Testing general linear hypotheses under a high-dimensional multivariate regression model with spiked noise covariance

Haoran Li*[1], Alexander Aue†[2], Debashis Paul‡[2], and Jie Peng§[2]

[1]Auburn University
[2]University of California at Davis

Abstract

We consider the problem of testing linear hypotheses under a multivariate regression model with a high-dimensional response and spiked noise covariance. The proposed family of tests consists of test statistics based on a weighted sum of projections of the data onto the estimated latent factor directions, with the weights acting as the regularization parameters. We establish asymptotic normality of the test statistics under the null hypothesis. We also establish the power characteristics of the tests and propose a data-driven choice of the regularization parameters under a family of local alternatives. The performance of the proposed tests is evaluated through a simulation study. Finally, the proposed tests are applied to the *Human Connectome Project* data to test for the presence of associations between volumetric measurements of human brain and behavioral variables.

Keywords: Factor models, Regularized tests, Local alternatives, Random matrix theory.

*Li is Assistant Professor, Department of Mathematics and Statistics, Auburn University, 211 Parker Hall, Auburn University, Auburn, AL, 36849. Email: hzl0152@auburn.edu.

†Aue is Professor, Department of Statistics, University of California at Davis, One Shields Avenue, Davis, CA 95616. Email: aaue@ucdavis.edu.

‡Paul is Professor, Department of Statistics, University of California at Davis, One Shields Avenue, Davis, CA 95616. Email: debpaul@ucdavis.edu. Paul was partially supported by NSF grant DMS-1915894 and DMS-1713120.

§Peng is Professor, Department of Statistics, University of California at Davis, One Shields Avenue, Davis, CA 95616. Email: jiepeng@ucdavis.edu. Peng was partially supported by NSF grant DMS-1915894.

1 Introduction

Large dimensional factor models are ubiquitous in econometrics and various branches of science. See, for example, Bai and Ng (2002); Bai (2003); Pesaran (2006); Price et al. (2006); Abraham and Inouye (2014); Zheng et al. (2012); Viviani et al. (2005); Andersen et al. (1999). In many applications, we are able to observe a group of explanatory variables that can be treated as observed factors, while the remaining factors are latent or unobserved. See for example, Fama and French (1992, 1993). In this paper, we are interested in testing for effects of observed factors in the presence of latent factors. Specifically, we consider a latent factor linear regression model with high-dimensional responses. We are interested in testing linear hypotheses on the regression coefficients of the observed factors (predictors). The problem is of statistical importance and can be used for many purposes, for example, testing significance of predictors, testing of linear trends or seasonal cycles, two-sample tests, MANOVA and others. We assume that the p-dimensional responses $\mathbf{y}_i$'s can be expressed as

$$\mathbf{y}_i = \mathbf{B}\mathbf{x}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, N, \quad \text{where} \tag{1.1}$$

$$\boldsymbol{\varepsilon}_i = \mathbf{D}\mathbf{f}_i + \sigma\mathbf{e}_i, \quad i = 1, \ldots, N. \tag{1.2}$$

We make the following structural assumptions:

(a) $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ is an $m \times N$ known deterministic predictor matrix of rank m ($< N$);

(b) $\mathbf{B}$ is a $p \times m$ unknown parameter matrix;

(c) $\mathbf{D}$ is a $p \times K$ unobserved matrix of *loadings* of rank K ($< p$);

(d) $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_N]$ is a $K \times N$ (unobserved) random matrix of independent latent *factors*;

(e) $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_N]$ is a $p \times N$ (unobserved) random matrix of i.i.d. entries with mean 0 and variance 1;

(f) $\mathbf{F}$ and $\mathbf{E}$ are independent and normally distributed;

(g) $\mathbb{E}\mathbf{f}_i = \mathbf{0}$, $\mathrm{Cov}(\mathbf{f}_i) = I_K$ and the matrix $\mathbf{D}^T\mathbf{D}$ is diagonal with distinct diagonal elements. The condition is to ensure the identifiability of $\mathbf{F}$ and $\mathbf{D}$ (see Anderson (1958)).

We are interested in testing general linear hypotheses under (1.1)–(1.2) of the form

$$H_0 \colon \mathbf{BC} = \mathbf{0} \quad \text{against} \quad H_a \colon \mathbf{BC} \neq \mathbf{0}. \tag{1.3}$$

Here, $\mathbf{C}$ is an arbitrary $m \times q$ pre-specified matrix of fixed dimensions ($q \leq m$), subject to the requirement that $\mathbf{BC}$ is estimable. Without loss of generality, $\mathbf{C}$ is taken to be of rank q. With appropriate specifications of $\mathbf{X}$ and $\mathbf{C}$, the above testing formulation encapsulates a broad range of inferential questions, including the ones mentioned above (1.1). For example, the two-sample test corresponds to the case when the $\mathbf{y}_i$'s are sampled from two populations and $\mathbf{B}$ consists of the corresponding population means, while the $\mathbf{x}_i$'s are group membership indicators. Of interest is to test whether the means are equal, which can be formulated as (1.3) by selecting $\mathbf{C} = [1, -1]^T$, $m = 2$, $q = 1$.

The $p \times 1$ vectors $\boldsymbol{\varepsilon}_i = \mathbf{D}\mathbf{f}_i + \sigma\mathbf{e}_i$ represent the observational noise and has zero mean and covariance $\Sigma_p = \mathbf{D}\mathbf{D}^T + \sigma^2 I_p$. Notice that the noise covariance Σ_p has the so called "spiked eigenvalue" structure under which the *spectrum*, i.e., the ordered sequence of eigenvalues of Σ_p, is of the form $(\ell_1, \ldots, \ell_K, \sigma^2, \cdots, \sigma^2)$ where $\ell_1 \geq \cdots \geq \ell_K$ are the K *spikes* and $\sigma^2 (< \ell_K)$ represents the variance of background (idiosyncratic) noise. Henceforth, we denote by $\mathbf{h}_j$ the eigenvector associated with ℓ_j. Throughout, we assume m and K to be fixed even as p and N diverge to infinity. Moreover, denote by $n = N - m$ the effective sample size. The

spiked covariance model has been extensively studied in the context of high-dimensional principal components analysis (PCA) (Johnstone, 2001; Baik and Silverstein, 2006; Paul, 2007; Onatski, 2012; Paul and Aue, 2014; Johnstone and Paul, 2018).

In classical multivariate analysis, such problems are typically addressed within the framework of likelihood ratio tests (LRT) (see Section 2.1). The LRT test statistic is computed based on the maximum likelihood estimates of Σ_p under the null and alternative hypotheses, given that $p < n$. The LRT performs well when n is much larger than p. However, when p is comparable to (but smaller than) n, the MLEs are inconsistent causing the performance of the LRT to deteriorate. Moreover, the computation of the null distribution of the LRT requires complex analytic approximations that are feasible only in moderate dimensions (He et al., 2021).

To the best of our knowledge, the question of linear hypothesis testing in such high-dimensional latent factor regression models is relatively unexplored in the statistical literature. The main focus of this paper is to demonstrate that significant performance enhancements over LRT can be made by implementing appropriate regularization that takes into account both the spiked covariance structure and the dimensionality of the response. The major goal here is to address the testing problem involving linear functions of the regression coefficients when the dimension p of the response and the number of observations N are of comparable size, so that $p/N \to \gamma \in (0, \infty)$ as $N \to \infty$.

To position our work in the literature, the question of testing general linear hypothesis has been explored when in (1.1), the response has an arbitrary covariance matrix Σ_p. In the special case of two-sample tests, in a seminal work, Bai and Saranadasa (1996) proposed to replace $\hat{\mathbf{\Sigma}}_{\text{Full}}$ (unrestricted MLE of Σ_p; see Section 2.1 for details) with the identity matrix I_p. Significant extensions were proposed by Chen and Qin (2010). In the past few years,

the general linear hypothesis problem has also been studied. For example, Bai et al. (2013) corrected the scaling of the LRT when m (rank of $\mathbf{X}$) and q (number of hypotheses) are proportional to p (dimension of the responses). Li et al. (2020a) and Li et al. (2020b) addressed the problem by applying spectral shrinkage to $\hat{\mathbf{\Sigma}}_{\text{Full}}$. He et al. (2021) proposed to reduce the dimensionality of the response and then apply a corrected likelihood ratio test on the reduced observations.

When the covariance of the response Σ_p has spiked eigenvalues, to the best of our knowledge, the linear hypothesis problem is only studied in the literature for the special case of two-sample tests. Aoshima and Yata (2018) suggested applying the test in Bai and Saranadasa (1996) when the spikes are mild ($\ell_1 = o(\sqrt{p})$). When the spikes are strong ($\liminf_{p\to\infty} \ell_1/\sqrt{p} > 0$), they proposed a test by projecting the responses onto the estimated eigensubspace associated with the idiosyncratic noise. A similar approach was taken by Wang and Xu (2018).

The testing problem (1.3) has not been studied in its fully generality when Σ_p has spiked eigenvalues. In particular, the presence of such structures opens up the possibility of adopting regularization schemes to mitigate the effects of high dimension, which is the central topic here. The main contribution of this paper is to propose a class of rotationally invariant regularized tests, introduced in Section 2, for the general linear hypothesis (1.3) under model (1.1)–(1.2), when the response dimension p and sample size N are comparable. We then investigate the power characteristics of the tests under a class of local alternatives in Section 3 and choose the regularization parameter using decision-theoretic principles with the aim of maximizing the local power in Section 4.

Note that the proposed tests, like most of the well-known classical tests for linear hypotheses including the likelihood ratio test, are rotationally invariant. Indeed, invariance

with respect to a group action as a guiding principle for designing statistical tests has a long and successful history. For a comprehensive discussion see Eaton and George (2021). A particularly desirable feature of rotationally invariant tests is that they do not depend on the coordinate system. This is in contrast with some other principles such as sparsity. Under the spiked covariance model, it also implicitly conducts dimension reduction by reducing the analytical properties of the test to that of a lower dimensional object, namely the spiked eigenvalues, the variance associated with the background noise, and certain projections of the spiked eigenvectors. Moreover, when there is a large-number of coordinates with small-sized signals, the proposed tests have the benefit of combining information across coordinates.

