_{i=1}^{m}y_{i}^{\frac{|p-n_{p}|-1}{2}}\prod_{i<j}^{m}\left(y_{i}-y_{j}\right)}{\prod_{j=1}^{r}\left(1+h_{j}\right)^{n_{p}/2}}\int_{0}^{\infty}x^{\frac{n_{p}p}{2}-1}\int\limits_{\mathcal{O}(p)}e^{-\frac{n_{p}}{2}x\operatorname{tr}(\Pi Q'\mathcal{Y}Q)}\left(\mathrm{d}Q\right)\mathrm{d}x, \tag{2}$$

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

Consider the likelihood ratios $L_p\left(h;\lambda_p\right)=f_{\lambda p}\left(\lambda_p;h\right)/f_{\lambda p}\left(\lambda_p;0\right)$ and $L_p\left(h;\mu\right)=f_{\mu p}\left(\mu_p;h\right)/f_{\mu p}\left(\mu_p;0\right)$. Formulae (1) and (2) imply the following proposition.

Proposition 1 *Let $\mathcal{O}\left(p\right)$ be the set of all $p\times p$ orthogonal matrices. Denote by $(\mathrm{d}Q)$ the invariant measure on the orthogonal group $\mathcal{O}\left(p\right)$ normalized to make the total measure unity. Put $\Lambda_p=\operatorname{diag}\left(\lambda_{p1},...,\lambda_{pp}\right)$, $S_p=\lambda_{p1}+...+\lambda_{pp}$, and let D_p be the $p\times p$ diagonal matrix $\operatorname{diag}\left(\frac{1}{2c_p}\frac{h_1}{1+h_1},...,\frac{1}{2c_p}\frac{h_r}{1+h_r},0,...,0\right)$, where $c_p=p/n_p$.*

Then,

$$L_p(h;\lambda_p) = \prod_{j=1}^{r} (1+h_j)^{-\frac{n_p}{2}} \int\limits_{\mathcal{O}(p)} e^{p\,\mathrm{tr}(D_p Q' \Lambda_p Q)} \, (\mathrm{d}Q) \;\; \textit{and} \tag{3}$$

$$L_p\big(h;\mu_p\big) = \prod_{j=1}^{r} (1+h_j)^{-\frac{n_p}{2}} \frac{\left(\frac{n_p}{2}\right)^{\frac{n_p p}{2}}}{\Gamma\left(\frac{n_p p}{2}\right)} \int_0^\infty x^{\frac{n_p p}{2}-1} e^{-\frac{n_p}{2}x} \int\limits_{\mathcal{O}(p)} e^{p\frac{x}{S_p}\,\mathrm{tr}(D_p Q' \Lambda_p Q)} \, (\mathrm{d}Q) \mathrm{d}x. \tag{4}$$

In the special case where $r = 1$, the rank of the matrix D_p equals one, and the integrals over the orthogonal group in (3) and (4) can be rewritten as integrals over a p-dimensional sphere. OMH show how such spherical integrals can be represented in the form of contour integrals, and apply Laplace approximation to these contour integrals to establish the asymptotic properties of $L_p(h;\lambda_p)$ and $L_p\big(h;\mu_p\big)$. In the $r > 1$ case, the integrals in (3) and (4) 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, as p goes to infinity, of integrals of the form $\int_{\mathcal{O}(p)} e^{p\,\mathrm{tr}(DQ'\Lambda Q)}\,(\mathrm{d}Q)$ was analyzed in Guionnet and Maida (2005) (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 Guionnet and Maida'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 $L_p(h;\lambda_p)$ and $L_p\big(h;\mu_p\big)$.

Let $\hat{F}_p^{\lambda}$ be the empirical distribution of $\lambda_{p1}, ..., \lambda_{pp}$, and denote by F_p^{MP} the

Marchenko-Pastur distribution function, with density

$$f_p^{MP}(x) = \frac{1}{2\pi c_p x}\sqrt{(b_p - x)(x - a_p)}, \tag{5}$$

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. As is well known, the difference between $\hat{F}_p^{\lambda}$ and F_p^{MP} weakly converges to zero a.s. as $p, n_p \to_c \infty$. Moreover, $\lambda_{p1} \overset{a.s}{\to} \left(1 + \sqrt{c}\right)^2$, and $\lambda_{pp} \overset{a.s}{\to} \left(1 - \sqrt{c}\right)^2$ if $c > 1$, and $\lambda_{pp} \overset{a.s}{\to} 0$ if $c \leq 1$.

Consider the Hilbert transform of F_p^{MP}, $H_p^{MP}(x) = \int (x - \lambda)^{-1} \mathrm{d}F_p^{MP}(\lambda)$. 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 (5), we get

$$H_p^{MP}(x) = \frac{x + c_p - 1 - \sqrt{(x - c_p - 1)^2 - 4c_p}}{2c_p x}, \tag{6}$$

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

$$K_p^{MP}(x) = \frac{1}{x} + \frac{1}{1 - c_p x},\ x \in H_p^{MP}\left(\mathbb{R}\backslash \operatorname{supp}\left(F_p^{MP}\right)\right). \tag{7}$$

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

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

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

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

From (6), $H_p^{MP}\left(\mathbb{R}\backslash\operatorname{supp}\left(F_p^{MP}\right)\right)=(-\infty,0)\cup\left(0,\frac{1}{\sqrt{c_p}\left(1+\sqrt{c_p}\right)}\right)\cup\left(\frac{1}{\sqrt{c_p}\left(\sqrt{c_p}-1\right)},\infty\right)$ when $c_p>1$, $\left(-\frac{1}{\sqrt{c_p}\left(1-\sqrt{c_p}\right)},0\right)\cup\left(0,\frac{1}{\sqrt{c_p}\left(1+\sqrt{c_p}\right)}\right)$ when $c_p<1$, and $(-\infty,0)\cup(0,1/2)$ when $c_p=1$. Therefore, $\Omega_{\varepsilon\eta}\subset H_p^{MP}\left(\mathbb{R}\backslash\operatorname{supp}\left(F_p^{MP}\right)\right)$ with probability approaching one as $n_p,p\rightarrow_c\infty$.

Proposition 2 *Let* $\{\Theta_p\}$ *be a sequence of random* $p\times p$ *diagonal matrices* $\operatorname{diag}(\theta_{p1},...,\theta_{pr},0,...,0)$, *where* $\theta_{pj}\neq 0$, $j=1,...,r$. *Further, let* $v_{pj}=R_p^{MP}(2\theta_{pj})$, *where* $R_p^{MP}(x)=1/(1-c_px)$ *is the R-transform of the Marchenko-Pastur distribution* F_p^{MP}. *Assume that, for some* $\varepsilon>0$ *and* $\eta>0$, $2\theta_{pj}\in\Omega_{\varepsilon,\eta}$ *with probability approaching one as* $n_p,p\rightarrow_c\infty$. *Then,*

$$\int_{\mathcal{O}(p)} e^{p\operatorname{tr}(\Theta_pQ'\Lambda_pQ)}(\mathrm{d}Q) = e^{p\sum_{j=1}^{r}\left[\theta_{pj}v_{pj}-\frac{1}{2p}\sum_{i=1}^{p}\ln(1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_{p,i})\right]}$$
$$\times\prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1-4\left(\theta_{pj}v_{pj}\right)\left(\theta_{ps}v_{ps}\right)c_p}\,(1+o(1))\ \ a.s.,$$

where $o(1)$ *is uniform over all sequences* $\{\Theta_p\}$ *satisfying the assumption.*

This proposition extends Theorem 3 of Guionnet and Maida (2005) to cases when $\operatorname{rank}(\Theta_p)>1$, θ_{pj} depends on p, and Λ_p is random. When $r=1$, $\theta_{p1}=\theta>0$ and $v_{p1}=v$ are fixed, it is straightforward to verify that $\sqrt{1-4\theta^2v^2c_p}=\sqrt{4\theta^2}/\sqrt{Z}$, where $Z=\int\left(K_p^{MP}(2\theta)-\lambda\right)^{-2}\mathrm{d}F_p^{MP}(\lambda)$. In Guionnet and Maida's (2005) Theorem 3, the expression $\sqrt{4\theta^2}/\sqrt{Z}$ should have been used instead of $\sqrt{Z-4\theta^2}/\theta\sqrt{Z}$, which is a typo.

Setting $r=1$ and $\theta_{p1}=\frac{1}{2c_p}\frac{h}{1+h}$ in Proposition 2 and using formula (3) from Proposition 1 gives us an expression for $L_p(h;\lambda_p)$ which is an equivalent of 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 $\theta_{pj} = h_j / 2c_p (1 + h_j)$ and

$$H_\delta = \begin{cases} [-1+\delta, 0) \cup (0, \sqrt{c} - \delta] & \text{for } c > 1, \\ [-\sqrt{c}+\delta, 0) \cup (0, \sqrt{c} - \delta] & \text{for } c \leq 1. \end{cases} \tag{8}$$

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

With the above setting for θ_{pj}, we have $v_{pj} = 1 + h_j$ and $K_p^{MP}(2\theta_{pj}) = (c_p + h_j)(1 + h_j) / h_j = z_{j0}$, say, as in Theorem 7 in OMH. Define

$$\Delta_p (z_{j0}) = \sum_{i=1}^{p} \ln (z_{j0} - \lambda_{pi}) - p \int \ln (z_{j0} - \lambda) \, \mathrm{d}F_p^{MP} (\lambda) . \tag{9}$$

Theorem 3 *Suppose that the null hypothesis is true ($h = 0$). Let δ be any fixed number such that $0 < \delta < \sqrt{c}$, and let $C [0, \sqrt{c} - \delta]^r$ be the space of real-valued continuous functions on $[0, \sqrt{c} - \delta]^r$ equipped with the supremum norm. Then, as $p, n_p \to_c \infty$,*

$$L_p(h; \lambda_p) = \prod_{j=1}^{r} \exp\left\{ -\frac{1}{2} \Delta_p (z_{j0}) + \frac{1}{2} \sum_{s=1}^{j} \ln\left(1 - \frac{h_j h_s}{c_p} \right) \right\} (1 + o(1)) \ \text{ and } \tag{10}$$

$$L_p\left(h; \mu_p\right) = L_p(h; \lambda_p) \exp\left\{ \frac{1}{4c_p} \left(\sum_{j=1}^{r} h_j \right)^2 - \frac{S_p - p}{2c_p} \sum_{j=1}^{r} h_j \right\} (1 + o(1)), \tag{11}$$

almost surely, where the $o(1)$ terms are uniform in $h \in [0, \sqrt{c} - \delta]^r$. Furthermore, $\ln L_p(h; \lambda_p)$ and $\ln L_p\left(h; \mu_p\right)$, viewed as random elements of $C [0, \sqrt{c} - \delta]^r$, converge weakly to $\mathcal{L}_\lambda (h)$ and $\mathcal{L}_\mu (h)$ with Gaussian finite-dimensional distributions such that $\mathrm{E} (\mathcal{L}_\lambda (h)) = -\frac{1}{2} \mathrm{Var} (\mathcal{L}_\lambda (h))$, $\mathrm{E} (\mathcal{L}_\mu (h)) = -\frac{1}{2} \mathrm{Var} (\mathcal{L}_\mu (h))$, and, for any

$h, \tilde{h} \in [0, \sqrt{c} - \delta]^r$,

$$\mathrm{Cov}\left(\mathcal{L}_\lambda(h), \mathcal{L}_\lambda\left(\tilde{h}\right)\right) = -\frac{1}{2}\sum_{i,j=1}^{r} \ln\left(1 - \frac{h_i \tilde{h}_j}{c}\right), \textit{ and} \tag{12}$$

$$\mathrm{Cov}\left(\mathcal{L}_\mu(h), \mathcal{L}_\mu\left(\tilde{h}\right)\right) = -\frac{1}{2}\sum_{i,j=1}^{r} \left(\ln\left(1 - \frac{h_i \tilde{h}_j}{c}\right) + \frac{h_i \tilde{h}_j}{c}\right). \tag{13}$$

Theorem 3 and Le Cam's first lemma (van der Vaart (1998), p.88) imply that the joint distributions of $\lambda_1, ..., \lambda_m$ (as well as those of $\mu_1, ..., \mu_{m-1}$) under the null and under the alternative are mutually contiguous for any $h \in [0, \sqrt{c})^r$. By applying Le Cam's third lemma (van der Vaart (1998), p.90), we can study the "local" powers of tests detecting signals in noise. The requirement that h_j be positive under alternatives corresponds to situations where the signals contained in the data are independent from the noise. If dependence between the signals and the noise is allowed, one might consider two-sided alternatives of the form $H_1 : h_j \neq 0$ for some j. Values of h_j between -1 and 0 correspond to alternatives under which the noise variance is reduced along certain directions. In view of Proposition 2, it should not be difficult to generalize Theorem 3 to the case of fully ($h_j \neq 0$, all j's) or partially ($h_j \neq 0$, some j's) two-sided alternatives. This problem will not be discussed here, and is left for future research.

