Electronic Journal of Statistics
Vol. 16 (2022) 6489–6511
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2089

Directional testing for high dimensional multivariate normal distributions*

Caizhu Huang[1], **Claudia Di Caterina**[2] **and Nicola Sartori**[1]

[1] *Department of Statistical Sciences, University of Padova*
e-mail: caizhu.huang@phd.unipd.it; nicola.sartori@unipd.it

[2] *Department of Economics, University of Verona*
e-mail: claudia.dicaterina@univr.it

Abstract: Thanks to its favorable properties, the multivariate normal distribution is still largely employed for modeling phenomena in various scientific fields. However, when the number of components p is of the same asymptotic order as the sample size n, standard inferential techniques are generally inadequate to conduct hypothesis testing on the mean vector and/or the covariance matrix. Within several prominent frameworks, we propose then to draw reliable conclusions via a directional test. We show that under the null hypothesis the directional p-value is exactly uniformly distributed even when p is of the same order of n, provided that conditions for the existence of the maximum likelihood estimate for the normal model are satisfied. Extensive simulation results confirm the theoretical findings across different values of p/n, and show that under the null hypothesis the directional test outperforms not only the usual first and higher-order finite-p solutions but also alternative methods tailored for high dimensional settings. Simulation results also indicate that the power performance of the different tests depends on the specific alternative hypothesis.

MSC2020 subject classifications: Primary 62F03; secondary 62H15.
Keywords and phrases: Bartlett correction, exponential family, higher-order asymptotics, likelihood ratio statistic, saddlepoint approximation.

Received March 2022.

1. Introduction

Hypothesis testing on the multivariate normal distribution is a subject of great interest in multivariate statistical analysis [see, e.g., 1, 25]. It is widely applied in various fields, such as social sciences, biomedical sciences and finance. Under fixed dimension p of the observation vector and large sample size n, standard asymptotic results are available for testing hypotheses on the mean vector and/or the covariance matrix. For instance, for sufficiently large n, the classical log-likelihood ratio statistic, its Bartlett correction [3], and the large-deviation modification to the log-likelihood ratio statistic proposed by Skovgaard [29] have an approximate χ^2_d null distribution, with d equal to the number of constraints on the parameters imposed under the null hypothesis. Yet, in many modern

*Caizhu Huang thanks the support of Guangzhou University–University of Padova Joint PhD Program. Nicola Sartori's work was supported by a grant from the University of Padova (BIRD203991).

applications the dimension p, while being lower than n, is large and often comparable with n.

Taking inspiration from a classification recently proposed by Battey and Cox [4], we distinguish between three asymptotic regimes where p and n diverge, namely low dimensional, high dimensional and ultra-high dimensional. In the first case p/n goes to zero, in the second case $p/n \to \kappa \in (0,1]$, while in the third case p/n tends to infinity or to a limit greater than one. It is well known that inferential problems arise in regression settings where the number of covariates increases with the sample size, both in low and high dimensional regimes; see, for instance, [11], [30] and [31] for results on logistic regression, and [32] for results on exponential family models.

Even for the special p-variate normal distribution case, the classical likelihood-based testing procedures may be already invalid in low dimensional settings. Indeed, He, Meng, Zeng and Xu [14] considered the case where $p = O(n^{\alpha})$, $0 < \alpha \leq 1$, and showed that the log-likelihood ratio statistic's distribution approximates to a χ^2_d if and only if $p = o(n^{1/2})$, while the analogous condition for the Bartlett correction is $p = o(n^{2/3})$. On the other hand, for the high dimensional setting Jiang and Yang [18] derived a central limit theorem that allows to construct reliable tests for hypotheses on the mean vector and/or the covariance matrix. Specifically, they proved that the distribution of the log-likelihood ratio statistic, suitably standardized, converges to a standard normal when both p and n tend to infinity, provided that $p/n \to \kappa \in (0,1]$.