The finite sample properties of the proposed test are discussed in Section 5, where we report empirical performance of the tests through a simulation study, and in Section 6, where we apply the proposed test to a *Human Connectome Project* data set. In Section 7, we discuss the possibility of generalizations to non-Gaussian data. Additional details of the simulation study and proofs of the main results are deferred to a *Supplementary Material*.

2 Testing general linear hypothesis

In this section, we first present a classical ANOVA decomposition of the total variability in the response under the unrestricted model when the observation error has arbitrary covariance that forms the backbone of classical inference procedures including the likelihood ratio test. Next, we propose a family of regularized tests that is built upon the structure of the classical testing procedure, but uses alternative scaling strategies for measuring the departure from the null hypothesis. This proposal also takes into account the factor model structure (1.2). Throughout, we assume that K, the number of latent factors, is known.

Data-driven selections of K are discussed in Section S.3 of the Supplementary Material.

Define the residual covariance matrix under the full model and the *hypothesis sums of squares and cross-products matrix*, respectively, as

$$\mathbf{S}_{\text{Full}} = \frac{1}{n}\mathbf{Y}(I_N - \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X})\mathbf{Y}^T \quad \text{and} \quad \mathbf{S}_{\text{H}} = \frac{1}{n}\mathbf{Y}Q_N Q_N^T\mathbf{Y}^T, \tag{2.1}$$

where $Q_N = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{C}[\mathbf{C}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{C}]^{-1/2}$. Note that $Q_N Q_N^T$ is the projection matrix onto the space of the null hypothesis and $\mathbf{S}_{\text{H}}$ quantifies the degree of departure of the observations from the null hypothesis. The residual covariance matrix under the reduced model $H_0\colon \mathbf{BC} = 0$ can be defined following the additive relationship $\mathbf{S}_{\text{Red}} = \mathbf{S}_{\text{Full}} + \mathbf{S}_{\text{H}}$. Note that $(n/N)\mathbf{S}_{\text{Full}}$ and $(n/N)\mathbf{S}_{\text{Red}}$ are, respectively, MLEs of Σ_p under the full model and the reduced model when Σ_p is unrestricted and allowed to be a general nonnegative definite matrix. Moreover, $T_0 = \log[\det(\mathbf{S}_{\text{Red}}\mathbf{S}_{\text{Full}}^{-1})]$ is the log-likelihood ratio test statistic, and T_0 has an asymptotic χ^2 limit under the null hypothesis when $n \gg p$ (see Fujikoshi (2016) for details). The matrix $\mathbf{S}_{\text{Red}}\mathbf{S}_{\text{Full}}^{-1}$ can be seen as a *generalized F-ratio statistic* measuring the relative magnitude of the "regression sum-of-squares-and-product" matrix $\mathbf{S}_{\text{Red}}$ and the "within sample variance" matrix $\mathbf{S}_{\text{Full}}$.

Classical methods such as the likelihood ratio test described above perform well when n is much larger than p. However, when p is comparable to n, the MLE of Σ_p is inconsistent even under the correct factor structure. Consequently, the performance of the LRT for unrestricted Σ_p and for factor models (see Section 2.1) suffer from poor power properties.

2.1 LRT under factor model

In this subsection, we describe a version of the likelihood ratio test (LRT) that takes into account the structure of Σ_p imposed by (1.2). Let $\tau_1 \geq \cdots \geq \tau_p \geq 0$ be the ordered

eigenvalues of $\mathbf{S}_{\text{Full}}$ and $\mathbf{p}_j$'s be the associated eigenvectors. Further, let $\alpha_1 \geq \cdots \geq \alpha_p \geq 0$ be the ordered eigenvalues of $\mathbf{S}_{\text{Red}}$ and $\mathbf{q}_j$'s be the associated eigenvectors. Define

$$\widetilde{\sigma}^2_{\text{Full}} = \frac{1}{p-K} \sum_{j=K+1}^{p} \tau_j,$$
$$\widetilde{\sigma}^2_{\text{Red}} = \frac{1}{p-K} \sum_{j=K+1}^{p} \alpha_j.$$

Note that, after the scaling of (n/N), τ_j's, $\mathbf{p}_j$'s and $\widetilde{\sigma}^2_{\text{Full}}$, respectively, are the MLEs of ℓ_j's, $\mathbf{h}_j$'s and σ^2 under the full model. Correspondingly, after the scaling of (n/N), α_j's, $\mathbf{q}_j$'s and $\widetilde{\sigma}^2_{\text{Red}}$ are the MLEs under the reduced model. Further, define

$$\hat{\mathbf{\Sigma}}_{\text{Full}} = \sum_{j=1}^{K} \tau_j \mathbf{p}_j \mathbf{p}_j^T + \widetilde{\sigma}^2_{\text{Full}} \sum_{j=K+1}^{p} \mathbf{p}_j \mathbf{p}_j^T, \tag{2.2}$$
$$\hat{\mathbf{\Sigma}}_{\text{Red}} = \sum_{j=1}^{K} \alpha_j \mathbf{q}_j \mathbf{q}_j^T + \widetilde{\sigma}^2_{\text{Red}} \sum_{j=K+1}^{p} \mathbf{q}_j \mathbf{q}_j^T. \tag{2.3}$$

Then, the MLE of Σ_p are $(n/N)\hat{\mathbf{\Sigma}}_{\text{Full}}$ and $(n/N)\hat{\mathbf{\Sigma}}_{\text{Red}}$, respectively, under the full model and the reduced model $H_0\colon \mathbf{BC} = 0$, when $\Sigma_p = \mathbf{D}\mathbf{D}^T + \sigma^2 I_p$. For more details see Anderson and Rubin (1956). The log-likelihood ratio test statistic for the hypothesis (1.3) is then

$$T_{\text{LRT}} = \log\det(\hat{\mathbf{\Sigma}}_{\text{Red}} \hat{\mathbf{\Sigma}}^{-1}_{\text{Full}}). \tag{2.4}$$

The χ^2-approximation of the LRT, derived under the fixed-dimensional setting, involves rejecting the null hypothesis at asymptotic level α if $NT_{\text{LRT}} > \chi^2_{1-\alpha}(pq)$, where $\chi^2_{1-\alpha}(pq)$ is the $1-\alpha$ upper quantile of the χ^2 distribution with pq degrees of freedom. He et al. (2021) indicate that the χ^2-approximation fails when the dimension p is comparable to the effective sample size $n = N - m$.

2.2 Regularized tests

Our proposal is built upon the observation that the likelihood ratio statistic in the general setting can be expressed as $T_0 = \log\det(I_q + \mathbf{S}_{\mathrm{H}}\mathbf{S}_{\mathrm{Full}}^{-1})$. Note that the classical testing procedures for H_0 are all based on appropriate linear functionals of the eigenvalues of the matrix $\mathbf{S}_{\mathrm{H}}\mathbf{S}_{\mathrm{Full}}^{-1}$. Here, $\mathbf{S}_{\mathrm{H}}$ captures the signal magnitude, i.e., the degree of departure from the null hypothesis, while $\mathbf{S}_{\mathrm{Full}}^{-1}$ is an estimator of Σ_p^{-1}, which sets the scale of individual coordinates. However, in high-dimensional regimes, $\mathbf{S}_{\mathrm{Full}}^{-1}$ either does not exist or is a poor estimator of Σ_p^{-1}. Therefore, we replace $\mathbf{S}_{\mathrm{Full}}^{-1}$ with a regularized version. Specifically, our approach is to exploit the factor model structure (1.2) to come up with a flexible class of "scaling matrices" that replace $\mathbf{S}_{\mathrm{Full}}^{-1}$ in the test statistics with

$$\mathbf{\Omega}(\Lambda) = \sum_{j=1}^{K} \lambda_j \mathbf{p}_j \mathbf{p}_j^T + \lambda_0 I_p, \tag{2.5}$$

where $\mathbf{p}_j$ is the eigenvector corresponding to the j-th largest eigenvalue of $\mathbf{S}_{\mathrm{Full}}$ and $\Lambda = (\lambda_0, \lambda_1, \ldots, \lambda_K) \in \mathbb{R}_+ \times \mathbb{R}^K$ is a vector of regularization parameters. Note that we do not restrict the λ_j's to be nonnegative. Indeed, setting $\lambda_j = -\lambda_0$, $j = 1, \ldots, K$, leads to $\mathbf{\Omega}(\Lambda) = \lambda_0(I_p - \sum_{j=1}^{K} \mathbf{p}_j \mathbf{p}_j^T)$ and the resulting test statistic involves projecting $\mathbf{S}_{\mathrm{H}}$ onto the estimated eigensubspace associated with the idiosyncratic noise. The proposals of Aoshima and Yata (2018) and Wang and Xu (2018) are along this line.

The regularized replacement of $\mathbf{S}_{\mathrm{H}}\mathbf{S}_{\mathrm{Full}}^{-1}$ is $\mathbf{S}_{\mathrm{H}}\mathbf{\Omega}(\Lambda)$. For mathematical convenience, we work with the symmetrized version,

$$\mathbf{M}(\Lambda) = \frac{1}{p} Q_N^T \mathbf{Y}^T \mathbf{\Omega}(\Lambda) \mathbf{Y} Q_N, \tag{2.6}$$

since $\mathbf{M}(\Lambda)$ has the same nonzero eigenvalues as $(p/n)\mathbf{S}_{\mathrm{H}}\mathbf{\Omega}(\Lambda)$ by (2.1). Note that, the

scaling factor $1/p$ in the expression of $\mathbf{M}(\Lambda)$ is to assure the elements of the matrix are neither diverging nor vanishing as p, n goes to infinity simultaneously at the same rate.

We propose three families of statistics

$$T_0^{\mathrm{LR}}(\Lambda) = \log\Big[\det(I_q + \mathbf{M}(\Lambda))\Big], \qquad T_0^{\mathrm{LH}}(\Lambda) = \mathrm{Tr}\Big[\mathbf{M}(\Lambda)\Big],$$
$$T_0^{\mathrm{BNP}}(\Lambda) = \mathrm{Tr}\Big[\mathbf{M}(\Lambda)\{I_q + \mathbf{M}(\Lambda)\}^{-1}\Big],$$

that generalize the classical likelihood ratio (LR), Lawley–Hotelling trace (LH) and Bartlett–Nanda–Pillai's normalized trace (BNP) tests.