3 Asymptotic power analysis

Denote by $\beta_\lambda(h)$ and $\beta_\mu(h)$, respectively, the asymptotic powers of the asymptotically most powerful λ- and μ-based tests of size α of the null $h = 0$ against a point alternative $h = (h_1, ..., h_r) \neq 0$ with $h_j < \sqrt{c}$, $j = 1, ..., r$. As functions of h, β_λ and β_μ are called the *asymptotic power envelopes.*

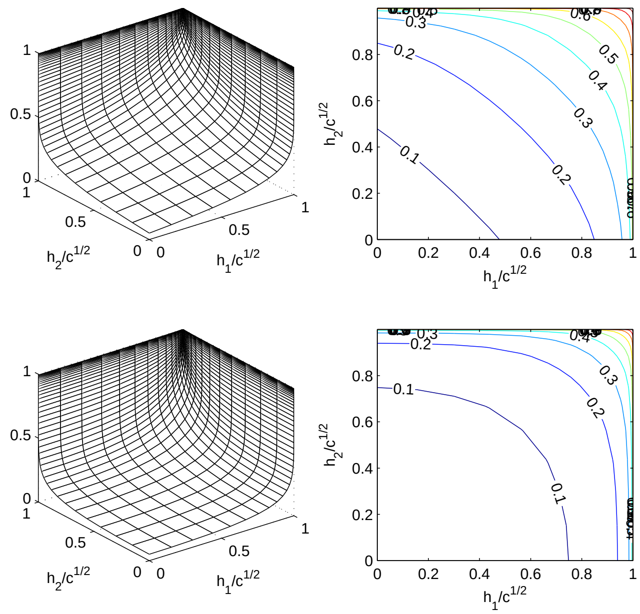

Figure 1: The power envelopes $\beta_\lambda(h)$ (upper panel) and $\beta_\mu(h)$ (lower panel) for $\alpha = 0.05$, as functions of $h/\sqrt{c} = (h_1, h_2)/\sqrt{c}$.

Proposition 4 *Let Φ denote the standard normal distribution function. Then,*

$$\beta_\lambda(h) = 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \sqrt{-\frac{1}{2}\sum_{i,j=1}^{r}\ln\left(1-\frac{h_i h_j}{c}\right)}\right] \quad \textit{and} \tag{14}$$

$$\beta_\mu(h) = 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \sqrt{-\frac{1}{2}\sum_{i,j=1}^{r}\left(\ln\left(1-\frac{h_i h_j}{c}\right)+\frac{h_i h_j}{c}\right)}\right]. \tag{15}$$

Figure 1 shows the asymptotic power envelopes $\beta_\lambda(h)$ and $\beta_\mu(h)$ as functions of $h_1/\sqrt{c}$ and $h_2/\sqrt{c}$ when $h = (h_1, h_2)$ is two-dimensional.

It is important to realize that the asymptotic power envelopes derived in Proposition 4 are valid 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 $\|A\|_F = \mathrm{tr}\,(A'A)$ 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 $h_1 = ... = h_r = 0$, and let H_1 be any of the following alternatives: $H_1 : h_j > 0$ for some $j = 1, ..., r$, or $H_1 : \Sigma \neq \sigma^2 I_p$, or $H_1 : \left\{\Sigma : \|\Sigma - \sigma^2 I_p\|_F > \varepsilon_{n,p}\right\}$, or $H_1 : \left\{\Sigma : \|\Sigma - \sigma^2 I_p\|_2 > \varepsilon_{n,p}\right\}$, where $\varepsilon_{n,p}$ is a positive constant that may depend on n and p.

Proposition 5 *For specified σ^2, consider tests of H_0 against H_1 that are invariant with respect to the left orthogonal transformations of the data $X = [X_1, ..., X_n]$. For any such test, there exists a test based on λ with the same 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 test based on μ with the same power function.*

Examples of the former tests include the tests of $H_0 : \Sigma = I$ studied in Chen et al (2010) and Cai and Ma (2012). An example of the latter test is the test of sphericity studied in Chen et al (2010). The tests studied in Chen et al (2010) and Cai and Ma (2012) are invariant, although they are not λ- or μ-based.

For $r = 1$, OMH show that the asymptotic power envelopes are closely approached by the asymptotic powers of the λ- and μ-based likelihood ratio tests. Our goal here is to explore the asymptotic power of those likelihood ratio tests for $r > 1$. Unfortunately, as r grows, it becomes increasingly difficult to compute the asymptotic critical values for the likelihood ratio tests by simulation. For example, $r = 2$ requires simulating a 2-dimensional Gaussian random field with the covariance function and the mean function described in Theorem 3.

For $r = 2$, Figure 2 shows sections of the power envelope (dotted lines) and the power of the likelihood ratio test based on λ for various fixed values of $h_1/\sqrt{c}$ under the alternative. Figure 3 shows the same plots for the tests based on μ. To enhance

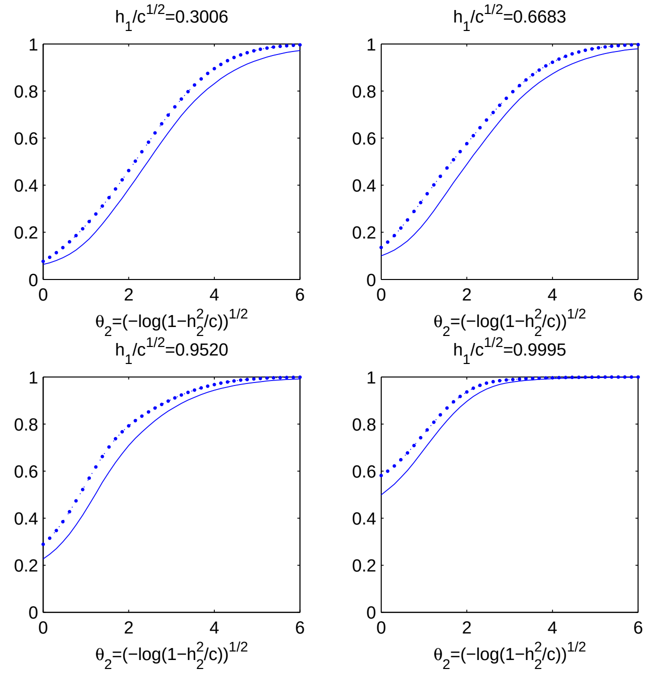

Figure 2: Profiles of the asymptotic power of the λ-based LR test (solid lines) relative to the asymptotic power envelope (dotted lines) for different values of $h_1/\sqrt{c}$ under the alternative; $\alpha = 0.05$.

readability, we use a different parametrization: $\theta_j = \sqrt{-\ln\left(1 - h_j^2/c\right)}$, $i = 1, ..., r$. As h_j varies in the region of contiguity $[0, \sqrt{c})$, θ_j spans the entire half-line $[0, \infty)$. Note that the asymptotic mean and autocovariance functions of the log likelihood ratios derived in Theorem 3 depend on h_j only through $h_j/\sqrt{c} = \sqrt{1 - e^{-\theta_j^2}}$. Therefore, under the new parametrization, they depend only on $\theta = (\theta_1, ..., \theta_r)$. The parameter θ plays the classical role of a "local parameter" in our setting.

Figure 4 further explores the relationship between the asymptotic powers of the λ- and μ-based LR test and the corresponding asymptotic power envelopes when $r = 2$. We pick all values of $h = (h_1, h_2)$ satisfying inequality $h_1 \geq h_2$ and such that the asymptotic power envelope for λ-based tests is exactly 25, 50, 75, and 90%. Then, we compute and plot the corresponding power of the λ-based LR test (solid lines) against h_2/h_1. The dashed lines show similar graphs for the

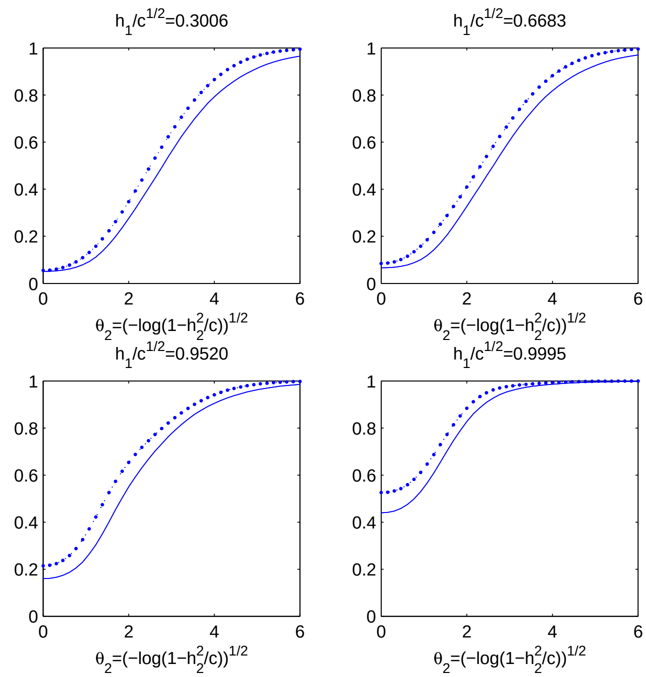

Figure 3: Profiles of the asymptotic power of the μ-based LR test (solid lines) relative to the asymptotic power envelope (dotted lines) for different values of $h_1/\sqrt{c}$ under the alternative; $\alpha = 0.05$.

μ-based LR test. The value $h_2/h_1 = 0$ corresponds to single-spiked alternatives $h_1 > 0$, $h_2 = 0$, the value $h_2/h_1 = 1$ corresponds to equi-spiked alternatives $h_1 = h_2 > 0$. The intermediate values of h_2/h_1 link the two extreme cases. We do not consider values $h_2/h_1 > 1$, as the power function is symmetric about the 45-degree line in the (h_1, h_2) space.

Somewhat surprisingly, the power of the LR test along the set of alternatives (h_1, h_2) corresponding to the same values of the asymptotic power envelope is not a monotone function of h_2/h_1. Equi-spiked alternatives typically seem to be particularly difficult to detect by the LR tests. However, for the set of alternatives corresponding to an asymptotic power envelope value of 90%, the single-spiked alternatives are even harder to detect.

A natural question is: how does the asymptotic power of the λ- and μ-based LR tests depend on the choice of r, that is, how do those tests perform when

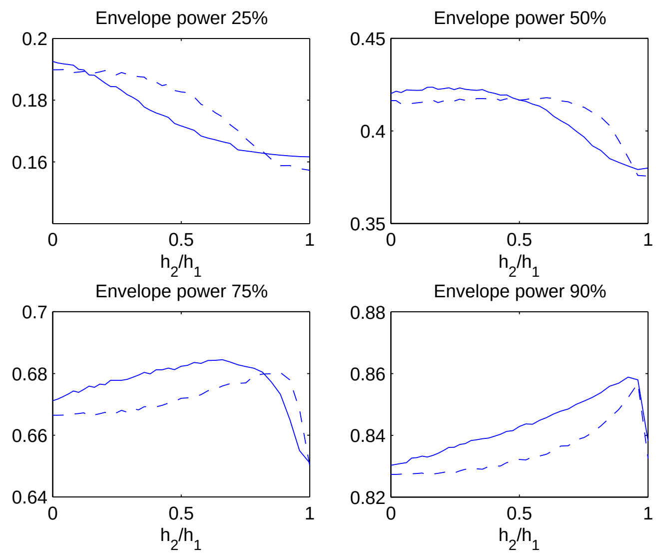

Figure 4: Power of λ-based (solid lines) and μ-based (dashed lines) LR tests plotted against h_2/h_1, where (h_1, h_2) are such that the respective asymptotic power envelopes $\beta_\lambda(h)$ and $\beta_\mu(h)$ equal 25, 50, 75 and 90%.

the actual r does not coincide with the value the test statistic is based on? For example, to detect a single signal, one can, in principle, use LR tests of the null hypothesis against alternatives with $r = 1, r = 2$, etc. How does the asymptotic powers of such tests compare? Figure 5 reports the asymptotic powers of the λ- and μ-based LR tests designed to detect alternatives with $r = 1$ (solid line) and $r = 2$ (dashed line), under single-spiked alternatives. As in Figures 2 and 3, we use the parametrization $\theta = \sqrt{-\ln(1 - h^2/c)}$ for the single-spiked alternative. It appears that the two asymptotic powers are very close to each other; interestingly, neither of them dominates the other. Using LR tests designed against alternatives with $r > 1$ seems to be beneficial for detecting a single-spiked alternative with relatively small θ (and h).