Under the same high dimensional setting of Jiang and Yang [18], we propose the use of a directional approach for testing general hypotheses on multivariate normal distributions. Directional inference on a vector parameter of interest was first introduced by Fraser and Massam [12] and later developed by Skovgaard [28] and Cheah, Fraser and Reid [6] using saddlepoint approximations for the distribution and analytical approximations for the required tail probability integrals in regular asymptotic scenarios. More recently, still assuming the classical scenario with fixed parameter dimension and increasing sample size, Davison, Fraser, Reid and Sartori [9] and Fraser, Reid and Sartori [13] proposed to compute the directional p-value by replacing the analytical approximations with one-dimensional numerical integration. The empirical evidence in these papers showed the excellent performance of the proposed method in many examples of practical interest. In fact, later the directional p-value was proven to coincide with that of an exact F-test in a few prominent modeling frameworks [23].

For multivariate normal distributions, a first example of the directional approach was given in [9, Example 5.3], who considered testing some conditional independence among the components starting from the full dependence structure. Although their simulations illustrated that the empirical extreme accuracy of directional inference is not limited to simple low-dimensional situations, those results were not formally justified within the asymptotic framework of the high dimensional regime.

We prove that the directional p-value is exact when testing a number of hypotheses on the multivariate normal distribution, even in the high dimensional scenario. Precisely, it is only required that $n \geq p + 2$, which is the condition for the existence of the maximum likelihood estimate for the covariance matrix. Consequently, we shall not deal here with the ultra-high dimensional regime. The exact uniform null distribution of the directional p-value follows from the exactness of the normalized saddlepoint approximation to the distribution of the canonical sufficient statistic in the multivariate normal model. We focus here on hypotheses that compare multiple multivariate normal distributions, while results on hypotheses regarding a single distribution are reported in the Supplementary Material [15].

Several simulation studies are conducted for comparing the proposed method with the usual χ^2_d approximations for the log-likelihood ratio statistic, the Bartlett correction, the modifications of the log-likelihood ratio statistic by Skovgaard [29] and the central limit theorem method by Jiang and Yang [18]. The results confirm the theoretical properties of the directional test under the null hypothesis. Indeed, the method proves uniformly more accurate than the alternative approaches. Only the central limit theorem test gives a comparable accuracy when the number of components p is large, but it is less reliable for small to moderate values of p. The various methods are assessed also in terms of power, after adjusting for Type I error. The approach leading to higher power depends on the specific alternative setting. In this respect the directional test, which does not need any adjustment, is competitive with the central limit theorem test and overperforms the other candidates across various alternative scenarios.

2. Background

2.1. Notation

Assume that the model for the data $y = (y_1, \dots, y_n)^T$ is an exponential family with canonical parameter $\varphi = \varphi(\theta)$ and canonical sufficient statistic $u = u(y)$. The distribution of y can then be expressed as

$$f(y;\theta) = \exp\left[\varphi(\theta)^T u(y) - K\{\varphi(\theta)\}\right] h(y),$$

with corresponding log-likelihood function $\ell(\theta; y) = \log f(y;\theta)$. When the dimension q of θ is equal to the dimension of φ and $\varphi(\theta)$ is one-to-one, the statistic $u(y)$ has a full natural exponential family distribution in the canonical parameterization. Hence, we can write $f(u;\varphi) = \exp\left\{\varphi^T u - K(\varphi)\right\} \tilde{h}(u)$ with associated log-likelihood $\ell(\varphi; u) = \varphi^T u - K(\varphi)$. However, $\tilde{h}(u)$ can rarely be derived explicitly.