The proposed families of tests are *rotation-invariant*, which means if an orthogonal transformation is applied to the observations, the test statistics remain unchanged. It is a desirable property in the absence of other structural constraints such as sparsity. Moreover, even though under a high-dimensional regime the sample eigenvectors p_j's are biased, the proposed tests only involve projection to a lower dimensional space which is a good proxy of its population counterpart as shown in Section 3.

The proposed test statistics are also connected with the LRT (2.4) under the factor model since the matrices $\mathbf{\Omega}(\Lambda)$ have the same eigen-subspace as $\hat{\mathbf{\Sigma}}_{\mathrm{Full}}$. Indeed, $\mathbf{M}(\Lambda)$ is expressible as a weighted sum of the projections of the signal-bearing component of the response $\mathbf{Y}$ (in reference to H_0) onto the eigensubspaces of $\hat{\mathbf{\Sigma}}_{\mathrm{Full}}$, while the regularization parameters Λ act as weights.

3 Asymptotic analysis

In this section, we first introduce some results in the *Random Matrix Theory* literature in Section 3.1 associated with a spiked covariance model. The asymptotic null distribution

of the proposed tests is derived in Section 3.2. In Section 3.3, we compute the asymptotic power of the proposed tests under a class of local alternatives. In Section 3.4, we derive empirically normalized tests and determine their critical values at any significance level.

Besides Assumptions (a)–(g) stated in Section 1, the following assumptions are made.

A1 Asymptotic regime: $n, p \to \infty$ simultaneously such that $\gamma_n = p/n \to \gamma \in (0, \infty)$ and $\sqrt{n}|\gamma_n - \gamma| \to 0$.

A2 Significant spike: The number of spikes K is fixed, and $\ell_1 > \cdots > \ell_K$ are fixed as $n, p \to \infty$, and all spikes are significant in the sense that $\ell_K > \sigma^2(1 + \sqrt{\gamma})$.

A3 Asymptotically full rank: $\mathbf{X}$ is of full rank and $n^{-1}\mathbf{X}\mathbf{X}^T$ converges to a positive definite $m \times m$ matrix.

A4 Asymptotically estimable: $\liminf_{n\to\infty} \rho_{\min}(\mathbf{C}^T(n^{-1}\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{C}) > 0$, where $\rho_{\min}(\cdot)$ is the smallest eigenvalue of a symmetric matrix.

Condition **A1** indicates that the dimensions of the response and sample size grow at the same rate. The $o(n^{-1/2})$ convergence rate of γ_n to γ is a technical condition needed to ensure the distributional convergence of the appropriately normalized test statistics. Condition **A2** guarantees that the spiked eigenvalues are detectable from the observed data. Condition **A3** is used to eliminate any redundancy in the specification of the model. Condition **A4** is to ensure that the parameter of interest $\mathbf{BC}$ is estimable.

3.1 Asymptotics of eigenvalues and eigenvectors of $\hat{\mathbf{\Sigma}}_{\text{Full}}$

In this subsection, we present an important result on the behavior of the eigenvalues and eigenvectors of the MLE $\hat{\mathbf{\Sigma}}_{\text{Full}}$ of the covariance matrix under the spiked model (1.1)–(1.2). Since the eigensubspaces of $\mathbf{\Omega}(\Lambda)$ and $\hat{\mathbf{\Sigma}}_{\text{Full}}$ are the same, these results have a direct bearing

on the asymptotic behavior of the proposed tests. First we define two auxiliary functions

$$\psi(x,\gamma) = x + \frac{\gamma x}{x-1},\ x \in (1,\infty); \quad \zeta(x,\gamma) = \Big[\frac{1-\gamma/(x-1)^2}{1+\gamma/(x-1)}\Big]^{1/2},\ x \in [1+\sqrt{\gamma},\ \infty). \tag{3.1}$$

The following theorem combines results in Paul (2007), Onatski (2012) and Passemier et al. (2017). Let $\xrightarrow{a.s.}$ denote almost sure convergence and $\Longrightarrow$ weak convergence.

Theorem 3.1 *Suppose that Model* (1.1)– (1.2)*, Conditions (a)–(g) and* ***A1****–****A3*** *hold.*

(1) The leading sample eigenvalues are biased upwards as follows:

$$\tau_j \xrightarrow{a.s.} \sigma^2\psi(\ell_j/\sigma^2,\gamma) = \ell_j + \frac{\gamma\ell_j}{\ell_j/\sigma^2 - 1}\,, \quad j = 1,\dots,K. \tag{3.2}$$

(2) The angle between a sample eigenvector and a population eigenvector is such that

$$|\langle \mathbf{p}_j, \mathbf{h}_i\rangle| = |\mathbf{p}_j^T\mathbf{h}_i| \xrightarrow{a.s.} \mathbb{1}_0(j-i)\zeta(\ell_j/\sigma^2,\gamma), \qquad 1 \le i,j \le K,$$

where $\mathbb{1}_0(x) = 1$ *if* $x = 0$ *and* $\mathbb{1}_0(x) = 0$ *otherwise.*

(3) The following decomposition of $\mathbf{p}_j$ *holds. Recall* $\mathbf{H}_K = [\mathbf{h}_1, \cdots, \mathbf{h}_K]$*. Denote* $\mathbf{H}_\perp$ *to be the orthogonal complement of* $\mathbf{H}_K$*. We have*

$$\mathbf{p}_j = w_j\mathbf{h}_j + \frac{1}{\sqrt{n}}w_j\mathbf{H}_K\mathbf{v}_j + \sqrt{1-\tilde{w}_j^2}\ \mathbf{H}_\perp\mathbf{u}_j, \quad j = 1,\dots,K, \quad \textit{where} \tag{3.3}$$

(i) $\mathbf{v}_j$ *is a* K*-variate random vector such that*

$$\mathbf{v}_j \Longrightarrow \mathcal{N}(0,\Sigma_j(\ell_j)), \quad \textit{where}\ \ \Sigma_j(\ell_j) = \frac{(\ell_j/\sigma^2-1)^2}{(\ell_j/\sigma^2-1)^2-\gamma}\sum_{1\le i\ne j\le K}\frac{\ell_i\ell_j}{(\ell_i-\ell_j)^2}\mathbf{o}_i\mathbf{o}_i^T.$$

Here, $\mathbf{o}_i$ *is the* i*th canonical vector of dimension* K *with 1 in the* i*th coordinate and*

0 *elsewhere.*

(ii) $w_j = \zeta(\ell_j/\sigma^2, \gamma_n) + O_p(n^{-1/2})$.

(iii) $\tilde{w}_j$ *is such that*

$$\tilde{w}_j^2 = w_j^2(1 + \frac{2}{\sqrt{n}} v_{jj} + \frac{1}{n}\|\mathbf{v}_j\|^2),$$

where v_{jj} *is the* j*-th element of* $\mathbf{v}_j$*. Without loss of generality, we can set* $\tilde{w}_j \geq 0$.

$$\tilde{w}_j = w_j + O_p(n^{-1/2}) = \zeta(\ell_j/\sigma^2, \gamma_n) + O_p(n^{-1/2}).$$

(iv) $\mathbf{u}_j$ *is a* $(p-K)$*-variate random vector following the uniform distribution on the unit* $(p-K-1)$*-sphere.*

(4) The MLE $\widetilde{\sigma}^2_{\text{Full}}$ *of noise variance is biased downwards as*

$$\frac{p-K}{\sigma^2\sqrt{2\gamma}}\Big(\widetilde{\sigma}^2_{\text{Full}} - \sigma^2\Big) + b(\sigma^2) \Longrightarrow \mathcal{N}(0,1), \quad b(\sigma^2) = \sqrt{(\gamma/2)}\Big\{K + \sum_{j=1}^{K} (\ell_j/\sigma^2 - 1)^{-1}\Big\}.$$

Theorem 3.1 shows that τ_j and $\mathbf{p}_j$ are biased systematically and $\widetilde{\sigma}^2_{\text{Full}}$ underestimates σ^2. Furthermore, (3.3) gives a characterization of the bias in the sample eigenvector $\mathbf{p}_j$ as an estimator of $\mathbf{h}_j$. Here, we assume without loss of generality that $\langle \mathbf{p}_j, \mathbf{h}_j \rangle > 0$.

3.2 Asymptotic null distribution

In this subsection, we first present a useful representation of the matrix $\mathbf{M}(\Lambda)$ under the null hypothesis for any fixed Λ. Denote $L = \text{diag}(\ell_1, \ldots, \ell_K)$ and $\Lambda_+ = \text{diag}(\lambda_1, \ldots, \lambda_K)$. Further, define $\zeta_j = \zeta(\ell_j/\sigma^2, \gamma_n)$ with $\gamma_n = p/n$, and $\Psi_1 = \text{diag}(\zeta_1, \ldots, \zeta_K)$, $\Psi_2 = \text{diag}\left(\sqrt{1-\zeta_1^2}, \ldots, \sqrt{1-\zeta_K^2}\right)$.