In the remaining part of this section, we consider examples of some of the tests that have been proposed previously in the literature, and, in Proposition 6, derive their asymptotic power functions.

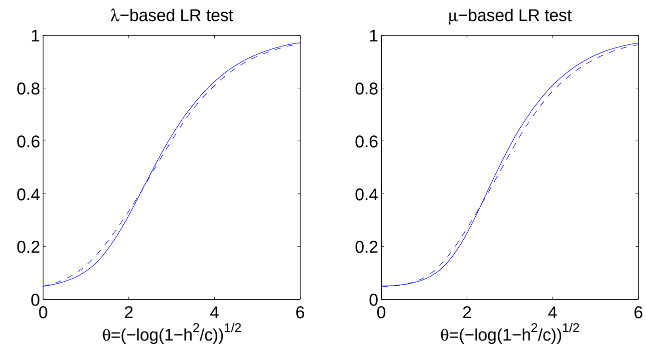

Figure 5: Asymptotic power of the λ-based (left panel) and μ-based (right panel) LR tests. Solid line: power when $r = 1$ is correctly assumed. Dashed line: power when $r = 2$ is incorrectly assumed.

Example 1 (John's (1971) test of sphericity $H_0 : \Sigma = \sigma^2 I$.) *John (1971) proposes testing the sphericity hypothesis $\theta = 0$ against general alternatives based on the test statistic*

$$U = \frac{1}{p}\operatorname{tr}\left[\left(\frac{\hat{\Sigma}}{(1/p)\operatorname{tr}\left(\hat{\Sigma}\right)} - I_p\right)^2\right], \tag{16}$$

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 $n, p \to_c \infty$. They prove that, under the null, $nU - p \xrightarrow{d} N(1, 4)$. Hence, the test with asymptotic size α rejects the null of sphericity whenever $\frac{1}{2}(nU - p - 1) > \Phi^{-1}(1 - \alpha)$.

Example 2 (The Ledoit-Wolf (2002) test of $H_0 : \Sigma = I$.) *Ledoit and Wolf (2002) propose*

$$W = \frac{1}{p}\operatorname{tr}\left[\left(\hat{\Sigma} - I\right)^2\right] - \frac{p}{n}\left[\frac{1}{p} tr\hat{\Sigma}\right]^2 + \frac{p}{n} \tag{17}$$

as a test statistic for testing the hypothesis that the population covariance matrix is the unit matrix. They show that, under the null, $nW - p \xrightarrow{d} N(1, 4)$. As in the previous example, the null 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 of $H_0 : \Sigma = I$.) *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 the equality of the population covariance matrix to the identity matrix against general alternatives. Under the null, $CLR \xrightarrow{d} N\left(-\frac{1}{2}\ln(1-c), -2\ln(1-c) - 2c\right)$ (still, as $n, p \to_c \infty$). 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 (Tracy-Widom-type tests of $H_0 : \Sigma = I$.) *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 is determined by the functional form of $\varphi(\cdot)$ and the fact that*

$$(\sigma_{n,c}(\lambda_1 - \nu_c), ..., \sigma_{n,c}(\lambda_r - \nu_c)) \xrightarrow{d} TW(r), \tag{18}$$

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 (18).

Example 5 (The Cai-Ma (2012) minimax test of $H_0 : \Sigma = I$.) *Cai and Ma (2012) propose to use a 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. Under the null, as $n, p \to_c \infty$, $T_n \xrightarrow{d} N(0, 4c^2)$. The null hypothesis is rejected at asymptotic level α whenever T_n is larger than $2\sqrt{p(p+1)/n(n-1)}\Phi^{-1}(1-\alpha)$. Cai and Ma (2012) show that this test is rate-optimal against general alternatives from a minimax point of view.

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

Proposition 6 *The asymptotic power functions of the tests described in Examples 1-5 are*

$$\begin{aligned}
\beta_{TW}(h) &= \alpha, \qquad (19)\\
\beta_J(h) &= \beta_{LW}(h) = \beta_{CM}(h) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{1}{2}\sum_{j=1}^{r}\frac{h_j^2}{c}\right), \text{ and} \qquad (20)\\
\beta_{CLR}(h) &= 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \sum_{j=1}^{r}\frac{h_j - \ln(1+h_j)}{\sqrt{-2\ln(1-c) - 2c}}\right), \qquad (21)
\end{aligned}$$

for any $h = (h_1, ..., h_r) \neq 0$ *such that* $h_j \in [0, \sqrt{c})$ *for* $j = 1, ..., r$.

Formula (20) for $\beta_{CM}(h)$ directly follows from Proposition 2 of Cai and Ma (2012). The proof of the other formulae follows along the same lines as in the proof of Proposition 10 in OMH, and is omitted. Except for the Tracy-Widom tests of Example 4, all those asymptotic power functions 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 $\beta_J(h)$ (solid line) to the power envelope $\beta_\mu(h)$ (dotted line). The Ledoit-Wolf test, the "corrected" likelihood ratio test, and the Cai-Ma test are invariant with respect to orthogonal transformations of the data only, and Figure 7 thus compares the asymptotic power functions $\beta_{LW}(h) = \beta_{CM}(h)$ and $\beta_{CLR}(h)$ (solid and dashed lines, respectively) to the power envelope $\beta_\lambda(h)$ (dotted line). Note that $\beta_{CLR}(h)$ depends on c. As c converges to one, $\beta_{CLR}(h)$ converges to α, which corresponds to the case of trivial power. As c converges to zero, $\beta_{CLR}(h)$ converges to $\beta_{LW}(h) = \beta_{CM}(h)$. In Figure 7, we provide plots of $\beta_{CLR}(h)$ that correspond to $c = 0.5$.

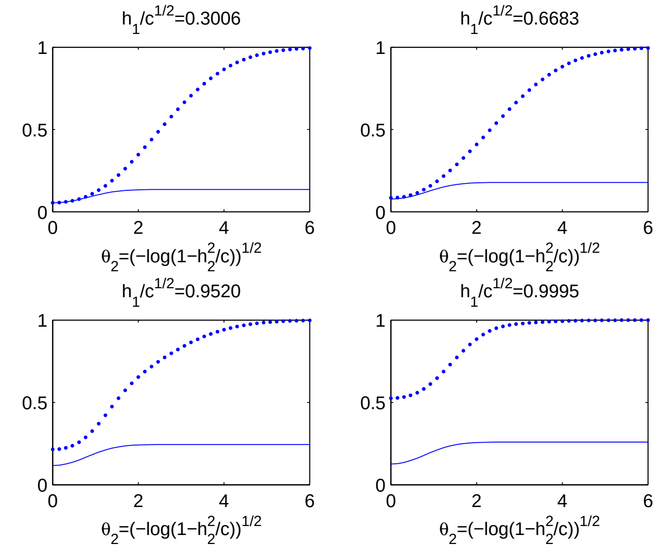

Figure 6: Profiles of the asymptotic power of John's test (solid lines) relative to the asymptotic power envelope β_μ (dotted lines) for different values of $h_1/\sqrt{c}$ under the alternative; $\alpha = 0.05$.

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 h_1 and h_2, and lie very far below the corresponding power envelope.

4 Conclusion

This paper extends Onatski, Moreira and Hallin's (2011) (OMH) 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 maximal asymptotic power envelope and for the asymptotic powers of several tests proposed in the literature. These asymptotic powers turn out to be very substantially below the envelope. We propose the likelihood ratio test based on the data reduced to the eigenvalues of the sample covariance matrix. Our computations show that the asymptotic power

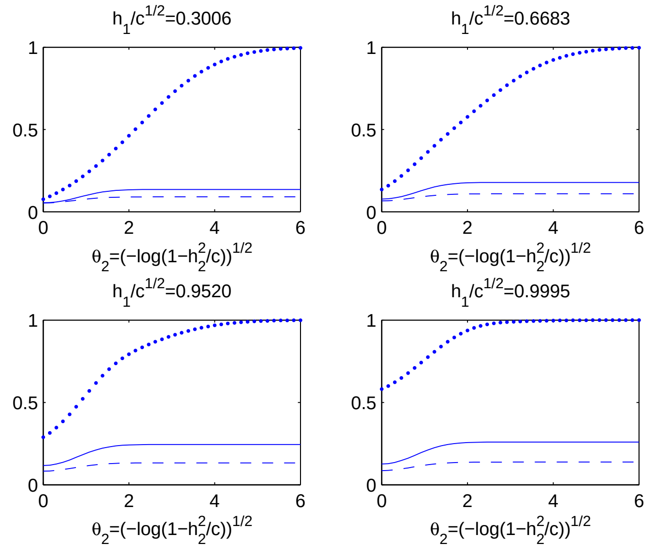

Figure 7: Profiles of the asymptotic power of the Ledoit-Wolf and Cai-Ma tests (solid lines) and the CLR test (dashed lines, for $c = 0.5$) relative to the asymptotic power envelope β_λ (dotted lines) for different values of $h_1/\sqrt{c}$ under the alternative; $\alpha = 0.05$.

of this test is close to the envelope.

5 Appendix

All convergence statements made below refer to the situation when $n_p, p \to_c \infty$. We start with two auxiliary results.

Lemma 7 *Let $d(\mu, \nu)$ be the Dudley distance between measures μ and ν defined over $(\mathbb{R}, \mathcal{B})$:*

$$d(\mu, \nu) = \sup \left\{ \left| \int f(d\mu - d\nu) \right| : f(x) \leq 1 \text{ and } \left| \frac{f(x) - f(y)}{x - y} \right| \leq 1, \forall x \neq y \right\}.$$

There exists a constant $\tau > 0$ such that $d\left(\hat{F}_p^\lambda, F_p^{MP}\right) = o\left(p^{-1} \log^\tau p\right)$ a.s..