For the development of the directional p-value in Section 2.2, it is notationally convenient to center the sufficient statistic at the observed data point y^0. Hence we let $s = u - u^0$, with $u^0 = u(y^0)$, and write

$$\ell(\varphi; s) = \varphi^T s + \ell^0(\varphi) = \varphi^T (u - u^0) + \ell(\varphi; u^0), \tag{2.1}$$

where $\ell^0(\varphi) = \ell(\varphi; s = 0_q) = \ell(\varphi; u = u^0)$, with 0_q denoting the q-dimensional vector of zeroes. This centering ensures that the observed value of s is $s^0 = 0_q$.

Suppose the parameter vector is partitioned as $\varphi = (\psi^T, \lambda^T)^T$, and thus (2.1) can be written as

$$\ell(\varphi; s) = \psi^T s_1 + \lambda^T s_2 + \ell^0(\psi, \lambda), \tag{2.2}$$

where ψ is a d-dimensional parameter of interest, λ is a $(q-d)$-dimensional nuisance parameter, and $(s_1^T, s_2^T)^T$ is the corresponding partition of s. Assume we are interested in testing the hypothesis H_ψ: $\psi(\varphi) = \psi$. To a first order of approximation, a parameterization-invariant measure of departure of s from H_ψ is given by the log-likelihood ratio statistic

$$W = 2\{\ell(\hat{\varphi}) - \ell(\hat{\varphi}_\psi)\},$$

where $\hat{\varphi}$ is the maximum likelihood estimate and $\hat{\varphi}_\psi = (\psi^T, \hat{\lambda}_\psi^T)^T$ denotes the constrained maximum likelihood estimate of φ under H_ψ. When q is fixed and $n \to +\infty$, the statistic W follows asymptotically a χ^2_d distribution with relative error of order $O(n^{-1})$ under H_ψ.

Higher-order improvements of likelihood inference for a vector parameter of interest are available. A first proposal is the Bartlett correction [3, 20], which rescales the log-likelihood ratio statistic by its expectation $E_\psi(W)$ under H_ψ, i.e.

$$W_{BC} = \frac{d}{E_\psi(W)} W,$$

and has a χ^2_d asymptotic null distribution with relative error of order $O(n^{-2})$ [8, Section 7.3; 24, Section 7.4]. Since the calculation of $E_\psi(W)$ is generally not feasible, Lawley [20] gave an asymptotic expansion for the exact expectation under H_ψ with error of order $O(n^{-1})$. However, the accuracy of Bartlett correction can be lost when the exact expectation is substituted with such asymptotic expansion [29, 9]. Increasing the computational cost, it is possible to replace analytical expansions by parametric bootstrap approximations [7, Section 2.7]. In the present framework, $E_\psi(W)$ can be computed exactly and the condition for validity of the χ^2_d approximation for the distribution of the Bartlett correction of W is $p = o(n^{2/3})$ [14]. The quantities needed for the Bartlett correction can be found in [18].

Starting from the extremely accurate r^* statistic for inference on a scalar ψ [2], for a vector parameter of interest Skovgaard [29] proposed two modifications of W designed to maintain high accuracy in the tails of the distribution:

$$W^* = W\left(1 - \frac{1}{W}\log\gamma\right)^2 \quad \text{and} \quad W^{**} = W - 2\log\gamma. \tag{2.3}$$

The statistics W^* and W^{**} are generally easier to calculate than the Bartlett correction and, under standard regularity conditions [see, e.g., 27, Section 3.4], they are also approximately distributed as χ^2_d when the null hypothesis holds. Even though the relative error is of order $O(n^{-1})$, as for W, they exhibit higher accuracy due to large-deviation properties of the saddlepoint approximation involved in their derivation [29]. Among the two forms, W^* has the advantages of being always non-negative and of reducing to the square of Barndorff-Nielsen's

r^* statistic when $d = 1$ [29]. The general expression of the correction factor γ in exponential families [29, Eq. (13)] simplifies to