Theorem 3.2 (Representation of $\mathbf{M}(\Lambda)$) *Suppose that Model* (1.1)–(1.2) *,Conditions (a)–(g) and* ***A1****–****A4*** *hold. Then, for any fixed* Λ*, under* H_0: $\mathbf{BC} = \mathbf{0}$*,*

$$\mathbf{M}(\Lambda) = \mathbf{M}_1(\Lambda) + \mathbf{M}_2(\Lambda) + O_p(n^{-3/2}), \text{ where} \tag{3.4}$$

$$\begin{aligned}\mathbf{M}_1(\Lambda) = {} & \frac{1}{p}\mathbf{V}_1^T(\lambda_0 I_K + \Lambda_+\Psi_1^2)L\mathbf{V}_1 + \frac{\sigma^2}{p}\mathbf{V}_2^T(\lambda_0 I_K + \Lambda_+\Psi_2^2)\mathbf{V}_2 \\ & + \frac{\sigma}{p}\mathbf{V}_1^T\Lambda_+\Psi_1\Psi_2 L^{1/2}\mathbf{V}_2 + \frac{\sigma}{p}\Big(\mathbf{V}_1^T\Lambda_+\Psi_1\Psi_2 L^{1/2}\mathbf{V}_2\Big)^T,\end{aligned} \tag{3.5}$$

$$\mathbf{M}_2(\Lambda) = \frac{\sigma^2\lambda_0}{p}\mathbf{V}_3^T\mathbf{V}_3. \tag{3.6}$$

Here, $\mathbf{V}_1$*,* $\mathbf{V}_2$ *and* $\mathbf{V}_3$ *are, respectively,* $K\times q$*,* $K\times q$ *and* $(p-2K)\times q$ *independent matrices with i.i.d.* $N(0,1)$ *entries. Notably,* $\mathbf{M}_1(\Lambda)$ *is independent of* $\mathbf{M}_2(\Lambda)$*.*

Corollary 3.1 (First and second moments of $\mathbf{M}(\Lambda)$) *Suppose that Model* (1.1)–(1.2) *and Conditions (a)–(g) and* ***A1****–****A4*** *hold. Denote the* (i,j)*-th element of* $\mathbf{M}(\Lambda)$ *as* $m_{ij}(\Lambda)$*.*

$$\mathbb{E}[m_{ij}(\Lambda)] = \mathbb{1}_0(i-j)\{\Theta_{1p}(\Lambda) + O(n^{-3/2})\},$$

$$\mathrm{Var}[m_{ij}(\Lambda)] = \frac{1+\mathbb{1}_0(i-j)}{p}\Theta_{2p}(\Lambda) + O(n^{-5/2}),$$

$$\mathrm{Cov}[m_{ij}(\Lambda), m_{i'j'}(\Lambda)] = O(n^{-5/2}), \text{ if } i\neq i' \text{ or } j\neq j', \quad \text{where},$$

$$\Theta_{1p}(\Lambda) = \frac{1}{p}\sum_{j=1}^{K}\lambda_j[\zeta_j^2\ell_j + (1-\zeta_j^2)\sigma^2] + \frac{\lambda_0}{p}\Big(\sum_{j=1}^{K}\ell_j + (p-K)\sigma^2\Big),$$

$$\begin{aligned}\Theta_{2p}(\Lambda) = {} & \frac{1}{p}\sum_{j=1}^{K}\Big[\lambda_j^2\{\zeta_j^2\ell_j + (1-\zeta_j^2)\sigma^2\}^2 + 2\lambda_0\lambda_j\{\zeta_j^2\ell_j^2 + (1-\zeta_j^2)\sigma^4\}\Big] \\ & + \frac{\lambda_0^2}{p}\Big(\sum_{j=1}^{K}\ell_j^2 + (p-K)\sigma^4\Big).\end{aligned}$$

We then derive the asymptotic null distribution of $\mathbf{M}(\Lambda)$ when $n,p\to\infty$ simultaneously as in Condition **A1**. We use $\mathbf{W} = [w_{ij}]_{i,j=1}^q$ to denote the *Gaussian Orthogonal Ensemble*

(**GOE**) characterized by (1) $w_{ij} = w_{ji}$; (2) $w_{ii} \sim \mathcal{N}(0,1)$, $w_{ij} \sim \mathcal{N}(0,1/2),\ i \neq j$; (3) w_{ij}'s are jointly independent for $1 \leq i \leq j \leq q$.

Theorem 3.3 (Asymptotic null distribution of $\mathbf{M}(\Lambda)$) *Suppose that Model* (1.1)–(1.2)*, Conditions (a)–(g) and* ***A1****–****A4*** *hold. Then, for any fixed* Λ *with* $\lambda_0 \neq 0$*, under* H_0: $\mathbf{BC} = \mathbf{0}$*, we have*

$$\mathbf{M}_1(\Lambda) = O_p(1/p) \qquad \textit{and} \qquad \frac{\sqrt{p}\{\mathbf{M}_2(\Lambda) - \Theta_{1p}^0(\Lambda) I_q\}}{\{2\Theta_{2p}^0(\Lambda)\}^{1/2}} \Longrightarrow \mathbf{W},$$

where $\Theta_{1p}^0(\Lambda) = (1 - 2(K/p))\lambda_0\sigma^2$ *and* $\Theta_{2p}^0(\Lambda) = (1 - 2(K/p))\lambda_0^2\sigma^4$*. Consequently,*

$$\frac{\sqrt{p}\{\mathbf{M}(\Lambda) - \Theta_{1p}(\Lambda) I_q\}}{\{2\Theta_{2p}(\Lambda)\}^{1/2}} \Longrightarrow \mathbf{W}.$$

The following corollary follows immediately from Theorem 3.3 and the δ-method.

Corollary 3.2 *Suppose that Model* (1.1)–(1.2)*, Conditions (a)–(g) and* ***A1****–****A4*** *hold. Then, for any fixed* Λ *with* $\lambda_0 \neq 0$*, under* H_0: $\mathbf{BC} = \mathbf{0}$*, the appropriately normalized versions of* $T_0^{\mathrm{LR}}(\Lambda)$*,* $T_0^{\mathrm{LH}}(\Lambda)$ *and* $T_0^{\mathrm{BNP}}(\Lambda)$ *are asymptotically normal:*

$$\begin{aligned}
T^{\mathrm{LR}}(\Lambda) &:= \frac{\sqrt{p}\{1 + \Theta_{1p}(\Lambda)\}}{\{2q\Theta_{2p}(\Lambda)\}^{1/2}} \Big[T_0^{\mathrm{LR}}(\Lambda) - q \log\{1 + \Theta_{1p}(\Lambda)\} \Big] \Longrightarrow \mathcal{N}(0,1), \\
T^{\mathrm{LH}}(\Lambda) &:= \frac{\sqrt{p}}{\{2q\Theta_{2p}(\Lambda)\}^{1/2}} \Big[T_0^{\mathrm{LH}}(\Lambda) - q\Theta_{1p}(\Lambda) \Big] \Longrightarrow \mathcal{N}(0,1), \qquad (3.7) \\
T^{\mathrm{BNP}}(\Lambda) &:= \frac{\sqrt{p}\{1 + \Theta_{1p}(\Lambda)\}^2}{\{2q\Theta_{2p}(\Lambda)\}^{1/2}} \Big[T_0^{\mathrm{BNP}}(\Lambda) - \frac{q\Theta_{1p}(\Lambda)}{1 + \Theta_{1p}(\Lambda)} \Big] \Longrightarrow \mathcal{N}(0,1).
\end{aligned}$$

Remark 3.1 *Theorem 3.2 indicates that for finite* n *and* p *the distribution of the eigenvalues of* $\mathbf{M}(\Lambda)$ *is skewed, especially when* λ_0 *is relatively small compared to the elements of* Λ_+*. Indeed, the eigenvalues of* $\mathbf{M}_1(\Lambda)$ *have the same distribution as the non-zero eigen-*

values of a mixture of Wishart matrices of degrees of freedom K, resulting in a skewness in the null distributions of the proposed test statistics.

3.3 Power under local alternatives

In this subsection, we investigate the power characteristics of the proposed regularized tests under a class of local alternatives. The alternatives are selected such that the asymptotic power under which is non-trivial and can be decomposed into the contributions of the signal $\mathbf{BC}$ projected onto eigen-subspaces of the latent factors (spikes) and the idiosyncratic noise. The results presented here hold true for all three types of tests, namely, $T^{\mathrm{LR}}(\Lambda)$, $T^{\mathrm{LH}}(\Lambda)$ and $T^{\mathrm{BNP}}(\Lambda)$. Hence, we use the unifying notation $T(\Lambda)$ and denote the power function of $T(\Lambda)$ with critical value ξ, given $\mathbf{BC}$, as $\Upsilon(\mathbf{BC}, \Lambda) = \mathbb{P}(T(\Lambda) > \xi | \mathbf{BC})$.

Recall $\mathbf{H}_K = [\mathbf{h}_1, \cdots, \mathbf{h}_K]$ are the spiked eigenvectors of Σ_p (1.2). Denote

$$\begin{aligned} \tilde{\pi}_j &= \sqrt{n}\mathbf{h}_j^T \mathbf{BC} R_n^{-1} \mathbf{C}^T \mathbf{B}^T \mathbf{h}_j, \quad j = 1, \ldots, K, \\ \tilde{\pi}_0 &= \sqrt{n}\mathrm{Tr}[(I - \mathbf{H}_K \mathbf{H}_K^T)\mathbf{BC} R_n^{-1} \mathbf{C}^T \mathbf{B}^T], \end{aligned} \tag{3.8}$$

where $R_n = \mathbf{C}^T (n^{-1}\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{C}$. Note that $\mathbf{BC}R_n^{-1/2}$ is the noncentrality parameter of $\mathbf{Y}Q_N$ when $\mathbf{BC}$ is nonzero, $\tilde{\pi}_j$ $(1 \le j \le K)$ represents the $\sqrt{n}$-scaled signal strength associated with the subspace corresponding to the j-th eigenvector $\mathbf{h}_j$ of Σ_p, and $\tilde{\pi}_0$ is the $\sqrt{n}$-scaled contribution of the signal strength associated with the subspace orthogonal to the spiked eigenvectors. Thus, $\tilde{\pi}_0$ can be seen as the scaled contribution to the signal strength by the subspace associated with the idiosyncractic noise in (1.2). Clearly, $\tilde{\pi}_j \ge 0$ for all j.

To achieve non-trivial asymptotic power, we consider the sequence of $\mathbf{BC}$ to be local in the sense that

$$\tilde{\pi}_j \to \pi_j, \ j = 0, 1, \ldots, K, \ \text{as } n \to \infty, \tag{3.9}$$

where π_j's are finite, nonnegative deterministic constants. Notice that it implies that $\mathrm{Tr}[\mathbf{BC}R_n^{-1}\mathbf{C}^T\mathbf{B}^T] = O(1/\sqrt{n})$. Let $\Pi = (\pi_0, \pi_1, \ldots, \pi_K)$.