Proof: Let us denote the cumulative distribution function corresponding to a measure μ as $F_\mu(x)$. Further, let us denote $\inf\{|x_2 - x_1| : \text{supp}(\mu) \subseteq [x_1, x_2]\}$ as $\text{diam}(\mu)$. Consider the following three distances between measures μ and ν : the

Kolmogorov distance $k(\mu,\nu) = \sup_x |F_\mu(x) - F_\nu(x)|$, the Wasserstein distance $w(\mu,\nu) = \sup\left\{\left|\int f(d\mu - d\nu)\right| : \left|\frac{f(x)-f(y)}{x-y}\right| \le 1, \forall x \ne y\right\}$, and the Kantorovich distance $\gamma(\mu,\nu) = \int |F_\mu(x) - F_\nu(x)|\, dx$. As is well known (see, for example, exercise 1 on p.425 of Dudley (2002)), $w(\mu,\nu) = \gamma(\mu,\nu)$. Therefore, we have

$$d\left(\hat{F}_p^\lambda, F_p^{MP}\right) \le w\left(\hat{F}_p^\lambda, F_p^{MP}\right) = \gamma\left(\hat{F}_p^\lambda, F_p^{MP}\right) \le k\left(\hat{F}_p^\lambda, F_p^{MP}\right)\left(\mathrm{diam}\left(\hat{F}_p^\lambda\right) + \mathrm{diam}\left(F_p^{MP}\right)\right).$$

As follows from Theorem 1.1 of Götze and Tikhomirov (2011), there exists a constant $\tau > 0$ such that $\sum_{p=1}^\infty \Pr\left(k\left(\hat{F}_p^\lambda, F_p^{MP}\right) > \varepsilon p^{-1}\log^\tau p\right) < \infty$ for all $\varepsilon > 0$. Thus, $k\left(\hat{F}_p^\lambda, F_p^{MP}\right) = o\left(p^{-1}\log^\tau p\right)$ a.s.. Since $\mathrm{diam}\left(F_p^{MP}\right)$ is $O(1)$ and $\mathrm{diam}\left(\hat{F}_p^\lambda\right) - \mathrm{diam}\left(F_p^{MP}\right) \to 0$ a.s., the result follows.□

Corollary 8 *Suppose that a sequence of functions $\{f_p(\lambda)\}$ is bounded Lipshitz on $\mathrm{supp}\left(F_p^{MP}\right) \cup \mathrm{supp}\left(\hat{F}_p^\lambda\right)$, uniformly over all sufficiently large p, a.s.. Then $\left|\int f_p(\lambda)\mathrm{d}\left(\hat{F}_p^\lambda(\lambda) - F_p^{MP}(\lambda)\right)\right| = o\left(p^{-1/2}\right)$, a.s..*

5.1 Proof of Proposition 2

Let us denote the integral $\int_{\mathcal{O}(p)} e^{p\,\mathrm{tr}(\Theta_p Q'\Lambda_p Q)}(\mathrm{d}Q)$ as $I_p(\Theta_p, \Lambda_p)$. As explained in Guionnet and Maida (2005, p.454), we can write

$$I_p(\Theta_p, \Lambda_p) = \mathbb{E}_{\Lambda_p}\exp\left\{p\sum_{j=1}^r \theta_{pj}\frac{\tilde{g}^{(j)\prime}\Lambda_p\tilde{g}^{(j)}}{\tilde{g}^{(j)\prime}\tilde{g}^{(j)}}\right\}, \tag{22}$$

where $\mathbb{E}_{\Lambda_p}$ denotes the expectation conditional on Λ_p, and the p-dimensional vectors $\left(\tilde{g}^{(1)}, ..., \tilde{g}^{(r)}\right)$ are obtained from standard Gaussian p-dimensional vectors $\left(g^{(1)}, ..., g^{(r)}\right)$, 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

$$\sum_{k=1}^{j-1} A_{jk}g^{(k)\prime}g^{(t)} = -g^{(j)\prime}g^{(t)} \text{ for } t = 1, ..., j-1. \tag{23}$$

In the spirit of the proof of Guionnet and Maida's (2005) Theorem 3, define

$$\gamma_{p1}^{(j,s)} = \sqrt{p}\left(\frac{1}{p}g^{(j)\prime}g^{(s)} - \delta_{js}\right) \text{ and } \gamma_{p2}^{(j,s)} = \sqrt{p}\left(\frac{1}{p}g^{(j)\prime}\Lambda_p g^{(s)} - v_{pj}\delta_{js}\right), \tag{24}$$

where $\delta_{js} = \mathbf{1}\{j = s\}$ stands for the classical Kronecker symbol. As will be shown below, after an appropriate change of measure, $\gamma_{p1}^{(j,s)}$ and $\gamma_{p2}^{(j,s)}$ are asymptotically centered Gaussian. Expressing the exponent in (22) 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 $\gamma_p = \left(\gamma_p^{(1,1)}, ..., \gamma_p^{(r,1)}, \gamma_p^{(2,2)}, ..., \gamma_p^{(r,2)}, \gamma_p^{(3,3)}, ..., \gamma_p^{(r,r)}\right)'$, where $\gamma_p^{(j,s)} = \left(\gamma_{p1}^{(j,s)}, \gamma_{p2}^{(j,s)}\right)$. Using this notation, (22), (23), and (24), we get, after some algebra,

$$I_p(\Theta_p, \Lambda_p) = \int f_{p,\theta}(\gamma_p)\, e^{p\sum_{j=1}^r \theta_{pj}\left(v_{pj} + \hat{\gamma}_p^{(j,j)} - v_{pj}\gamma_p^{(j,j)}\right)} \prod_{j=1}^r \prod_{i=1}^p \mathrm{d}\mathbb{P}\left(g_i^{(j)}\right), \tag{25}$$

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

$$\begin{aligned}
f_{p,\theta}(\gamma_p) &= \exp\left\{\sum_{j=1}^r \theta_{pj} \frac{N_{1j} + ... + N_{6j}}{D_j}\right\} \text{ with} \qquad (26)\\
N_{1j} &= -\gamma_{p1}^{(j,j)}\left(\gamma_{p2}^{(j,j)} - v_{pj}\gamma_{p1}^{(j,j)}\right),\\
N_{2j} &= \gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)} + I\right)^{-1}\left(G_{p2}^{(j)} + W_{pj}\right)\left(G_{p1}^{(j)} + I\right)^{-1}\gamma_{p1}^{(j,1:j-1)},\\
N_{3j} &= -2\gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)} + I\right)^{-1}\gamma_{p2}^{(j,1:j-1)},\\
N_{4j} &= v_{pj}\gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)} + I\right)^{-1}\gamma_{p1}^{(j,1:j-1)},\\
N_{5j} &= p^{-1/2}\gamma_{p2}^{(j,j)}\gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)} + I\right)^{-1}\gamma_{p1}^{(j,1:j-1)},\\
N_{6j} &= -p^{-1/2}v_{pj}\gamma_{p1}^{(1:j-1,j)\prime}\left(G_{p1}^{(j)} + I\right)^{-1}\gamma_{p1}^{(1:j-1,j)}\gamma_{p1}^{(j,j)}, \text{ and}\\
D_j &= 1 + p^{-1/2}\gamma_{p1}^{(j,j)} - p^{-1}\gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)} + I\right)^{-1}\gamma_{p1}^{(j,1:j-1)},
\end{aligned}$$

where $G_{pi}^{(j)}$ is a $(j-1)\times(j-1)$ matrix with (k,s)-th element $p^{-1/2}\gamma_{pi}^{(k,s)}$,

$$W_{pj} = \operatorname{diag}(v_{p1}, ..., v_{p,j-1}), \text{ and } \gamma_{pi}^{(j,1:j-1)} = \left(\gamma_{pi}^{(j,1)}, ..., \gamma_{pi}^{(j,j-1)}\right)'.$$

Next, define the event

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

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

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

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. 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

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

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

$$\begin{aligned} v_{pj} + \frac{1}{2\theta_{pj}} &> \left(1 + \sqrt{c}\right)^2 + \hat{\varepsilon} \text{ when } \theta_{pj} > 0 \text{ and} \\ v_{pj} + \frac{1}{2\theta_{pj}} &< -\hat{\varepsilon} \text{ when } \theta_{pj} < 0. \end{aligned}$$

Recall that $\lambda_{pp} \geq 0$, and $\lambda_{p1} \to (1 + \sqrt{c})^2$ a.s.. Therefore, still a.s., for sufficiently large p, $v_{pj} + \frac{1}{2\theta_{pj}} > \lambda_{p1}$ when $\theta_{pj} > 0$ and $v_{pj} + \frac{1}{2\theta_{pj}} < \lambda_{pp}$ when $\theta_{pj} < 0$. 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

$$I_p^{M,M'}(\Theta_p, \Lambda_p) = e^{p\sum_{j=1}^r\left[\theta_{pj}v_{pj} - \frac{1}{2p}\sum_{i=1}^p \ln(1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_{pi})\right]} J_p^{M,M'}, \tag{27}$$

where

$$J_p^{M,M'} = \int \mathbf{1}\{B_{M,M'}\} f_{p,\theta}(\gamma_p) \prod_{j=1}^r \prod_{i=1}^p \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right). \tag{28}$$

We now show that, under $\prod_{j=1}^r \prod_{i=1}^p d\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right)$, γ_p a.s. converges in distribution to a centered $r^2 + r$-dimensional Gaussian vector, so that $J_p^{M,M'}$ is asymptotically equivalent to an integral with respect to a Gaussian measure on $\mathbb{R}^{r^2+r}$.

First, let us find the mean $\mathbb{E}_p\gamma_p$, and the variance $\mathbb{V}_p\gamma_p$ of γ_p under measure $\prod_{j=1}^r\prod_{i=1}^p \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right)$. Note that $\mathbb{V}_p\gamma_p = \mathrm{diag}\left(\mathbb{V}_p\gamma_p^{(1,1)}, \mathbb{V}_p\gamma_p^{(2,1)}, ..., \mathbb{V}_p\gamma_p^{(r,r)}\right)$ and $\mathbb{E}_p\gamma_p = \left(\mathbb{E}_p\gamma_p^{(1,1)}, \mathbb{E}_p\gamma_p^{(2,1)}, ..., \mathbb{E}_p\gamma_p^{(r,r)}\right)'$. With probability one, for sufficiently large p, we have

$$
\begin{aligned}
\mathbb{E}_p\gamma_{p1}^{(k,s)} &= \sqrt{p}\delta_{ks}\left(\frac{1}{p}\sum_{i=1}^p \frac{1}{(1+2\theta_{pk}v_{pk}-2\theta_{pk}\lambda_{pi})} - 1\right) \\
&= \sqrt{p}\delta_{ks}\int \frac{(2\theta_{pk})^{-1}}{K_p^{MP}(2\theta_{pk})-\lambda}\mathrm{d}\left(\hat{F}_p^\lambda(\lambda) - F_p^{MP}(\lambda)\right),
\end{aligned}
$$

which, by Corollary 1, is $o(1)$ uniformly in $2\theta_{pk}\in\Omega_{\varepsilon\eta}$, a.s.. That Corollary 1 can be applied here follows from the form of expression (7) for $K_p^{MP}(x)$. Similarly,

$$
\mathbb{E}_p\gamma_{p2}^{(k,s)} = \sqrt{p}\frac{\delta_{ks}}{2\theta_{pk}}\int \frac{K_p^{MP}(2\theta_{pk})}{K_p^{MP}(2\theta_{pk})-\lambda}\mathrm{d}\left(\hat{F}_p^\lambda(\lambda) - F_p^{MP}(\lambda)\right) = o(1)
$$

uniformly in $2\theta_{pk}, 2\theta_{ps}\in\Omega_{\varepsilon\eta}$, a.s.. Thus,

$$
\sup_{\{2\theta_{pj}\in\Omega_{\varepsilon\eta}, j\le r\}} \mathbb{E}_p\gamma_p = o(1) \text{ a.s..} \tag{29}
$$

Next, with probability one, for sufficiently large p we have

$$
\mathbb{V}_p\gamma_{p1}^{(k,s)} = \frac{1}{p}\sum_{i=1}^p \frac{1+\delta_{ks}}{(1+2\theta_{pk}v_{pk}-2\theta_{pk}\lambda_{pi})(1+2\theta_{ps}v_{ps}-2\theta_{ps}\lambda_{pi})}.
$$

Let $\hat{H}_{p,ks}^{(2)} = \int \frac{\mathrm{d}\hat{F}_p^\lambda(\lambda)}{\left(K_p^{MP}(2\theta_{pk})-\lambda\right)\left(K_p^{MP}(2\theta_{ps})-\lambda\right)}$ and $H_{p,ks}^{(2)} = \int \frac{\mathrm{d}F_p^{MP}(\lambda)}{\left(K_p^{MP}(2\theta_{pk})-\lambda\right)\left(K_p^{MP}(2\theta_{ps})-\lambda\right)}$. Then, using Corollary 1, we get

$$
\mathbb{V}_p\gamma_{p1}^{(k,s)} = \frac{1+\delta_{ks}}{4\theta_{pk}\theta_{ps}}\hat{H}_{p,ks}^{(2)} = \frac{1+\delta_{ks}}{4\theta_{pk}\theta_{ps}}H_{p,ks}^{(2)} + o(1) \text{ a.s.,}
$$

uniformly in $2\theta_{pk}, 2\theta_{ps}\in\Omega_{\varepsilon\eta}$. Similarly, we have

$$
\begin{aligned}
\mathbb{V}_p\gamma_{p2}^{(k,s)} &= \frac{1}{p}\sum_{i=1}^p \frac{\lambda_{pi}^2(1+\delta_{ks})}{(1+2\theta_{pk}v_{pk}-2\theta_{pk}\lambda_{pi})(1+2\theta_{ps}v_{ps}-2\theta_{ps}\lambda_{pi})} \\
&= \frac{1+\delta_{ks}}{4\theta_{pk}\theta_{ps}}\left(1+K_p^{MP}(2\theta_{ps})K_p^{MP}(2\theta_{pk})H_{p,ks}^{(2)} - 2\theta_{pk}K_p^{MP}(2\theta_{pk}) - 2\theta_{ps}K_p^{MP}(2\theta_{ps})\right) + o(1),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}_p\left(\gamma_{p1}^{(k,s)},\gamma_{p2}^{(k,s)}\right) &= \frac{1}{p}\sum_{i=1}^{p}\frac{\lambda_{pi}\left(1+\delta_{ks}\right)}{\left(1+2\theta_{pk}\upsilon_{pk}-2\theta_{pk}\lambda_{pi}\right)\left(1+2\theta_{ps}\upsilon_{ps}-2\theta_{ps}\lambda_{pi}\right)}\\
&= \frac{\left(1+\delta_{ks}\right)}{4\theta_{pk}\theta_{ps}}\left(K_p^{MP}\left(2\theta_{ps}\right)H_{p,ks}^{(2)}-2\theta_{pk}\right)+o(1),
\end{aligned}
$$

uniformly in $2\theta_{pk},2\theta_{ps}\in\Omega_{\varepsilon\eta}$, a.s..