$$\gamma = \frac{\left\{(s - s_\psi)^T J_{\varphi\varphi}(\hat{\varphi}_\psi)^{-1}(s - s_\psi)\right\}^{d/2}}{W^{d/2-1}(\hat{\varphi} - \hat{\varphi}_\psi)^T(s - s_\psi)} \left\{\frac{|J_{\varphi\varphi}(\hat{\varphi}_\psi)|}{|J_{\varphi\varphi}(\hat{\varphi})|}\right\}^{1/2}, \tag{2.4}$$

where s_ψ is the expected value of the sufficient statistic s under H_ψ and $J_{\varphi\varphi}(\varphi) = -\partial^2\ell(\varphi; s)/\partial\varphi\ \partial\varphi^T$ is the observed information matrix for φ, which coincides with the expected Fisher information matrix since φ is the canonical parameter. In order to calculate the p-value, the quantity (2.4) is evaluated at $s = s^0 = 0_q$, corresponding to $y = y^0$.

2.2. Directional tests in linear exponential families

Directional tests for a vector parameter of interest in exponential family models were considered in [28], [6] and [9]. In particular, the latter proposed to compute the directional p-value via one-dimensional integration. Directional tests in linear exponential families are essentially developed in two dimension-reduction steps, since the sufficient statistic has the same dimension of the canonical parameter φ. Specifically, the first step consists of reducing the dimension of the sufficient statistic from q to the dimension of the parameter of interest d; indeed, the conditional distribution of the component relative to ψ of the sufficient statistic in (2.2), s_1, given the component relative to λ, s_2, can be accurately approximated using saddlepoint approximations. The second step further reduces the d-variate conditional distribution to a one-dimensional conditional distribution given the direction indicated by the observed data point. We review here the key methodological steps, already detailed in [9], to derive the directional p-value in linear exponential families.

The simplicity of exponential families makes conditional inference a practicable strategy. In particular, the theory guarantees that the conditional distribution of the component of interest s_1 of the canonical sufficient statistic, given s_2, depends only on ψ [see, e.g., 26, Theorem 5.6]. Indeed, we have

$$f(s_1|s_2;\psi) = \exp\left\{\psi^T s_1 - K_{s_2}(\psi)\right\} h_{s_2}(s_1),$$

where the cumulant generating function $K_{s_2}(\psi)$ and the marginal density $h_{s_2}(s_1)$ depend on the conditioning value s_2 and can rarely be derived explicitly. However, as in [9], a saddlepoint approximation [see, e.g., 26, Section 10.10] can be used instead. Under the null hypothesis H_ψ, the saddlepoint approximation to the density of s_1 given s_2 is expressed as

$$h(s;\psi) = c \exp\left[\ell(\hat{\varphi}_\psi; s) - \ell\{\hat{\varphi}(s); s\}\right] |J_{\varphi\varphi}\{\hat{\varphi}(s); s\}|^{-1/2}, \quad s \in L^0_\psi, \tag{2.5}$$

where c is a normalizing constant and L^0_ψ is a d-dimensional plane defined by setting s_2 to its observed value, i.e. $s_2 = 0_{q-d}$. All values of s in L^0_ψ have the same constrained maximum likelihood estimate $\hat{\lambda}_\psi = \hat{\lambda}^0_\psi$, while they have unconstrained maximum likelihood estimate $\hat{\varphi}(s)$.

We construct a directional test for H_ψ by considering the one-dimensional model based on the magnitude of s, $||s||$, conditional on its direction. This is done by defining a line L^*_ψ in L^0_ψ through the observed value of s, $s^0 = 0_q$, and the expected value of s under H_ψ, s_ψ, which depends on the observed data point y^0, i.e.

$$s_\psi = -\ell^0_\varphi\left(\hat{\varphi}^0_\psi\right) = \left\{-\ell^0_\psi\left(\hat{\varphi}^0_\psi\right)^T, 0^T_{q-d}\right\}^T. \tag{2.6}$$