Theorem 3.4 (Asymptotic power under local alternatives) *Suppose that Model* (1.1)–(1.2) *,Conditions (a)–(g) and* ***A1****–****A4*** *hold. If under* H_a, $\mathbf{BC}$ *satisfies* (3.9)*, then the power of the test* $T(\Lambda)$ *satisfies*

$$\Upsilon(\mathbf{BC}, \Lambda) - \Phi\Big(-\xi + \frac{\mathcal{H}(\Lambda, \gamma_n; \Pi)}{\{2\gamma_n q \Theta_{2p}(\Lambda)\}^{1/2}}\Big) \longrightarrow 0, \quad \textit{where} \tag{3.10}$$

$$\mathcal{H}(\Lambda, \gamma_n; \Pi) = \sum_{j=1}^{K} \pi_j \lambda_j \zeta^2(\ell_j/\sigma^2, \gamma_n) + \lambda_0 \sum_{j=0}^{K} \pi_j. \tag{3.11}$$

3.4 Empirically normalized tests and critical values

In order to make use of the normalized test statistics given in Corollary 3.2, we need consistent estimators of $\Theta_{1p}(\Lambda)$ and $\Theta_{2p}(\Lambda)$. In Algorithm 1, we first propose consistent estimators of ℓ_j and σ^2 by adjusting for the bias in the MLEs τ_j and $\widetilde{\sigma}^2_{\text{Full}}$ presented in Theorem 3.1. We then use these estimates through a simple plug-in strategy to get estimators of $\Theta_{1p}(\Lambda)$ and $\Theta_{2p}(\Lambda)$.

Algorithm 1 (Adjusted estimation of spikes and noise variance)

Step 1 *Utilizing (3.2), calculate a first-iteration estimator of* ℓ_j*, say* $\hat{\ell}_j^{\,(1)}$*, by solving the following equation*

$$\tau_j = \widetilde{\sigma}^2_{\text{Full}} \psi(\hat{\ell}_j^{\,(1)}/\widetilde{\sigma}^2_{\text{Full}}, \gamma_n) = \hat{\ell}_j^{\,(1)} \Big(1 + \frac{\gamma_n}{\hat{\ell}_j^{\,(1)}/\widetilde{\sigma}^2_{\text{Full}} - 1}\Big),$$

with $\gamma_n = p/n$*. The equation has two real-valued roots if* $\tau_j/\widetilde{\sigma}^2_{\text{Full}} \geq (1 + \sqrt{\gamma_n})^2$ *and two complex-valued roots if otherwise. Take* $\hat{\ell}_j^{\,(1)} = \max\{Re(r_1), Re(r_2)\}$*, where* r_1

and r_2 are the two roots of the above equation and $Re(r)$ is the real part of r.

Step 2 *Denote $b_0 = K + \sum_{j=1}^{K}(\widetilde{\sigma}_{\text{Full}}^{-2}\hat{\ell}_j^{(1)} - 1)^{-1}$. The adjusted estimator of σ^2 is*

$$\hat{\sigma}^2 = \widetilde{\sigma}_{\text{Full}}^2\Big(1 + \frac{\gamma_n b_0}{p-K}\Big). \tag{3.12}$$

Notably, $\hat{\sigma}^2$ is the bias-corrected estimator of σ^2 proposed by Passemier et al. (2017).

Step 3 *Replace $\widetilde{\sigma}_{\text{Full}}^2$ with $\hat{\sigma}^2$ and repeat **Step 1** to solve for the second-iteration estimator $\hat{\ell}_j^{(2)}$ of ℓ_j. In case $\tau_j/\hat{\sigma}^2 < (1+\sqrt{\gamma_n})^2$, set $\hat{\ell}_j^{(2)} = \hat{\sigma}^2(1+\sqrt{\gamma_n})$ which is the proposed adjusted estimator of ℓ_j. For simplicity, we shall use $\hat{\ell}_j$ to denote $\hat{\ell}_j^{(2)}$.*

The following proposition describes consistency of the adjusted estimators, while (3.13) is proved in Passemier et al. (2017).

Proposition 3.1 *Suppose that* (1.1)– (1.2)*, Conditions (a)–(g) and **A1**–**A4** hold. Then,*

$$\frac{p-K}{\sigma^2\sqrt{2\gamma_n}}(\hat{\sigma}^2 - \sigma^2) \Longrightarrow \mathcal{N}(0,1), \tag{3.13}$$

$$\hat{\ell}_j - \ell_j = O_p(n^{-1/2}), \text{ for } j = 1, \ldots, K. \tag{3.14}$$

We then construct plug-in estimators of $\Theta_{1p}(\Lambda)$ and $\Theta_{2p}(\Lambda)$, denoted by $\hat{\Theta}_{1p}(\Lambda)$ and $\hat{\Theta}_{2p}(\Lambda)$, through substituting ℓ_j's and σ^2 in their definitions (Corollary 3.1) by the adjusted estimators $\hat{\ell}_j$'s and $\hat{\sigma}^2$, respectively.

Proposition 3.2 *Suppose that* (1.1)– (1.2)*, Conditions (a)–(g) and **A1**–**A4** hold. Then,*

$$\hat{\Theta}_{1p}(\Lambda) - \Theta_{1p}(\Lambda) = O_p(n^{-1}),$$

$$\hat{\Theta}_{2p}(\Lambda) - \Theta_{2p}(\Lambda) = O_p(n^{-1}).$$

The empirically normalized tests are denoted by $\hat{T}^{\mathrm{LR}}(\Lambda)$, $\hat{T}^{\mathrm{LH}}(\Lambda)$, and $\hat{T}^{\mathrm{BNP}}(\Lambda)$, where $\Theta_{1p}(\Lambda)$ and $\Theta_{2p}(\Lambda)$ are replaced by $\hat{\Theta}_{1p}(\Lambda)$ and $\hat{\Theta}_{2p}(\Lambda)$ in (3.7). Proposition 3.2 indicates that with the empirically normalized tests, both the asymptotic normality under the null hypothesis (Theorem 3.3) and power characteristics (Theorem 3.4) still hold.

Theorem 3.2, Corollary 3.2 and Proposition 3.2 suggest two ways of determining the critical values for any of the proposed tests at a given significance level α.

1. *Asymptotic normal quantiles:* The null hypothesis is rejected at asymptotic level α, if $\hat{T}(\Lambda) > \xi(\alpha)$, where $\xi(\alpha)$ is the $(1-\alpha)$ quantile of the standard normal distribution.

2. *Parametric bootstrapped quantiles:* With estimated L, Ψ_1 and Ψ_2 replacing the true quantities in (3.5) and (3.6), we obtain bootstrap replicates of $\mathbf{M}_1(\Lambda) + \mathbf{M}_2(\Lambda)$ by generating $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$ as independent matrices of i.i.d. $N(0,1)$ entries. We then approximate the null distribution of $\hat{T}(\Lambda)$ by replacing $\mathbf{M}(\Lambda)$ with bootstrapped $\mathbf{M}_1(\Lambda) + \mathbf{M}_2(\Lambda)$. The null hypothesis is rejected at asymptotic level α if $\hat{T}(\Lambda) > \xi_{\mathrm{boot}}(\alpha)$, where $\xi_{\mathrm{boot}}(\alpha)$ is the $(1-\alpha)$ quantile of the bootstrap samples of $\hat{T}(\Lambda)$.

Both procedures require consistent estimators of ℓ_j's and σ_0^2, presented in Section 3.4. Simulation studies suggest that normal quantiles lead to well controlled empirical sizes when λ_0 is comparable to λ_j's, $j \geq 1$. When λ_0 is relatively small, due to pronounced skewness of the sampling distributions of the regularized test statistics, the parametric bootstrap procedure provides better approximations to the null distributions.

4 Selection of regularization parameter

Our objective in this section is to propose a data-driven choice for the regularization parameter Λ, under a class of local alternatives as described in Section 3.3, that leads to

nontrivial local power. Here, the guiding principle is to select the minimax test among the class of tests $\{T(\Lambda)\colon \Lambda \in \mathbb{R}_+ \times \mathbb{R}^K\}$. Note however that we are not aiming to derive the optimal test among *all possible* tests. For this purpose, we rely on the expression of the local power in terms of the vector $\Pi = (\pi_0, \pi_1, \ldots, \pi_K)$, as stated in Theorem 3.4. Note that the proposed tests have the same asymptotic power under any two sequences of alternatives that lead to the same Π. Consider all alternatives with the same overall limiting signal strength such that $\sum_{j=0}^K \pi_j = \lim \sqrt{n}\mathrm{Tr}[\mathbf{B}\mathbf{C}R_n^{-1}\mathbf{C}^T\mathbf{B}^T] = 1$. Therefore, with the normalization $\sum_{j=0}^K \pi_j = 1$, we can treat Π as an *equivalence class of priors* on the parameter space. This interpretation is helpful since it allows us to borrow the techniques from classical decision theory, in terms of deriving a minimax procedure as a Bayes procedure under a least favorable prior corresponding to an associated Bayes risk.

The following results hold equally for $\hat{T}^{\mathrm{LR}}(\Lambda)$, $\hat{T}^{\mathrm{LH}}(\Lambda)$ and $\hat{T}^{\mathrm{BNP}}(\Lambda)$. The unifying notation $\hat{T}(\Lambda)$ is used to refer to any of the test statistics. Notice that $\hat{T}(\Lambda) = \hat{T}(\Lambda/\|\Lambda\|_2)$ for any non-zero Λ. Therefore, for the purpose of selecting Λ, it suffices to restrict Λ to the set $\mathbb{S} = \{\mathbf{d} \in \mathbb{R}^{K+1}\colon \|\mathbf{d}\|_2 = 1\}$.

For each equivalence class of prior Π, the associated (Bayes) risk is defined as the asymptotic Typer II error rate under the corresponding alternatives.