A straightforward calculation, using formula (7), shows that

$$
H_{p,ks}^{(2)}=\left(\frac{1}{4\theta_{pk}\theta_{ps}}-c_p\upsilon_{pk}\upsilon_{sk}\right)^{-1}\text{, and }\mathbb{V}_p\gamma_p^{(k,s)}=V_p^{(k,s)}+o(1), \tag{30}
$$

uniformly in $2\theta_{pk},2\theta_{ps}\in\Omega_{\varepsilon\eta}$, a.s., where the matrix $V_p^{(k,s)}$ has elements

$$
V_{p,11}^{(k,s)} = (1+\delta_{ks})\left(1-4\theta_{pk}\upsilon_{pk}\theta_{ps}\upsilon_{sk}c_p\right)^{-1}, \tag{31}
$$

$$
V_{p,12}^{(k,s)} = V_{p,21}^{(k,s)}=(1+\delta_{ks})\,\upsilon_{pk}\upsilon_{sk}\left(1-4\theta_{pk}\upsilon_{pk}\theta_{ps}\upsilon_{sk}c_p\right)^{-1},\text{ and} \tag{32}
$$

$$
V_{p,22}^{(k,s)} = (1+\delta_{ks})\left[c_p\upsilon_{pk}\upsilon_{sk}+\upsilon_{pk}^2\upsilon_{sk}^2\left(1-4\theta_{pk}\upsilon_{pk}\theta_{ps}\upsilon_{sk}c_p\right)^{-1}\right]. \tag{33}
$$

This implies that

$$
\det\left(V_p^{(k,s)}\right)=\prod_{k\geq s}^{r}(1+\delta_{ks})^2\,c_p\upsilon_{pk}\upsilon_{sk}\left(1-4\theta_{pk}\upsilon_{pk}\theta_{ps}\upsilon_{sk}c_p\right)^{-1}, \tag{34}
$$

which is bounded away from zero and infinity for sufficiently large p, uniformly over $\{2\theta_{pj}\in\Omega_{\varepsilon\eta},j\leq r\}$, a.s..

By construction, γ_p is a sum of p independent random vectors having uniformly bounded third and fourth absolute moments under measure $\prod_{j=1}^{r}\prod_{i=1}^{p}\mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right)$. Therefore, a central limit theorem applies. Moreover, since the function $f_{p,\theta}\left(\gamma_p\right)$ is Lipshitz over $B_{M,M'}$, uniformly in $\{2\theta_{pj}\in\Omega_{\varepsilon\eta},j\leq r\}$, Theorem 13.3 of Bhattacharya and Rao (1976), which describes the accuracy of the Gaussian approximations to integrals of the form (28) in terms of the oscillation measures of the integrand, implies that

$$
J_p^{M,M'}=\int_{B_{M,M'}}f_{p,\theta}\left(x\right)\mathrm{d}\Phi\left(x;\mathbb{E}_p\gamma_p,\mathbb{V}_p\gamma_p\right)+o_{M,M'}\left(1\right), \tag{35}
$$

where $\Phi\left(x;\mathbb{E}_p\gamma_p,\mathbb{V}_p\gamma_p\right)$ denotes the Gaussian distribution function with mean

$\mathbb{E}_p\gamma_p$ and variance $\mathbb{V}_p\gamma_p$, and $o_{M,M'}(1)$ converges to zero uniformly in $\{2\theta_{pj} \in \Omega_{\varepsilon\eta}, j \leq r\}$ as $p \to \infty$, a.s.. The rate of such a convergence may depend on the values of M and M'.

Note that, in $B_{M,M'}$, as $p \to \infty$, the difference $f_{p,\theta}(\gamma_p) - \overline{f}_{p,\theta}(\gamma_p)$ converges to zero uniformly over $\{2\theta_{pj} \in \Omega_{\varepsilon\eta}, j \leq r\}$, where

$$\begin{aligned}
\overline{f}_{p,\theta}(\gamma_p) &= \exp\left\{\sum_{j=1}^{r} \theta_{pj}\left(\bar{N}_{1j} + ... + \bar{N}_{4j}\right)\right\}, \text{ with} \qquad (36) \\
\bar{N}_{1j} &= -\gamma_1^{(j,j)}\left(\gamma_2^{(j,j)} - v_{pj}\gamma_1^{(j,j)}\right), \bar{N}_{2j} = \gamma_1^{(j,1:j-1)\prime} W_{pj}\gamma_1^{(j,1:j-1)}, \\
\bar{N}_{3j} &= -2\gamma_1^{(j,1:j-1)\prime}\gamma_2^{(j,1:j-1)}, \text{ and } \bar{N}_{4j} = v_{pj}\gamma_1^{(j,1:j-1)\prime}\gamma_1^{(j,1:j-1)}.
\end{aligned}$$

Such a convergence, together with (29), (30), and (35) implies that

$$J_p^{M,M'} = \int_{B_{M,M'}} \overline{f}_{p,\theta}(x)\, \mathrm{d}\Phi(x; 0, V_p) + o_{M,M'}(1), \qquad (37)$$

where $V_p = \operatorname{diag}\left(V_p^{(1,1)}, V_p^{(2,1)}, ..., V_p^{(r,r)}\right)$.

Note that the difference $\int_{B_{M,M'}} \overline{f}_{p,\theta}(x)\, \mathrm{d}\Phi(x;0,V_p) - \int_{\mathbb{R}^{r^2+r}} \overline{f}_{p,\theta}(x)\, \mathrm{d}\Phi(x;0,V_p)$ converges to zero as $M, M' \to \infty$, uniformly in p for p sufficiently large. On the other hand,

$$\int_{\mathbb{R}^{r^2+r}} \overline{f}_{p,\theta}(x)\, \mathrm{d}\Phi(x; 0, V_p) = \prod_{j=1}^{r}\prod_{s=1}^{j}\int_{\mathbb{R}^2} \frac{\exp\left[-\frac{1}{2}y'\left(W_p^{(j,s)}\right)^{-1}y\right]}{2\pi\sqrt{\det\left(V_p^{(j,s)}\right)}}\mathrm{d}y, \qquad (38)$$

where

$$\left(W_p^{(j,s)}\right)^{-1} = \left(V_p^{(j,s)}\right)^{-1} + (1+\delta_{js})^{-1}\begin{pmatrix} -2\theta_{pj}(v_{pj}+v_{ps}) & 2\theta_{pj} \\ 2\theta_{pj} & 0 \end{pmatrix}.$$

Using (31-33), we verify that, for sufficiently large p, $W_p^{(j,s)}$ is a.s. positive definite, and

$$\begin{aligned}
\det\left(W_p^{(j,s)}\right) &= (1+\delta_{js})^2 c_p v_{pj} v_{ps}, \text{ and} \qquad (39) \\
\det\left(V_p^{(j,s)}\right) &= (1+\delta_{js})^2 c_p v_{pj} v_{ps} \left(1 - 4(\theta_{pj}v_{pj})(\theta_{ps}v_{ps})\, c_p\right)^{-1}. \qquad (40)
\end{aligned}$$

Therefore,

$$\int_{\mathbb{R}^{r^2+r}} \overline{f}_{p,\theta}(x)\, \mathrm{d}\Phi(x;0,V_p) = \prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1-4\left(\theta_{pj}\upsilon_{pj}\right)\left(\theta_{ps}\upsilon_{ps}\right)c_p}$$

and, uniformly in p for p sufficiently large,

$$\lim_{M,M'\to\infty}\left\{\int_{B_{M,M'}} \overline{f}_{p,\theta}(x)\, \mathrm{d}\Phi(x;0,V_p) - \prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1-4\left(\theta_{pj}\upsilon_{pj}\right)\left(\theta_{ps}\upsilon_{ps}\right)c_p}\right\} = 0. \quad (41)$$

Equations (27), (37), and (41) describe the behavior of $I_p^{M,M'}(\Theta_p,\Lambda_p)$ for large p, M, and M'.

Let us now turn to the analysis of $I_p(\Theta_p,\Lambda_p) - I_p^{M,M'}(\Theta_p,\Lambda_p)$. Let B_M be the event $\left\{\left|\gamma_{p1}^{(j,s)}\right| \le M \text{ for all } j,s\le r\right\}$, and let

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

As explained in Guionnet and Maida (2005, p.455), $\gamma_{p1}^{(j,s)}$, $j,s=1,...,r$ are independent of $\tilde{g}^{(j)\prime}\Lambda\tilde{g}^{(j)}/\tilde{g}^{(j)\prime}\tilde{g}^{(j)}$, $j=1,...,r$. Therefore,

$$I_p^M(\Theta_p,\Lambda_p) = \mathbb{E}_{\Lambda_p}\left(\mathbf{1}\{B_M\}\right)I_p(\Theta_p,\Lambda_p) = \left(1-\mathbb{E}_{\Lambda_p}\left(\mathbf{1}\{B_M^c\}\right)\right)I_p(\Theta_p,\Lambda_p).$$

Denoting again by $\mathbb{P}$ the centered standard Gaussian measure on $\mathbb{R}$, we have

$$\mathbb{E}_{\Lambda_p}\left(\mathbf{1}\left\{\left|\gamma_{p1}^{(j,s)}\right|\ge M\right\}\right) = \int \mathbf{1}\left\{\left|\gamma_{p1}^{(j,s)}\right|\ge M\right\}\prod_{j=1}^{r}\prod_{i=1}^{p}\mathrm{d}\mathbb{P}\left(g_i^{(j)}\right).$$

For $j\ne s$ and $\tau\in\left(-\frac{1}{2}\sqrt{p},\frac{1}{2}\sqrt{p}\right)$,

$$\begin{aligned}\int e^{\tau\gamma_{p1}^{(j,s)}}\prod_{j=1}^{r}\prod_{i=1}^{p}\mathrm{d}\mathbb{P}\left(g_i^{(j)}\right) &= \frac{1}{(2\pi)^p}\int e^{\tau\frac{1}{\sqrt{p}}g^{(j)\prime}g^{(s)}}e^{-\frac{1}{2}\left(g^{(j)\prime}g^{(j)}+g^{(s)\prime}g^{(s)}\right)}\prod_{i=1}^{p}\left(\mathrm{d}g_i^{(j)}\mathrm{d}g_i^{(s)}\right)\\ &= \left(1-\frac{\tau^2}{p}\right)^{-\frac{p}{2}} \le e^{2\tau^2}.\end{aligned}$$

Therefore, using Chebyshev's inequality, for $j\ne s$ and $\tau\in\left(-\frac{1}{2}\sqrt{p},\frac{1}{2}\sqrt{p}\right)$,

$$\int \mathbf{1}\left\{\gamma_{p1}^{(j,s)}\ge M\right\}\prod_{j=1}^{r}\prod_{i=1}^{p}\mathrm{d}\mathbb{P}\left(g_i^{(j)}\right)\le\frac{e^{2\tau^2}}{e^{M\tau}}.$$

Setting $\tau = M/4$ (here we assume that $M<2\sqrt{p}$), we get

$$\int \mathbf{1}\left\{\gamma_{p1}^{(j,s)}\ge M\right\}\prod_{j=1}^{r}\prod_{i=1}^{p}\mathrm{d}\mathbb{P}\left(g_i^{(j)}\right)\le e^{-M^2/8}.$$

Similarly, we show that the same inequality holds when $\gamma_{p1}^{(j,s)}$ is replaced by $-\gamma_{p}^{(j,s)}$, and thus

$$\int \mathbf{1}\left\{\left|\gamma_{p1}^{(j,s)}\right| \geq M\right\} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}\left(g_i^{(j)}\right) \leq 2e^{-M^2/8}. \tag{42}$$

For $j = s$, the same line of arguments yields

$$\int \mathbf{1}\left\{\left|\gamma_{p}^{(j,j)}\right| \geq M\right\} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}\left(g_i^{(j)}\right) \leq 2e^{-M^2/16}. \tag{43}$$

Inequalities (42) and (43) imply that $\mathbb{E}_{\Lambda_p}\left(\mathbf{1}\left\{B_M^c\right\}\right) \leq 2r^2 e^{-M^2/16}$, and therefore, for sufficiently large p,

$$I_p\left(\Theta_p, \Lambda_p\right) \geq I_p^M\left(\Theta_p, \Lambda_p\right) \geq \left(1 - 2r^2 e^{-M^2/16}\right) I_p\left(\Theta_p, \Lambda_p\right). \tag{44}$$

Note that

$$I_p^M\left(\Theta_p, \Lambda_p\right) = e^{p\sum_{j=1}^{r}\left[\theta_{pj}\upsilon_{pj} - \frac{1}{2p}\sum_{i=1}^{p}\ln(1+2\theta_{pj}\upsilon_{pj}-2\theta_{pj}\lambda_i)\right]}\left(J_p^{M,M'} + J_p^{M,M',\infty}\right), \tag{45}$$

where

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

We will now derive an upper bound for $J_p^{M,M',\infty}$.