We parameterize this line by $t \in \mathbb{R}$, namely $s(t) = s_\psi + t(s^0 - s_\psi)$. In particular, $t = 0$ and $t = 1$ correspond, respectively, to the expected value s_ψ and to the observed value s^0. The conditional distribution of $||s||$ given the unit vector $s/||s||$ is obtained from (2.5) by a change of variable from s to $(||s||, s/||s||)$. The Jacobian of the transformation is proportional to t^{d-1}. The directional p-value to measure the departure from H_ψ along the line L^*_ψ is defined as the probability that $s(t)$ is as far or farther from s_ψ than is the observed value s^0, that is,

$$p(\psi) = \frac{\int_1^{t_{\sup}} t^{d-1} h\{s(t);\psi\}dt}{\int_0^{t_{\sup}} t^{d-1} h\{s(t);\psi\}dt}, \tag{2.7}$$

where the denominator is a normalizing constant. See [9, Section 3.2] for more details. The upper limit of the integrals in (2.7) is the largest value of t for which the maximum likelihood estimate $\hat{\varphi}(t)$ corresponding to $s(t)$ exists; depending on the case, it can be found analytically or approximated numerically. The scalar integrals in (2.7) can be accurately computed via numerical integration. The error in (2.7) is therefore essentially given by the error from the saddelpoint approximation used in (2.5). Some results on such an error when p increases with n are given in [33]. However, in all settings described in Section 2.3, (2.5) holds exactly, up to the normalizing constant c. Thus, since c simplifies in the ratio (2.7), also the directional p-value is exact. The results are formally derived in Section 3.

2.3. Multivariate normal distribution

Let $y_1, \ldots, y_n$ be a sample of independent observations from a multivariate normal distribution $N_p(\mu, \Lambda^{-1})$, where both the mean vector μ and the concentration matrix Λ, symmetric and positive definite, are unknown. Let $y = [y_1 \cdots y_n]^T$ denote the $n \times p$ data matrix and $\mathrm{tr}(M)$ denote the trace of a square matrix M. Define by $\mathrm{vec}(M)$ the operator which transforms the matrix M into a vector by stacking its columns one underneath the other. When M is symmetric it is useful to consider also $\mathrm{vech}(M)$, which is obtained from $\mathrm{vec}(M)$ by eliminating all supradiagonal elements of M. The two operators satisfy $D_p\mathrm{vech}(M) = \mathrm{vec}(M)$, where D_p is the duplication matrix [22, Section 3.8]. The log-likelihood for the parameter $\theta = \{\mu^T, \mathrm{vech}(\Lambda^{-1})^T\}^T$ is

$$\ell(\theta; y) = \mu^T\Lambda y^T 1_n - \frac{1}{2}\mathrm{tr}(\Lambda y^T y) + \frac{n}{2}\log|\Lambda| - \frac{n}{2}\mu^T\Lambda\mu.$$

The canonical parameter in this exponential family model is given by $\varphi = \{\xi^T, \text{vech}(\Lambda)^T\}^T = \{\mu^T\Lambda, \text{vech}(\Lambda)^T\}^T$ with canonical sufficient statistic $u = \{n\bar{y}^T, -\frac{1}{2}\text{vech}(y^Ty)^T D_p^T D_p\}^T$, and the corresponding log-likelihood is

$$\begin{aligned}\ell(\varphi; y) &= n\xi^T\bar{y} - \frac{1}{2}\text{tr}(\Lambda y^T y) + \frac{n}{2}\log|\Lambda| - \frac{n}{2}\xi^T\Lambda^{-1}\xi \\ &= \xi^T n\bar{y} - \text{vech}(\Lambda)^T\left\{\frac{1}{2}D_p^T D_p \text{vech}(y^T y)\right\} + \frac{n}{2}\log|\Lambda| - \frac{n}{2}\xi^T\Lambda^{-1}\xi,\end{aligned}$$