Definition 4.1 (Risk) *Consider a proposed test $\delta(\alpha, \Lambda) = \mathbb{1}(\hat{T}(\Lambda) > \xi(\alpha))$ at asymptotic level α where $\xi(\alpha)$ is the critical value at level α determined as in Section 3.4. Given the prior Π, we define its Bayes risk function corresponding to the prior Π as*

$$\mathfrak{R}_p(\delta(\alpha, \Lambda), \Pi) = 1 - \Phi\Big(-\xi(\alpha) + \frac{\mathcal{H}(\Lambda, \gamma_n; \Pi)}{\{2\gamma_n q \Theta_{2p}(\Lambda)\}^{1/2}}\Big). \tag{4.1}$$

Definition 4.2 (Bayes procedure) *Given the prior Π, a proposed test $\delta(\alpha, \Lambda_B(\Pi)) = \mathbb{1}(\hat{T}(\Lambda_B(\Pi)) > \xi(\alpha))$ at asymptotic level α is said to be the Bayes test with respect to Π, if*

the minimum of $\mathfrak{R}_p(\delta(\alpha,\Lambda),\Pi)$ *is obtained at* $\Lambda_B(\Pi)$. *We call* $\Lambda_B(\Pi)$ *the Bayes selection of* $\Lambda \in \mathbb{S}$ *with respect to prior* Π, *i.e.,*

$$\Lambda_B(\Pi) = \arg\min_{\Lambda\in\mathbb{S}} \mathfrak{R}_p(\delta(\alpha,\Lambda),\Pi) = \arg\max_{\Lambda\in\mathbb{S}} \frac{\mathcal{H}(\Lambda,\gamma_n;\Pi)}{\Theta_{2p}^{1/2}(\Lambda)}. \tag{4.2}$$

Proposition 4.1 *For a given* Π, *the Bayes selection of* Λ *is*

$$\Lambda_{\mathrm{B}}(\Pi) = \frac{\widetilde{\Lambda}_{\mathrm{B}}}{\|\widetilde{\Lambda}_{\mathrm{B}}\|_2}, \quad \textit{where} \;\; \widetilde{\Lambda}_{\mathrm{B}} = \mathbf{G}^{-1}\mathbf{b}(\Pi), \tag{4.3}$$

with $\mathbf{G}$ *and* $\mathbf{b}(\Pi)$ *defined as*

$$\mathbf{b}(\Pi) = (\sum_{j=0}^{K} \pi_j,\ \zeta_1^2\pi_1,\ \ \ldots,\ \zeta_K^2\pi_K)^T,$$

$$\mathbf{a} = \Big(\frac{1}{p}\{\zeta_1^2\ell_1^2 + (1-\zeta_1^2)\sigma^4\},\ \ \ldots\ ,\ \frac{1}{p}\{\zeta_K^2\ell_K^2 + (1-\zeta_K^2)\sigma^4\}\Big)^T,$$

$$\mathbf{J} = Diag\Big(\frac{1}{p}\{\zeta_1^2\ell_1 + (1-\zeta_1^2)\sigma^2\}^2,\ \ \ldots\ ,\ \frac{1}{p}\{\zeta_K^2\ell_K + (1-\zeta_K^2)\sigma^2\}^2\Big),$$

$$\mathbf{G} = \begin{bmatrix} c & \mathbf{a}^T \\ \mathbf{a} & \mathbf{J} \end{bmatrix},\ \textit{with}\ c = \frac{1}{p}\sum_{j=1}^{K}\ell_j^2 + \frac{p-K}{p}\sigma^4.$$

The risk at the Bayesian selection $\Lambda_B(\Pi)$, henceforth referred to as the *optimal Bayes risk* with respect to Π, is

$$\mathfrak{R}_p(\delta(\alpha,\Lambda_B(\Pi)),\Pi) = 1 - \Phi\Big(-\xi(\alpha) + \frac{1}{(2\gamma_n q)^{1/2}}\Big[\mathbf{b}^T(\Pi)\mathbf{G}^{-1}\mathbf{b}(\Pi)\Big]^{1/2}\Big). \tag{4.4}$$

4.1 Minimax selection of Λ

Consider a family of priors Π as

$$\mathfrak{P}(\varrho) = \Big\{\Pi\colon \sum_{j=0}^{K} \pi_j = 1, \;\; 0 \le \pi_0 \le \varrho, \;\; \text{and} \;\; 0 \le \pi_j \le 1, \;\; j = 1, \ldots, K\Big\},$$

where ϱ is the maximum fraction of the signal strength associated with the orthogonal complement to the subspace of spiked eigenvectors (i.e., associated with the idiosyncratic noise). The user-specified parameter ϱ can be seen as a hyperparameter that imposes restrictions on the class of alternatives. A larger value of ϱ leads to a bigger class of alternatives, and the value $\varrho = 1$ makes the class of alternatives unrestricted in terms of how much of the signal resides in the subspace associated with idiosyncratic noise.

Definition 4.3 (Minimax selection) *For a given $0 < \varrho \le 1$, a proposed test $\delta(\alpha, \Lambda^*(\varrho)) = \mathbb{1}(\hat{T}(\Lambda^*(\varrho)) > \xi(\alpha))$ at asymptotic level α is said to be minimax within the class $\{\delta(\alpha, \Lambda)\colon \Lambda \in \mathbb{S}\}$ with respect to the prior family $\mathfrak{P}(\varrho)$ if the minimum value of $\max_{\Pi\in\mathfrak{P}(\varrho)} \mathfrak{R}_p(\delta(\alpha, \Lambda), \Pi)$ is obtained at $\Lambda^*(\varrho)$, that is, $\Lambda^*(\varrho) = \arg\min_{\Lambda\in\mathbb{S}} \max_{\Pi\in\mathfrak{P}(\varrho)} \mathfrak{R}_p(\delta(\alpha, \Lambda), \Pi)$. We call $\Lambda^*(\varrho)$ the minimax selection of $\Lambda \in \mathbb{S}$ with respect to $\mathfrak{P}(\varrho)$.*

Using *Sion's Minimax Theorem* (Sion, 1958), we have the following proposition.

Proposition 4.2 *For a given $0 < \varrho \le 1$, the minimax risk is obtained at $(\Lambda^*(\varrho), \Pi^*(\varrho))$, where $\Pi^*(\varrho)$ is the least favorable prior defined as $\Pi^*(\varrho) = \arg\max_{\Pi\in\mathfrak{P}(\varrho)} \min_{\Lambda\in\mathbb{S}} \mathfrak{R}_p(\delta(\alpha, \Lambda), \Pi)$. And the minimax selection $\Lambda^*(\varrho)$ is the Bayes selection with respect to $\Pi^*(\varrho)$, i.e., $\Lambda^*(\varrho) = \Lambda_B(\Pi^*(\varrho)) = \arg\min_{\Lambda\in\mathbb{S}} \mathfrak{R}_p(\delta(\alpha, \Lambda), \Pi^*(\varrho))$.*

Proposition 4.2 suggests that to solve for the minimax selection $\Lambda^*(\varrho)$, we could first find the least favorable prior $\Pi^*(\varrho)$. $\Lambda^*(\varrho)$ is then the Bayes selection with respect to $\Pi^*(\varrho)$. A

detailed proof is presented in the Supplementary material. Motivated by this, and using (4.4), we have the following algorithm.

Algorithm 2 (Minimax selection of Λ) *Given $0 < \varrho \leq 1$:*

Step 1. ***Least favorable prior:*** *The least favorable prior $\Pi^*(\varrho)$ is the solution of the following quadratic programming problem.*

$$\textit{Minimize} \quad \mathbf{b}^T(\Pi)\mathbf{G}^{-1}\mathbf{b}(\Pi) \textit{ with respect to } \Pi, \quad \textit{subject to} \quad \Pi \in \mathfrak{P}(\varrho). \tag{4.5}$$

Step 2. ***Minimax selection:*** *The minimax choice of Λ with respect to $\mathfrak{P}(\varrho)$ is $\Lambda^*(\varrho) = \Lambda_B(\Pi^*(\varrho))$, where $\Lambda_B(\Pi^*(\varrho))$ is the Bayes selection with respect to $\Pi^*(\varrho)$ as in (4.3).*

In closing, we make the following interesting connection with a natural generalization of the test proposed by Bai and Saranadasa (1996) in the context of two-sample test for equality of population means. Note that the test proposed by Bai and Saranadasa (1996) replaces the estimator of Σ_p by the identity matrix in the likelihood ratio test.

Proposition 4.3 *The test proposed by Bai and Saranadasa (1996) is equivalent to the minimax selection of Λ when we restrict the space of the normalized regularization parameter Λ to be all unit vectors with nonnegative coordinates, denoted by $\mathbb{S}^+$ and when $\varrho = 1$, so that the space of priors Π is the entire K-dimensional unit simplex.*

As a remark, the parameter space $\mathbb{S}^+$ excludes tests proposed by Aoshima and Yata (2018) and Wang and Xu (2018), both of which allow negative values of λ_j, $j = 1, \ldots, K$.

We present a detailed analysis of the dependence of the regularization parameter Λ on ϱ in Section S.2 of the Supplementary Material in the single spike ($K = 1$) case.

5 Simulation study

The performance of the proposed tests is examined numerically through an extensive simulation study. Due to limited sapce, we only highlight key results here and defer the detailed simulation settings and additional results to Section S.4 of the Supplementary Material.

5.1 Empirical null distribution

The empirical null distributions of T^{LR} are shown in Figure 5.1 – 5.3. Corresponding figures for T^{LH} and T^{BNP} are included in the Supplementary Material. These figures indicate that the null distribution of the proposed tests are robust to the noise distribution. The distributions are mildly skewed when ϱ is relatively large, but significantly so when ϱ is small.

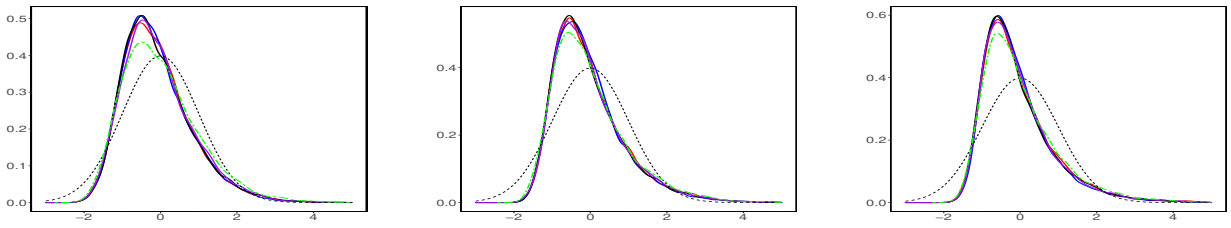

Figure 5.1: Empirical null distribution of T^{LR} with $\varrho = 0.2$ when $(N_1, N_2) = (40, 60)$ for the following noise settings: Normal (red solid), $t(4)$ (blue solid), $t(5)$ (black solid), $t(6)$ (purple solid). From left to right: $p = 50, 200, 1000$. The standard normal p.d.f. (theoretical limiting null distribution) is depicted as black dashed line. The oracle bootstrap null distribution when $\mathbf{M}(\Lambda^*)$ is approximated by $\mathbf{M}_1(\Lambda^*) + \mathbf{M}_2(\Lambda^*)$ is depicted in green dashed line.