From the definition of $f_{p,\theta}\left(\gamma_p\right)$, we see that there exist positive constants β_1 and β_2, which may depend on r, ε and η, such that for any θ_{pj} satisfying $2\theta_{pj} \in \Omega_{\varepsilon\eta}$, $j \leq r$ and for sufficiently large p, when B_M holds,

$$f_{p,\theta}\left(\gamma_p\right) \leq \exp\left\{\beta_1 M \sum_{s,k=1}^{r}\left|\gamma_{p2}^{(k,s)}\right| + \beta_2 M^2\right\}.$$

Let $B_{M,M'}^{(k,s)} = B_M \cap \left\{\left|\gamma_{p2}^{(k,s)}\right| = \max_{j,m\leq r}\left|\gamma_{p2}^{(j,m)}\right| > M'\right\}$. Clearly, $B_M \backslash B_{M,M'} = \bigcup_{k,s=1}^{r} B_M^{(k,s)}$. Therefore,

$$\begin{aligned}
J_p^{M,M',\infty} &\leq \sum_{k,s=1}^{r}\int_{B_{M,M'}^{(k,s)}} e^{\beta_1 M r^2\left|\gamma_{p2}^{(k,s)}\right| + \beta_2 M^2} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right) \\
&\leq \sum_{j,m=1}^{r}\int_{\left|\gamma_{p2}^{(k,s)}\right|\geq M'} e^{\beta_1 M r^2\left|\gamma_{p2}^{(k,s)}\right| + \beta_2 M^2} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right).
\end{aligned}$$

First assume $k \neq s$. Denote $\lambda_{pi}\left(1 - 2\theta_{pk}\lambda_{pi} + 2\theta_{pk}\upsilon_{pk}\right)^{-1/2}\left(1 - 2\theta_{ps}\lambda_{pi} + 2\theta_{ps}\upsilon_{ps}\right)^{-1/2}$ as $\tilde{\lambda}_{pi}$ and $\left(1 - 2\theta_{pj}\lambda_{pi} + 2\theta_{pj}\upsilon_{pj}\right)^{1/2} g_i^{(j)}$ as $\tilde{g}_i^{(j)}$. Note that, under $\mathbb{P}_{pi}^{(j)}$, $\tilde{g}_i^{(j)}$ is a standard normal random variable. Further, as long as $2\theta_{pj} \in \Omega_{\varepsilon\eta}$ for $j \le r$, $\tilde{\lambda}_{pi}$ considered as a function of λ_i is continuous on $\lambda_i \in \operatorname{supp} \hat{F}_p^\lambda$ for sufficiently large p, a.s.. Hence, the empirical distribution of $\tilde{\lambda}_i$ converges. Moreover, $\tilde{\lambda}_{\max} = \max_{i=1,\dots,p}(\tilde{\lambda}_{pi})$ and $\tilde{\lambda}_{\min} = \min_{i=1,\dots,p}(\tilde{\lambda}_{pi})$ a.s. converge to finite real numbers. Now, for τ such that $|\tau| < \sqrt{p}/(2\tilde{\lambda}_{\max})$, we have

$$\int e^{\tau\gamma_{p2}^{(k,s)}} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right) = \mathbb{E}e^{\tau\sqrt{p}\frac{1}{p}\sum_{i=1}^{p}\tilde{\lambda}_{pi}\tilde{g}_i^{(k)}\tilde{g}_i^{(s)}}$$

$$= \prod_{i=1}^{p}\mathbb{E}e^{\tau\frac{1}{\sqrt{p}}\tilde{\lambda}_{pi}\tilde{g}_i^{(k)}\tilde{g}_i^{(s)}} = \prod_{i=1}^{p}\left(1-\tau^2\frac{\tilde{\lambda}_{pi}^2}{p}\right)^{-1/2} \le e^{2\tilde{\lambda}_{\max}^2\tau^2}$$

for sufficiently large p, a.s.. Using this inequality, we get, for sufficiently large p and any positive t such that $\beta_1 r^2 M + t < \sqrt{p}/(2\tilde{\lambda}_{\max})$,

$$\int_{\gamma_{p2}^{(k,s)}\ge M'} e^{\beta_1 r^2 M\gamma_{p2}^{(k,s)}} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right) \le \int e^{\beta_1 r^2 M\gamma_{p2}^{(k,s)} + t\left(\gamma_{p2}^{(k,s)} - M'\right)} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right)$$

$$= e^{-tM'}\int e^{\left(\beta_1 r^2 M + t\right)\gamma_{p2}^{(k,s)}} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right) \le e^{-tM'}e^{2\tilde{\lambda}_{\max}^2\left(\beta_1 r^2 M + t\right)^2}.$$

Setting $t = \frac{M'}{4\tilde{\lambda}_{\max}^2} - \beta_1 r^2 M$ (here we assume that M and M' are such that t satisfies the above requirements), we get

$$\int_{\gamma_{p2}^{(k,s)}\ge M'} e^{\beta_1 r^2 M\gamma_{p2}^{(k,s)}} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right) \le e^{-\frac{(M')^2}{8\tilde{\lambda}_{\max}^2} + \beta_1 r^2 MM'}.$$

Replacing $\gamma_{p2}^{(k,s)}$ by $-\gamma_{p2}^{(k,s)}$ in the above derivations and combining the result with the above inequality, we get

$$\int_{\left|\gamma_{p2}^{(k,s)}\right|\ge M'} e^{\beta_1 r^2 M\left|\gamma_{p2}^{(k,s)}\right|} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right) \le 2e^{-\frac{(M')^2}{8\tilde{\lambda}_{\max}^2} + \beta_1 r^2 MM'}.$$

When $k = s$, following a similar line of arguments, we obtain

$$\int_{\left|\gamma_{p2}^{(k,k)}\right|\ge M'} e^{\beta_1 r^2 M\left|\gamma_{p2}^{(k,k)}\right|} \prod_{j=1}^{r}\prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right) \le 4e^{-\frac{(M')^2}{16\tilde{\lambda}_{\max}^2} + \beta_1 r^2 MM'}.$$

and thus, for sufficiently large p,

$$J_p^{M,M',\infty} \le 4r^2 e^{-\frac{(M')^2}{16\tilde{\lambda}_{\max}^2} + \beta_1 r^2 MM'}. \tag{46}$$

Finally, combining (44), (45), and (46), we obtain for

$$J_p = I_p\left(\Theta_p, \Lambda_p\right) e^{-p\sum_{j=1}^{r}\left[\theta_{pj}v_{pj} - \frac{1}{2p}\sum_{i=1}^{p}\ln(1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_i)\right]} \tag{47}$$

the following upper and lower bounds:

$$J_p^{M,M'} \leq J_p \leq \left(1 - 2r^2 e^{-\frac{M^2}{16}}\right)^{-1}\left(J_p^{M,M'} + 4r^2 e^{-\frac{(M')^2}{16\bar{\lambda}_{\max}^2}+\beta_1 r^2 MM'}\right). \tag{48}$$

Let $\tau > 0$ be an arbitrarily small number. Equations (37) and (41) imply that there exist $\bar{M}$ and $\bar{M}'$ such that, for any $M > \bar{M}$ and $M' > \bar{M}'$,

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

for all sufficiently large p. Let us choose $M > \bar{M}$ and $M' > \bar{M}'$ so that

$$\begin{aligned}\left(1 - 2r^2 e^{-\frac{M^2}{16}}\right)^{-1} &< 2,\\ \left(1 - 2r^2 e^{-\frac{M^2}{16}}\right)^{-1} 4r^2 e^{-\frac{(M')^2}{16\bar{\lambda}_{\max}^2}+\beta_1 r^2 MM'} &< \frac{\tau}{4},\end{aligned}$$

and

$$\left[\left(1 - 2r^2 e^{-\frac{M^2}{16}}\right)^{-1} - 1\right] \sup_{\{2\theta_{pj}\in\Omega_{\varepsilon\eta}, j\leq r\}} \prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1-4\left(\theta_{pj}v_{pj}\right)\left(\theta_{ps}v_{ps}\right)c_p} < \frac{\tau}{4}$$

for all sufficiently large p, a.s.. Then, (48) implies that

$$\left|J_p - \prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1-4\left(\theta_{pj}v_{pj}\right)\left(\theta_{ps}v_{ps}\right)c_p}\right| < \tau \tag{49}$$

for all sufficiently large p, a.s.. Since τ can be chosen arbitrarily, we have, from (47) and (49),

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

where $o(1) \to 0$ as $p \to \infty$ uniformly in $\{2\theta_{pj} \in \Omega_{\varepsilon\eta}, j \leq r\}$, a.s..□

5.2 Proof of Theorem 3

Setting $\theta_{pj} = \frac{1}{2c_p}\frac{h_j}{1+h_j}$, we have $\upsilon_{pj} = 1 + h_j$, $\theta_{pj}\upsilon_{pj} = \frac{h_j}{2c_p}$, and

$$\ln\left(1+2\theta_{pj}\upsilon_{pj}-2\theta_{pj}\lambda_{pi}\right) = \ln\left(\frac{1}{c_p}\frac{h_j}{1+h_j}\right) + \ln\left(z_{0j} - \lambda_{pi}\right).$$

Further, by Lemma 11 and formula (3.3) of OMH, $\int \ln\left(z_{j0} - \lambda\right) \mathrm{d}F_p^{MP}(\lambda) = \frac{h_j}{c_p} - \frac{1}{c_p}\ln\left(1 + h_j\right) + \ln\frac{(1+h_j)c_p}{h_j}$ for sufficiently large p, a.s.. With these auxiliary results, formula (10) is a straightforward consequence of (3) and Proposition 2.

Turning to the proof of (11), consider the integrals

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

In what follows, we omit the subscript p in n_p to simplify notation. Note that $\mathcal{I}(0, \infty)$ is the integral appearing in expression (4) for $L_p\big(h; \mu_p\big)$. Let us now prove that, for some constant $\alpha > 0$,

$$\mathcal{I}(0, \infty) = \mathcal{I}\left(p - \alpha\sqrt{p}, p + \alpha\sqrt{p}\right)\left(1 + o(1)\right),\ \text{a.s.} \tag{50}$$

where $o(1)$ is uniform in $h \in [0, \sqrt{c} - \delta]^r$.