where $\bar{y} = y^T 1_n/n$ with 1_n a n-dimensional vector of ones. The score function with respect to the canonical parameter φ is

$$\begin{aligned}\ell_\varphi(\varphi) &= \left\{\ell_\xi(\varphi)^T, \ell_{\text{vech}(\Lambda)}(\varphi)^T\right\}^T \\ &= \left\{n\bar{y}^T - n\xi^T\Lambda^{-1}, \ \frac{n}{2}\text{vech}\left(\Lambda^{-1} - y^T y/n + \Lambda^{-1}\xi\xi^T\Lambda^{-1}\right)^T\right\}^T.\end{aligned}$$

The maximum likelihood estimates for μ and Λ^{-1} are $\hat{\mu} = \bar{y}$ and $\hat{\Lambda}^{-1} = y^T y/n - \bar{y}\bar{y}^T$, respectively; thus, $\hat{\xi} = \hat{\Lambda}\hat{\mu}$. Moreover, the observed information matrix for components ξ and $\text{vech}(\Lambda)$ of φ can be written in block form as

$$J_{\varphi\varphi}(\varphi) = \begin{bmatrix} n\Lambda^{-1} & -n(\xi^T\Lambda^{-1} \otimes \Lambda^{-1})D_p \\ -nD_p^T(\Lambda^{-1}\xi \otimes \Lambda^{-1}) & \frac{n}{2}D_p^T\{\Lambda^{-1}(\mathrm{I}_p + 2\xi\xi^T\Lambda^{-1}) \otimes \Lambda^{-1}\}D_p \end{bmatrix},$$

where $\otimes$ denotes the Kronecker product [see, e.g., 19, Section 5.1]. Finally, the determinant of the observed information matrix, appearing in (2.5), satisfies $|J_{\varphi\varphi}(\varphi)| \propto |\Lambda^{-1}|^{p+2}$ (see Supplementary Material S1.1).

3. Directional test for multiple-sample hypotheses

We consider now testing two hypotheses on the parameters of the multivariate normal model presented in Section 2.3. In particular, we concentrate here on: (I) equality of covariance matrices in k independent groups; (II) equality of multivariate normal distributions in k independent groups. We also obtained similar theoretical results for four one-sample hypotheses about: (III) sphericity of the covariance matrix; (IV) block-independence; (V) complete independence; (VI) specified values for the mean vector and the covariance matrix. The detailed results for cases (III)-(VI) are available in Supplementary Material S2. In all hypotheses, it is shown that the saddlepoint approximation (2.5) is exact, consequently leading to an exact directional p-value, up to the error from the scalar numerical integrations in (2.7).

3.1. Testing the equality of covariance matrices in k independent groups

Suppose $y_{i1}, \dots, y_{in_i}$, for $i \in \{1, \dots, k\}$, $k \geq 2$, are independent realizations of $N_p(\mu_i, \Lambda_i^{-1})$. We focus on testing the null hypothesis

$$H_\psi : \Lambda_1 = \cdots = \Lambda_k. \tag{3.1}$$

In the following, with a slight abuse of notation, let y_i denote the $n_i \times p$ data matrix of the i-th group. We then have $\bar{y}_i = y_i^T 1_{n_i}/n_i$ and $A_i = y_i^T y_i - n_i \bar{y}_i \bar{y}_i^T$. The unconstrained maximum likelihood estimates for all $i \in \{1, \dots, k\}$ are $\hat{\mu}_i = \bar{y}_i$ and $\hat{\Lambda}_i^{-1} = A_i/n_i$; the constrained maximum likelihood estimates are instead $\hat{\mu}_{0i} = \bar{y}_i$ and $\hat{\Lambda}_0^{-1} = \sum_{i=1}^k A_i/n$, where $n = \sum_{i=1}^k n_i$. Bartlett [3] suggested to use the modified maximum likelihood estimator of Λ^{-1}, that is $\tilde{\Lambda}_i^{-1} = A_i/(n_i - 1)$ and $\tilde{\Lambda}_0^{-1} = \sum_{i=1}^k A_i/(n-k)$. The modified log-likelihood ratio statistic is then equal to

$$\widetilde{W} = \sum_{i=1}^{k} -(n_i - 1) \log |\tilde{\Lambda}_i^{-1} \tilde{\Lambda}_0|.$$