Critical value	p	$\varrho = 0.2$			$\varrho = 0.5$			$\varrho = 0.8$			LRT
		BNP	LH	LR	BNP	LH	LR	BNP	LH	LR	
ξ_{norm}	50	3.55	6.41	4.90	2.93	5.70	4.30	2.93	5.70	4.30	8.20
	200	5.82	6.79	6.28	4.54	6.22	5.38	4.23	6.02	5.08	12.8
	1000	5.78	5.97	5.88	5.09	5.47	5.32	4.60	5.23	4.96	29.1
ξ_{boot}	50	4.80	4.80	4.80	4.42	4.42	4.42	4.40	4.40	4.40	8.20
	200	4.68	4.68	4.68	4.84	4.84	4.84	4.80	4.80	4.80	12.8
	1000	3.70	3.70	3.70	3.68	3.68	3.68	3.84	3.84	3.84	29.1

Table 5.1: Empirical Sizes $\times 100$ at 5% nominal significance level under the normal noise setting when $(N_1, N_2) = (40, 60)$, with normal critical values ξ_{norm} and bootstrapped critical values ξ_{boot}, respectively.

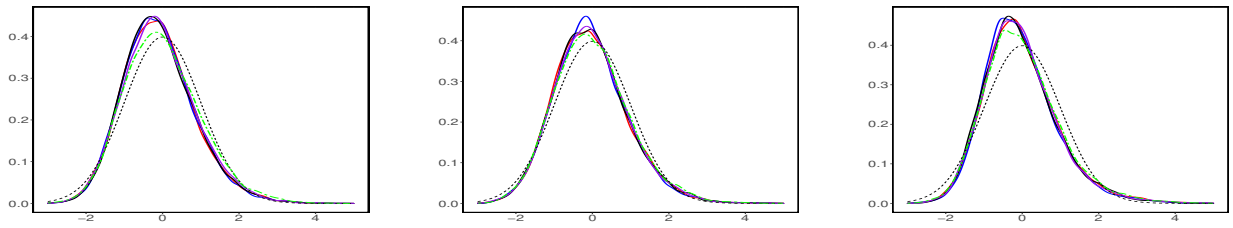

Figure 5.2: Same as Figure 5.1 but with $\varrho = 0.5$.

The empirical sizes under the normal noise setting are shown in Table 5.1 when the normal critical values and the bootstrapped critical values are used, respectively. The sizes when the observational noise is student-t distributed are reported in the Supplementary Material Tables S.4.1–Table S.4.12. These tables indicate that the empirical sizes of the proposed tests are reasonably controlled at the significance level 5% by both methods of choosing critical values. On the other hand, the sizes of LRT are inflated especially when p is large. As for the comparison among T^{LR}, T^{LH} and T^{BNP}, the results suggest that BNP is more conservative in terms of type I error rate, while LH is more liberal.

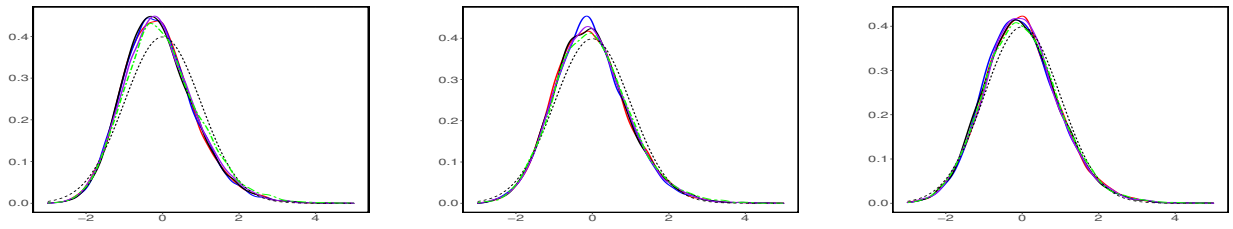

Figure 5.3: Same as Figure 5.1 but with $\varrho = 0.8$.

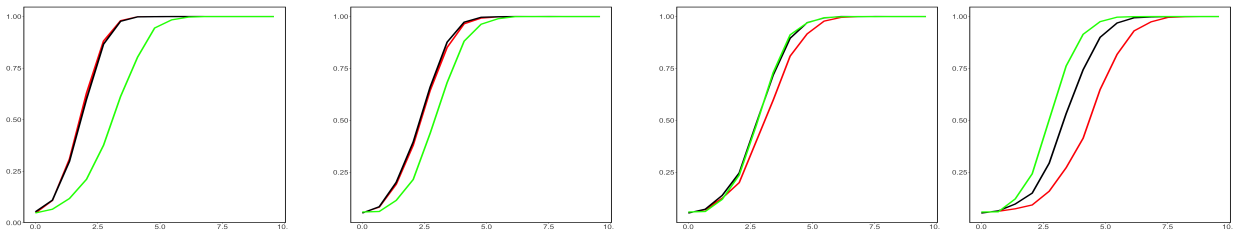

Figure 5.4: Size-adjusted empirical power when $p = 50$ and $(N_1, N_2) = (40, 60)$. From left to right: $\pi_0^{\mathrm{true}} = 0, 0.5, 0.8, 1$. LRT (Green), T^{LR} with $\varrho = 0.2$ (Red), $\varrho = 0.5$ (Blue), $\varrho = 0.8$ (Black).

5.2 Empirical power

The empirical power curves against the signal strength s are shown in Figures 5.4–5.6. To better compare the power across different tests, we utilize the size-adjusted critical values based on the Monte-Carlo null distribution computed on 10,000 independent replicates. The LR, LH and BNP tests behave similarly across simulation settings, as predicted by Theorem 3.4. So only the power of $T^{\mathrm{LR}}(\Lambda)$ is displayed for ease of visualization.

Figures 5.4–5.6 indicate that the power of the proposed tests is not too sensitive to the selection of ϱ, except when π_0^{true} is large. This means that unless the signal associated with the departure from the null hypothesis has little contribution from the leading eigen-directions of the noise covariance matrix Σ_p (i.e., $\pi_0^{\mathrm{true}} \approx 1$), the minimax test is robust to the choice of ϱ. When π_0^{true} is small and for larger dimension p, the proposed tests have significantly higher power than the LRT. On the other hand, when π_0^{true} is large, i.e., when a large portion of the signal is along the idiosyncratic noise eigen-directions, then the (size-adjusted) LRT outperforms the proposed tests. However, in practice LRT is only applicable when $p \ll n$ due to the lack of an approximation of its null distribution when p is comparable to or larger than n. Furthermore, the power of the proposed tests tends to improve if the specified ϱ is close to π_0^{true}.

In Section S.4.5 of the Supplementary Material, we consider two additional settings (1) when the leading spike is diverging and (2) when there are undetectable spikes. The settings are beyond **A2**. The reported results demonstrate that the proposed tests still have reasonable power, while the empirical sizes are reasonably controlled.

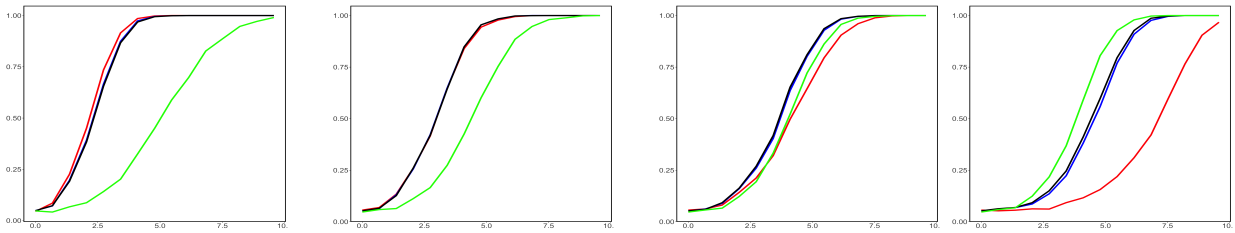

Figure 5.5: Same as Figure 5.4 but with $p = 200$.

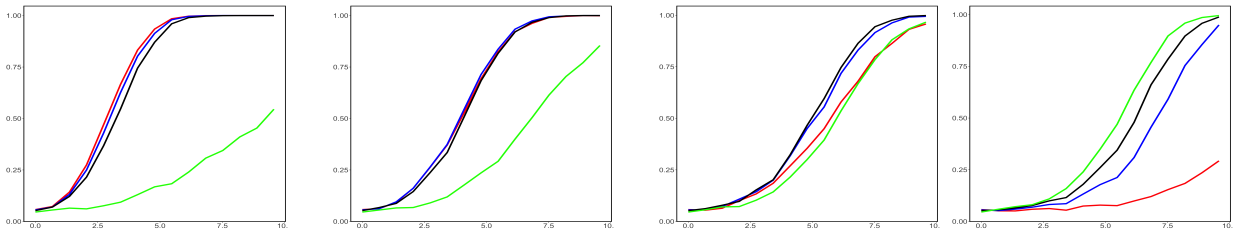

Figure 5.6: Same as Figure 5.4 but with $p = 1000$.

6 HCP application

The NIH funded *Human Connectome Project* (HCP) targets the characterization of the human connectome and its variability using cutting-edge neuroimaging technologies. As part of HCP, a consortium led by the Washington University and University of Minnesota aims to characterize human brain connectivity and functionality based on data collected from $N = 1113$ healthy young adults and to enable detailed comparisons between brain circuits, behavior, and genetics at the level of individual subjects. We refer to Van Essen et al. (2013) for details. Among the publicly available data are cerebral volumetric measurements and human behavior evaluation test scores. In this section, we use the proposed tests to study the association between cerebral measurements and human behaviors, using the aforementioned HCP young adults data.

The behavior evaluation scores belong to various behavioral domains: alertness, cognition, emotion, sensory and others. The alertness domain evaluates the cognitive status and sleep quality of the subjects based on a mental status exam and the *Pittsburgh Sleep Quality Index*. The cognition domain evaluates the subjects' cognitive abilities on vari-

ous aspects including episodic memory, cognitive flexibility, attention control and others. The emotion domain consists of indices of the ability to recognize emotions, psychological well-being, social relationships, stress and self efficacy. The motor domain measures cardiovascular endurance, manual dexterity, grip strength, and gait speed. The sensory domain includes auditive, olfaction, taste, and vision tests. After a pre-screening process that filters out highly correlated variables, we select 127 representative behavior variables and study whether the cerebral measurements are related to these variables.