Since, by Corollary 1, $S_p/p \to 1$ a.s., the set H_δ is bounded from below, and $\lambda_{p1} \to (1 + \sqrt{c})^2$ a.s., there exists a constant $A_1 > 0$ that depends only on δ and r, such that $\inf_{[0,\sqrt{c}-\delta]^r} px \operatorname{tr}\left(D_p Q'\Lambda_p Q\right)/S_p \geq -A_1 x/2$ for all $x \geq 0$ and all sufficiently large p, a.s.. Therefore, for all $h \in [0, \sqrt{c} - \delta]^r$,

$$2\mathcal{I}(0, \infty) \geq \int_0^\infty x^{\frac{np}{2}-1} e^{-\frac{n+A_1}{2}x}\mathrm{d}x = \left(\frac{n + A_1}{2}\right)^{-\frac{np}{2}} \Gamma\left(\frac{np}{2}\right),$$

and, using Stirling's approximation, we get

$$\begin{aligned}
\mathcal{I}(0, \infty) &\geq \left(\frac{n + A_1}{2}\right)^{-\frac{np}{2}} \left(\frac{np}{2}\right)^{\frac{np}{2}} e^{-\frac{np}{2}} \left(\frac{4\pi}{np}\right)^{1/2} \left(1 + o(1)\right) \\
&= p^{\frac{np}{2}} e^{-\left(\frac{n}{2} + \frac{A_1}{2} - \frac{1}{4}\frac{A_1^2}{n}\right)p} \left(\frac{4\pi}{np}\right)^{1/2} \left(1 + o(1)\right),\ \text{a.s.}
\end{aligned} \tag{51}$$

Next, there exists a constant $A_2 > 0$ such that, for all $x \geq 0$ and all sufficiently large p, $\sup_{h\in[0,\sqrt{c}-\delta]^r} px \operatorname{tr}\left(D_p Q'\Lambda_p Q\right)/S_p \leq A_2 x/2$, a.s.. Therefore, a.s., for all sufficiently large p,

$$\mathcal{I}\left(p+\alpha\sqrt{p},\infty\right)\leq\int_{p+\alpha\sqrt{p}}^{\infty}x^{\frac{np}{2}-1}e^{-\frac{n-A_2}{2}x}\mathrm{d}x=\left(\frac{n-A_2}{2}\right)^{-\frac{np}{2}}\Gamma\left(\frac{np}{2},y\right),$$

where $\Gamma\left(\frac{np}{2},y\right)$ is the complementary incomplete Gamma function (see Olver 1997, p.45) with $y=\left(p+\alpha\sqrt{p}\right)\left(\frac{n-A_2}{2}\right)$. Hence, for sufficiently large p, $y>np/2+n\alpha\sqrt{p}/4$, and we can continue

$$\mathcal{I}\left(p+\alpha\sqrt{p},\infty\right)<\left(\frac{n-A_2}{2}\right)^{-\frac{np}{2}}\Gamma\left(\frac{np}{2},\frac{np}{2}+\frac{\alpha n\sqrt{p}}{4}\right),\text{ a.s.}$$

Now, $\Gamma\left(\beta,\gamma\right)\leq e^{-\gamma}\gamma^{\beta}/(\gamma-\beta+1)$ whenever $\beta>1$ and $\gamma>\beta-1$ (Olver 1997, p.70). Therefore, we have, for sufficiently large p,

$$\begin{aligned}\mathcal{I}\left(p+\alpha\sqrt{p},\infty\right) &< \left(1-\frac{A_2}{n}\right)^{-\frac{np}{2}}\frac{e^{-\frac{np}{2}-\frac{\alpha n\sqrt{p}}{4}}p^{\frac{np}{2}}\left(1+\frac{\alpha}{2\sqrt{p}}\right)^{\frac{np}{2}}}{\alpha n\sqrt{p}/4+1}\\ &= p^{\frac{np}{2}}e^{\frac{A_2p}{2}+\frac{A_2^2p}{4n}}\frac{e^{-\frac{np}{2}-\frac{\alpha^2n}{16}+\frac{\alpha^3n}{48\sqrt{p}}-\frac{\alpha^4n}{128p}}}{\alpha n\sqrt{p}/4+1}\left(1+o(1)\right)\\ &< p^{\frac{np}{2}}e^{-\frac{np}{2}}\frac{e^{p\left(A_2-\frac{\alpha^2n}{32p}\right)}}{\alpha n\sqrt{p}/4+1}\left(1+o(1)\right),\text{ a.s.}\end{aligned}$$

Comparing this to (51), we see that α can be chosen so that

$$\mathcal{I}\left(p+\alpha\sqrt{p},\infty\right)=o(1)\mathcal{I}\left(0,\infty\right),\text{ a.s.}\tag{52}$$

Further, for sufficiently large p,

$$\begin{aligned}\mathcal{I}\left(0,p-\alpha\sqrt{p}\right) &\leq \int_0^{p-\alpha\sqrt{p}}x^{\frac{np}{2}-1}e^{-\frac{n-A_2}{2}x}\mathrm{d}x\\ &= \left(\frac{n-A_2}{2}\right)^{-\frac{np}{2}}\int_0^{y}t^{\frac{np}{2}-1}e^{-t}\mathrm{d}t,\text{ a.s.,}\end{aligned}$$

where $y=\left(p-\alpha\sqrt{p}\right)\frac{n-A_2}{2}<\frac{np}{2}-\frac{\alpha n\sqrt{p}}{4}$. Therefore, for any positive $z<\frac{np}{2}$ and sufficiently large p,

$$\begin{aligned}\mathcal{I}\left(0,p-\alpha\sqrt{p}\right) &\leq \left(\frac{n-A_2}{2}\right)^{-\frac{np}{2}}\int_0^{\frac{np}{2}-\frac{\alpha n\sqrt{p}}{4}}t^{\frac{np}{2}-1}e^{-t}\mathrm{d}t\\ &< \left(\frac{n-A_2}{2}\right)^{-\frac{np}{2}}\left(\frac{np}{2}-\frac{\alpha n\sqrt{p}}{4}\right)^{z}\Gamma\left(\frac{np}{2}-z\right).\end{aligned}$$

Setting $z=\alpha n\sqrt{p}/4$ and using Stirling's approximation, we have, a.s.,

$$\left(\frac{np}{2}-\frac{\alpha n\sqrt{p}}{4}\right)^{z}\Gamma\left(\frac{np}{2}-z\right)=\left(\frac{np}{2}-\frac{\alpha n\sqrt{p}}{4}\right)^{\frac{np}{2}-\frac{1}{2}}e^{-\frac{np}{2}+\frac{\alpha n\sqrt{p}}{4}}\sqrt{2\pi}\left(1+o\left(1\right)\right)$$

so that

$$\begin{aligned}\mathcal{I}\left(0,p-\alpha\sqrt{p}\right) &< \left(\frac{n-A_2}{2}\right)^{-\frac{np}{2}}\left(\frac{np}{2}-\frac{\alpha n\sqrt{p}}{4}\right)^{\frac{np}{2}-\frac{1}{2}}e^{-\frac{np}{2}+\frac{\alpha n\sqrt{p}}{4}}\sqrt{2\pi}\left(1+o\left(1\right)\right)\\ &< p^{\frac{np}{2}}e^{-\frac{np}{2}}e^{p\left(\frac{A_2}{2}+\frac{A_2^2}{4n}-\frac{\alpha^2 n}{16p}\right)}\left(1+o\left(1\right)\right),\text{ a.s..}\end{aligned}$$

Comparing this to (51), we see that α can be chosen so that

$$\mathcal{I}\left(0,p-\alpha\sqrt{p}\right)=o(1)\mathcal{I}\left(0,\infty\right), \tag{53}$$

a.s.. Combining (52) and (53), we get (50).

Now, letting $\tilde{\theta}_{pj}=\frac{x}{S_p}\theta_{pj}=\frac{x}{S_p}\frac{1}{2c_p}\frac{h_j}{1+h_j}$, note that there exist $\varepsilon>0$ and $\eta>0$ such that $\left\{2\tilde{\theta}_{pj}:h_j\in[0,\sqrt{c}-\delta]\text{ and }x\in\left[p-\alpha\sqrt{p},p+\alpha\sqrt{p}\right]\right\}\subseteq\Theta_{\varepsilon\eta}$ for all sufficiently large p, a.s.. Hence, by (50), and Proposition 2, a.s.,

$$\begin{aligned}\mathcal{I}\left(0,\infty\right) &= \int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}}x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{p\sum_{j=1}^{r}\left[\tilde{\theta}_{pj}\tilde{v}_{pj}-\frac{1}{2p}\sum_{i=1}^{p}\ln\left(1+2\tilde{\theta}_{pj}\tilde{v}_{pj}-2\tilde{\theta}_{pj}\lambda_{pi}\right)\right]}\\ &\quad\times\left(\prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1-4\left(\tilde{\theta}_{pj}\tilde{v}_{pj}\right)\left(\tilde{\theta}_{ps}\tilde{v}_{ps}\right)c_p}+o(1)\right)\mathrm{d}x,\end{aligned} \tag{54}$$

where $o(1)$ is uniform in $h\in[0,\sqrt{c}-\delta]^r$ and $x\in\left[p-\alpha\sqrt{p},p+\alpha\sqrt{p}\right]$.

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 $\frac{x}{p}-1$, we get

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

where B_0,B_1 and B_2 are $O(1)$ uniformly in $h\in[0,\sqrt{c}-\delta]^r$. Further, consider the integral

$$I^{(0)}=\int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}}x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{p\left(B_1\frac{x}{p}+B_2\left(\frac{x}{p}-1\right)^2\right)}\mathrm{d}x.$$

Splitting the domain of integration into segments $\left[p-\alpha\sqrt{p}, p-\alpha p^{\gamma}\right], \left[p-\alpha p^{\gamma}, p+\alpha p^{\gamma}\right]$ and $\left[p+\alpha p^{\gamma}, p+\alpha\sqrt{p}\right]$, where $0<\gamma<1/2$, and denoting the corresponding integrals by $I^{(1)}, I^{(2)}$ and $I^{(3)}$, respectively, we have

$$
\begin{aligned}
I^{(1)} &< e^{\alpha^2}\int_{p-\alpha\sqrt{p}}^{p-\alpha p^{\gamma}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{B_1 x}\mathrm{d}x < e^{\alpha^2}p^{\frac{np}{2}}\left(1-\frac{2B_1}{n}\right)^{\frac{np}{2}}\int_0^{1-\frac{\alpha}{2}p^{\gamma-1}} y^{\frac{np}{2}-1}e^{-\frac{np}{2}y}\mathrm{d}y,\\
I^{(2)} &> \int_{p-\alpha p^{\gamma}}^{p+\alpha p^{\gamma}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{B_1 x}\mathrm{d}x > p^{\frac{np}{2}}\left(1-\frac{2B_1}{n}\right)^{\frac{np}{2}}\int_{1-\frac{\alpha}{2}p^{\gamma-1}}^{1+\frac{\alpha}{2}p^{\gamma-1}} y^{\frac{np}{2}-1}e^{-\frac{np}{2}y}\mathrm{d}y, \text{ and}\\
I^{(3)} &< e^{\alpha^2}\int_{p+\alpha p^{\gamma}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{B_1 x}\mathrm{d}x < e^{\alpha^2}p^{\frac{np}{2}}\left(1-\frac{2B_1}{n}\right)^{\frac{np}{2}}\int_{1+\frac{\alpha}{2}p^{\gamma-1}}^{\infty} y^{\frac{np}{2}-1}e^{-\frac{np}{2}y}\mathrm{d}y.
\end{aligned}
$$

Using the Laplace approximation, we have

$$
\begin{aligned}
\int_0^{1-\frac{\alpha}{2}p^{\gamma-1}} y^{\frac{np}{2}-1}e^{-\frac{np}{2}y}\mathrm{d}y &= o(1)\int_{1-\frac{\alpha}{2}p^{\gamma-1}}^{1+\frac{\alpha}{2}p^{\gamma-1}} y^{\frac{np}{2}-1}e^{-\frac{np}{2}y}\mathrm{d}y, \text{ and}\\
\int_{1+\frac{\alpha}{2}p^{\gamma-1}}^{\infty} y^{\frac{np}{2}-1}e^{-\frac{np}{2}y}\mathrm{d}y &= o(1)\int_{1-\frac{\alpha}{2}p^{\gamma-1}}^{1+\frac{\alpha}{2}p^{\gamma-1}} y^{\frac{np}{2}-1}e^{-\frac{np}{2}y}\mathrm{d}y,
\end{aligned}
$$

so that $I^{(2)}$ dominates $I^{(1)}$ and $I^{(3)}$ and

$$
\begin{aligned}
I^{(0)} &= (1+o(1))\int_{p-\alpha p^{\gamma}}^{p+\alpha p^{\gamma}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{p\left(B_1\frac{x}{p}+B_2\left(\frac{x}{p}-1\right)^2\right)}\mathrm{d}x\\
&= (1+o(1))\int_{p-\alpha p^{\gamma}}^{p+\alpha p^{\gamma}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{B_1 x}\mathrm{d}x\\
&= (1+o(1))\int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{B_1 x}\mathrm{d}x.
\end{aligned}
$$

This implies that

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

and hence, only 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 $\frac{x}{p}-1$ matter for the evaluation of $\mathcal{I}(0,\infty)$. Let us find these terms.