For the cerebral measurements, we focus on cortcial surface regions that belong to 14 cerebral lobes symmetrically located on the two brain hemispheres, and 38 subcortical anatomy structures. Figure S.5.2 of the Supplementary Material shows part of the gyral-based regions. We refer to Desikan et al. (2006) for details. Our analysis is on the level of cortical lobes and subcortical structures and focuses on lobe surface area, average lobe thickness, lobe gray-matter volume, and subcortical structure volume.

Available demographics information includes the age and gender of subjects. Specifically, the subjects are divided into four age groups, namely 22–25, 26–30, 31–35, and 36+. The data set is roughly balanced with respect to gender (606 females and 507 males).

The foregoing leads to the multivariate regression model

$$\mathbf{y}_i = \beta_0 + \beta_1^T \mathbf{D}_i + \beta_2^T \mathbf{SA}_i + \beta_3^T \mathbf{AT}_i + \beta_4^T \mathbf{GV}_i + \beta_5^T \mathbf{SC}_i + L\mathbf{f}_i + \sigma e_i, \quad i = 1, \ldots, 1113, \tag{6.1}$$

where (i) $\mathbf{y}_i$ is the vector of 127 behavior scores of subject i ; (ii) $\mathbf{D}_i$ are age and gender group dummy variables of subject i (4 in total); (iii) $\mathbf{SA}_i$, $\mathbf{AT}_i$, $\mathbf{GV}_i$ are surface area, average thickness, gray-matter volume variables of the 14 cortical lobes of subject i ($14 \times 3 = 42$ in total); (iv) $\mathbf{SC}_i$ are subcortical structure volume variables (38 in total); and (v) $L\mathbf{f}_i$ and σe_i are latent factors and idiosyncratic noise as in (1.1)-(1.2). The dimension of the

explanatory variables (including the intercept) is $m = 85$, the dimension of response $\mathbf{y}_i$ is $p = 127$ and the sample size is $N = 1113$.

We use **Method 2** (Kritchman and Nadler, 2008) described in the Supplementary Material Section S.3.1 to determine $K = 12$ spikes. Figure S.5.1 shows the empirical eigenvalues of the residual covariance matrix, estimated spike variance and estimated noise variance with methods discussed in Section 3.4.

We test individually for the significance of the coefficients of lobe measurement variables and subcortical variables. We consider $\hat{T}^{\mathrm{LR}}(\Lambda)$ with minimax selection of Λ and two choices of ϱ, namely, $\varrho = 0.001$ and $\varrho = 1$. The reason for considering two different and extreme values of the hyper-parameter ϱ is to investigate robustness of its specification. Indeed, for the estimated $\hat{\ell}_j$'s and $\hat{\sigma}^2$, under the unrestricted condition (i.e., $\varrho = 1$), the minimax selection Λ^* is obtained at where the associated least favorable prior Π^* is such that $\pi_0^* = 0.014$. Therefore, for $\varrho \geq 0.014$, the restriction of $\pi_0 \leq \varrho$ in $\mathfrak{P}(\varrho)$ is inactive. It implies that the minimax selection $\Lambda^*(\varrho)$ is identical for $0.014 \leq \varrho \leq 1$ and so would be the testing results. As a comparison, we also consider the likelihood ratio test (LRT). The p-values for the proposed method are calculated based on the asymptotic normal distribution. The p-values for the LRT are calculated based on χ^2-approximation as described in Section 2.1. The p-values under $\varrho = 1$ and $\varrho = 0.001$ for the propose method and those for the LRT are reported in Table S.5.1. For the proposed method, the results under $\varrho = 1$ and $\varrho = 0.001$ are not greatly dissimilar.

Among the significant coefficients by the proposed tests at 10% level, some are associated with the volumes of left amygdala and right hippocampus. The Amygdala performs primary roles in the formation and storage of memories associated with emotional events (Maren, 1999). The hippocampus plays an important role in the formation of new memories

about experienced events (Eichenbaum et al., 1993). Among the behavioral variables, there are emotion processing tasks that may lead to activation of amygdala and hippocampus (Barch et al., 2013). Among other significant regions, the medial temporal lobe contains the parahippocampal and entorhinal cortices which are among the primary regions deemed responsible for the formation of memories and spatial cognition. These cortices are anatomically adjacent to, and also functionally communicates with, amygdala and hippocampus (Koob et al., 2010).

7 Discussion

In this paper, we addressed the problem of testing general hypothesis in a high-dimensional latent factor linear regression model. We proposed a family of regularized tests where the signal is projected onto the estimated latent factor directions and the weights on these directions act as regularization parameters. We studied their asymptotic null distributions and the asymptotic power under a class of local alternatives. Taking this approach further, we established a minimax criterion to select the regularization parameters by considering an ensemble of priors.

Our asymptotic results rely on the Gaussianity mainly because that the decomposition of the leading eigenvectors as in (3.3) of Theorem 3.1 is known to be valid for Gaussian data in the literature, e.g., Paul (2007) and Onatski (2012), since Gaussianity encapsulates a rotational invariance of the sample eigenvector. This invariance enables a transparent asymptotic representation of the sample eigenvector associated with a spiked sample eigenvalue when the corresponding population eigenvalue is above the phase transition threshold.

Behaviors of both the spiked eigenvalues and the associated eigenvectors in non-Gaussian settings have been studied in Bai and Yao (2008), Benaych-Georges and Nadakuditi (2012),

Shi (2013), Bloemendal et al. (2016) and Bao et al. (2020). For non-Gaussian spiked models, Shi (2013) showed that when the first four moments of the data match the Gaussian case, the second-order fluctuations of both spiked sample eigenvalues and certain projections of the associated sample eigenvectors also match the Gaussian case. Recently, Bao et al. (2020) proved that $\langle \mathbf{a}, \mathbf{p}_j \rangle$, for an arbitrary unit vector $\mathbf{a}$, can be expressed up to second order as a quadratic functional of a fixed number of asymptotically Gaussian random variables. Moreover, if the first four moments of the observations match the Gaussian case, the limiting behavior of $\langle \mathbf{a}, \mathbf{p}_j \rangle$ mathes the Gaussian case. This mean that the behavior of sample spiked eigenvalues as well as linear functionals of the corresponding eigenvectors is similar to that in the Gaussian case up to the second order and suggests that the conclusions in this paper are likely to hold in non-Gaussian settings.

Supplementary material

Supplementary Material includes detailed proofs of the main theoretical results, additional details of the procedure and of the simulation study and real data application.

References

Abraham, G. and Inouye, M. (2014). Fast principal component analysis of large-scale genome-wide data. *PloS one*, 9(4):e93766.

Andersen, A. H., Gash, D. M., and Avison, M. J. (1999). Principal component analysis of the dynamic response measured by fmri: a generalized linear systems framework. *Magnetic resonance imaging*, 17(6):795–815.

Anderson, T. W. (1958). An introduction to multivariate statistical analysis.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 5, pages 111–150.

Aoshima, M. and Yata, K. (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica*, 28:43–62.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, Z., Jiang, D., Yao, J.-f., and Zheng, S. (2013). Testing linear hypotheses in high-dimensional regressions. *Statistics*, 47(6):1207–1223.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6:311–329.

Bai, Z. and Yao, J.-f. (2008). Central limit theorems for eigenvalues in a spiked population model. In *Annales de l'IHP Probabilités et statistiques*, volume 44, pages 447–474.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408.

Bao, Z., Ding, X., Wang, J., and Wang, K. (2020). Statistical inference for principal components of spiked covariance matrix. *arXiv preprint arXiv:2008.11903*.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., et al. (2013). Function in the human connectome: task-fmri and individual differences in behavior. *Neuroimage*, 80:169–189.

Benaych-Georges, F. and Nadakuditi, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate*

Analysis, 111:120–135.

Bloemendal, A., Knowles, A., Yau, H.-T., and Yin, J. (2016). On the principal components of sample covariance matrices. *Probability theory and related fields*, 164(1-2):459–552.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980.

Eaton, M. L. and George, E. I. (2021). Charles stein and invariance: Beginning with the hunt–stein theorem. *The Annals of Statistics*, 49(4):1815–1822.

Eichenbaum, H. et al. (1993). *Memory, amnesia, and the hippocampal system*. MIT press.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fujikoshi, Y. (2016). Likelihood ratio tests in multivariate linear model. *Applied Linear Algebra in Action*, page 139.

He, Y., Jiang, T., Wen, J., and Xu, G. (2021). Likelihood ratio test in multivariate linear regression. *Statistica Sinica*, 31(3):1215–1238.

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

Johnstone, I. M. and Paul, D. (2018). PCA in high dimensions: An orientation. *Proceedings of the IEEE*, 106(8):1277–1292.

Koob, G. F., Le Moal, M., and Thompson, R. F. (2010). *Encyclopedia of behavioral neuroscience*. Elsevier.

Kritchman, S. and Nadler, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, 94(1):19–32.

Li, H., Aue, A., and Paul, D. (2020a). High-dimensional general linear hypothesis tests via non-linear spectral shrinkage. *Bernoulli*, 26:2541–2571.

Li, H., Aue, A., Paul, D., Peng, J., and Wang, P. (2020b). An adaptable generalization of Hotelling's t^2 test in high dimension. *The Annals of Statistics*, 48:1815–1847.

Maren, S. (1999). Long-term potentiation in the amygdala: a mechanism for emotional learning and memory. *Trends in neurosciences*, 22(12):561–567.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258.

Passemier, D., Li, Z., and Yao, J. (2017). On estimation of the noise variance in high dimensional probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):51–67.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17:1617–1642.

Paul, D. and Aue, A. (2014). Random matrix theory in statistics. *Journal of Statistical Planning and Inference*, 150:1–29.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-

wide association studies. *Nature Genetics*, 38(8):904–909.

Shi, D. (2013). *Asymptotic Behavior of Eigenvalues and Eigenvectors of a Random Matrix*. PhD thesis, Stanford University.

Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The WU-Minn human connectome project: an overview. *Neuroimage*, 80:62–79.

Viviani, R., Grön, G., and Spitzer, M. (2005). Functional principal component analysis of fmri data. *Human Brain Mapping*, 24(2):109–129.

Wang, R. and Xu, X. (2018). On two-sample mean tests under spiked covariances. *Journal of Multivariate Analysis*, 167:225–249.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of snp data. *Bioinformatics*, 28(24):3326–3328.