By Corollary 1, $\frac{x}{S_p} - 1 = \frac{x}{p} - \frac{S_p}{p} + o(p^{-1})$ a.s.. Using this fact, after some algebra, we get

$$\tilde{\theta}_{pj}\widetilde{v}_{pj} = \theta_{pj}v_{pj} + \theta_{pj}v_{pj}^2\left(\frac{x}{p} - \frac{S_p}{p}\right) + O\left(\left(\frac{x}{p}-1\right)^2\right),$$

$$\ln\left(2\tilde{\theta}_{pj}\right) = \ln\left(2\theta_{pj}\right) + \left(\frac{x}{p} - \frac{S_p}{p}\right) + O\left(\left(\frac{x}{p}-1\right)^2\right),$$

and

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

It follows that

$$\begin{aligned}\mathcal{I}(0,\infty) &= \int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{p\sum_{j=1}^{r}\left[\theta_{pj}v_{pj}-\frac{1}{2p}\sum_{i=1}^{p}\ln(1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_{pi})\right]} \\ &\quad\times e^{\sum_{j=1}^{r}\theta_{pj}v_{pj}(x-S_p)}\left(\prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1-4\left(\theta_{pj}v_{pj}\right)\left(\theta_{ps}v_{ps}\right)c_p}+o(1)\right)\mathrm{d}x \\ &= (1+o(1))\prod_{j=1}^{r}\left(1+h_j\right)^{\frac{np}{2}}L_p(h;\lambda_p)\int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{\sum_{j=1}^{r}\theta_{pj}v_{pj}(x-S_p)}\mathrm{d}x,\end{aligned} \tag{55}$$

where the last equality in (55) follows from (3) and Proposition 2.

The last equality in (55), (4) and the fact that

$$\int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1}e^{-\frac{n}{2}x}e^{\sum_{j=1}^{r}\theta_{pj}v_{pj}(x-S_p)}\mathrm{d}x = e^{\sum_{j=1}^{r}-\frac{h_j}{2c_p}S_p}\left(\frac{n}{2}-\sum_{j=1}^{r}\frac{h_j}{2c_p}\right)^{-\frac{np}{2}}\Gamma\left(\frac{np}{2}\right)(1+o(1))$$

imply that

$$\begin{aligned}L_p\left(h;\mu_p\right) &= (1+o(1))\,L_p(h;\lambda_p)e^{\sum_{j=1}^{r}-\frac{h_j}{2c_p}S_p}\left(1-\sum_{j=1}^{r}\frac{h_j}{nc_p}\right)^{-\frac{np}{2}} \\ &= (1+o(1))\,L_p(h;\lambda_p)e^{-\frac{S_p-p}{2c_p}\sum_{j=1}^{r}h_j+\frac{1}{4c_p}\left(\sum_{j=1}^{r}h_j\right)^2},\end{aligned}$$

which establishes (11). The rest of the statements of Theorem 1 follow from (10), (11), and Lemmas 12 and A2 of OMH.□

5.3 Proof of Proposition 4

To save space, we only derive the asymptotic power envelope for the relatively more difficult case of real-valued data and μ-based tests. According to the Neyman-Pearson lemma, the most powerful test of $h = 0$ against the simple alternative $h = (h_1, ..., h_r)$ is the test which rejects the null when $L_p\big(h; \mu_p\big)$ is larger than a critical value C. It follows from Theorem 1 that, for such a test to have asymptotic size α, C must be

$$C = \sqrt{W(h)}\Phi^{-1}(1-\alpha) + m(h), \tag{56}$$

where

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

Now, according to Le Cam's third lemma and Theorem 1, under $h = (h_1, ..., h_r)$, $\ln L_p\big(h; \mu_p\big) \xrightarrow{d} N(m(h) + W(h), W(h))$. The asymptotic power (15) follows.$\square$

5.4 Invariance issues and Proof of Proposition 5

Before turning to the proof of Proposition 5, let us clarify the invariance issues in the problem under study. For basic definitions (invariant, maximal invariant, etc.), we refer to Chapter 6 of Lehmann and Romano (2005).

Suppose that X is a $p \times n$ random matrix with $\text{vec}(X) \sim N(0, I_n \otimes \Sigma)$. This model is clearly invariant under the group $\mathcal{G}_p$, acting on $\mathbb{R}^{p\times n}$, of left-hand multiplications by a $p \times p$ orthogonal matrix $x \mapsto Qx$, $x \in \mathbb{R}^{p\times n}$, $Q \in \mathcal{O}(p)$; so are the null hypothesis H_0 and the alternative H_1. Letting $m = \min(n, p)$, the m-tuple $\lambda(X) = (\lambda_1, ..., \lambda_m)$ of non-zero eigenvalues of $\frac{1}{n}XX'$ is clearly invariant under that group, since $\frac{1}{n}xx'$ and $\frac{1}{n}(Qx)(Qx)' = \frac{1}{n}Qxx'Q'$ share the same eigenvalues $\lambda(x)$ for any orthogonal matrix Q and any matrix $x \in \mathbb{R}^{p\times n}$. However, $\lambda(X)$ is not maximal invariant for $\mathcal{G}_p$, as xx' and $(xP)(xP)' = xPP'x' = xx'$, where P is an arbitrary $n \times n$ orthogonal matrix, share the same $\lambda(x) = \lambda(xP)$ although, in general, there is no $p \times p$ orthogonal matrix Q such that $xP = Qx$.

Now, the joint density of the elements of X is

$$f_{\Sigma}^{(n)}(x) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left(\Sigma^{-1}xx'\right)\right\}, \ x \in \mathbb{R}^{p\times n}.$$

By the factorization theorem, XX' is a sufficient statistic, and it is legitimate to restrict attention to XX'-measurable inference procedures. Left-hand orthogonal multiplications Qx of x yields, for xx', a transformation of the form $Qxx'Q'$. When Q range over the family $\mathcal{O}_p$ of $p \times p$ orthogonal matrices, those transformations also form a group, $\widetilde{\mathcal{G}}_p$, say, now acting on the space of $p \times p$ symmetric positive semidefinite real matrices of rank m. Clearly, $\lambda(x)$ is maximal invariant for $\widetilde{\mathcal{G}}_p$, as xx' and yy' share the same eigenvalues if and only if $yy' = Qxx'Q'$ for some $p \times p$ orthogonal matrix Q.

Combining the principles of sufficiency and invariance thus leads to considering λ-measurable tests only.

A similar reasoning applies in the case of unspecified σ^2, with a larger group combining multiplication by an arbitrary non-zero constant with the $p \times p$ left orthogonal transformations. Sufficiency and invariance then lead to restricting attention to μ-measurable tests.

Proof of Proposition 5.

With the same notation as above, write $T = T(X) = XX'$ for the sufficient statistic. Consider an arbitrary invariant (under the group $\mathcal{G}_p$ of left orthogonal transformations of $\mathbb{R}^{p\times n}$) test $\phi(X)$, and define $\psi(t) = E(\phi(X)|T = t)$. Then $\psi(T)$ is a T-measurable test with the same size and power function as $\phi(X)$. It follows from the proof of Theorem 6.5.3 (i) in Lehmann and Romano (2005) that

$\psi(T)$ is *almost invariant*. Moreover, since the conditions of Lemma 6.5.1 (same reference) hold, this test is invariant under the group $\widetilde{\mathcal{G}}_p$ (acting on T). Since the ordered m-tuple $\lambda_1, ..., \lambda_m$ of the eigenvalues of $\frac{1}{n}T = \frac{1}{n}XX'$ is maximal invariant for $\widetilde{\mathcal{G}}_p$, and since any invariant statistic is a measurable function of a maximal invariant one, $\psi(T)$ must be λ-measurable. Hence, $\psi(T)$ is a λ-measurable test and has the same power function as $\phi(X)$, as was to be shown.

The existence of a μ-measurable test with the same power function as that of a test $\phi(X)$ invariant under left orthogonal transformations and multiplication by non-zero constants is established similarly.□

References

[1] Bai, Z.D. and J.W. Silverstein (2004) "CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices", Annals of Probability 32, 553-605.

[2] Bai, Z.D., Jiang, D., Yao, J.F., and S. Zheng (2009) "Corrections to LRT on Large-dimensional Covariance Matrix by RMT", Annals of Statistics 37, 3822-3840.

[3] Baik, J., Ben Arous, G. and S. Peche. (2005) "Phase transition of the largest eigenvalue for non-null complex sample covariance matrices" Annals of Probability 33, 1643–1697.

[4] Baik, J. and J.W. Silverstein. (2006) "Eigenvalues of large sample covariance matrices of spiked population models". Journal of Multivariate Analysis 97, 1382–1408.

[5] Berthet, Q. and Rigollet, P. (2012). Optimal detection of sparse principal components in high dimension. arXiv:1202.5070

[6] Cai, T.T. and Z. Ma (2012) "Optimal Hypothesis Testing for High Dimensional Covariance Matrices", arXiv: 1205.4219

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

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

[9] Goodman, N. R. (1963) "Statistical analysis based on a certain multivariate complex Gaussian distribution, (An introduction)." Annals of Mathematical Statistics 34, 152-177.

[10] Harish-Chandra (1957) "Differential Operators on Semi-simple Lie Algebra", *American journal of Mathematics* 79, 87-120.

[11] Itzykson, C., and Zuber, J.B. (1980) "The Planar Approximation. II", *Journal of Mathematical Physics* 21, 411-421.

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

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

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

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

[16] Klinger, A. (1967) "The Vandermonde Matrix", *The American Mathematical Monthly* 74, 571-574.

[17] Ledoit, O. and M. Wolf (2002) "Some Hypothesis Tests for the Covariance Matrix when the Dimension is Large Compared to the Sample Size", Annals of Statistics 30, 1081-1102.

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

[19] Macdonald, I. G. (1995) Symmetric functions and Hall polynomials. Second edition. Oxford Mathematical Monographs. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York.

[20] Mo, M.Y. (2011) "The rank 1 real Wishart spiked model", arXiv:1101.5144v1

[21] Olver, F.W.J. (1997) Asymptotics and Special Functions, A K Peters, Natick, Massachusetts.

[22] Onatski, A. (2008) "The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices", *Annals of Applied Probability* 18, No. 2, 470-490

[23] Onatski, A. (2009) "Testing Hypotheses About the Number of Factors in Large Factor Models", Econometrica 77, 1447-1479.

[24] Onatski, A., Moreira, M. J., and Hallin, M. (2011) "Asymptotic Power of Sphericity Tests for High-dimensional Data", manuscript, Economics Faculty, University of Cambridge.

[25] Onatski, A. (2012) "Detection of Weak Signals in High-dimensional Complex Valued Data", arXiv:1207.7098.

[26] Ratnarajah, T. and R. Vaillancourt (2005) “Complex Singular Wishart Matrices and Applications”, Computers & Mathematics with Applications 50, 399–411.

[27] 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.

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

[29] Tulino, A. and Verdú, S. (2004). Random Matrix Theory and Wireless Communications. Foundations and Trends in Communications and Information Theory 1. Now Publishers, Hanover, MA.

[30] Uhlig, H. (1994) “On Singular Wishart and Singular Multivariate Beta Distributions”, Annals of Statistics 22, 395-405.

[31] van der Vaart, A.W. (1998) Asymptotic Statistics, Cambridge University Press.

[32] Wang, D. (2010) “The largest eigenvalue of real symmetric, Hermitian and hermitian self-dual random matrix models with rank one external source, part I.” arXiv:1012.4144