The null distribution of $\widetilde{W}$ is approximately χ_d^2 with $d = kp(p+1)/2 - p(p+1)/2 = p(p+1)(k-1)/2$ if and only if $p = o(n_i^{1/2})$, and the analogous condition for the Bartlett correction is $p = o(n_i^{2/3})$, for all $i \in \{1, \dots, k\}$ with finite k [14]. The expression for Skovgaard's modifications [29] can be found in Supplementary Material S1.2.

For the directional p-value, under H_ψ, the expectation of s has components

$$-\left\{ 0_p^T,\ \frac{n_i}{2} \text{vech} \left(\hat{\Lambda}_0^{-1} - \hat{\Lambda}_i^{-1} \right)^T \right\}^T, \quad i \in \{1, \dots, k\},$$

and the tilted log-likelihood, by group independence, can be written as $\ell(\varphi; t) = \sum_{i=1}^k \ell_i(\varphi_i; t)$ with the i-th group's contribution

$$\begin{aligned} \ell_i(\varphi_i; t) \quad &= \quad n_i \xi_i^T \bar{y}_i - \frac{n_i}{2} \text{tr} \left[\Lambda_i \left\{ \frac{y_i^T y_i}{n_i} + (1-t) \left(\hat{\Lambda}_0^{-1} - \hat{\Lambda}_i^{-1} \right) \right\} \right] \\ &\quad + \frac{n_i}{2} \log |\Lambda_i| - \frac{n_i}{2} \xi_i^T \Lambda_i^{-1} \xi_i. \end{aligned}$$

Maximizing the tilted log-likelihood leads to the estimates $\hat{\mu}_i(t) = \bar{y}_i$ and $\hat{\Lambda}_i(t)^{-1} = (1-t)\hat{\Lambda}_0^{-1} + t\hat{\Lambda}_i^{-1}$, $i \in \{1, \dots, k\}$. Hence, the saddlepoint approximation (2.5) takes the form

$$h\{s(t); \psi\} = c \exp \left\{ \sum_{i=1}^{k} \frac{n_i - p - 2}{2} \log |\hat{\Lambda}_i^{-1}(t)| \right\},$$

where c is a normalizing constant.

The value t_{sup} in (2.7) is the largest t for which $\hat{\Lambda}_i(t)^{-1}$ is positive definite and is equal to $\{1 - \min_{1 \le i \le k} \nu_{(1)}^i\}^{-1}$ where $\nu_{(1)}^i$ is the smallest eigenvalue of $\hat{\Lambda}_0 \hat{\Lambda}_i^{-1}$ (see Lemma 4.2 in Section 4.1).

Since $\bar{y}_i$ and $\hat{\Lambda}_i^{-1}$ in the multivariate normal distribution are independent, we have that the saddlepoint approximation (2.5) to the density of s is exact, and therefore the directional p-value follows exactly a uniform distribution, even in high dimensional settings with p allowed to grow with n_i. The exact condition for the validity of this result is given in the following theorem.

Theorem 3.1. *Assume that $p = p_n$ such that $n_i \geq p + 2$ for all $n_i \geq 3, i \in \{1, \ldots, k\}$, with k fixed. Then, under the null hypothesis H_ψ (3.1), the directional p-value (2.7) is exactly uniformly distributed.*

The proof of Theorem 3.1 is given in Appendix A.1. Theorem 3.1 only requires $n_i \geq p + 2$, $i \in \{1, \ldots, k\}$, for ensuring that the maximum likelihood estimate of the covariance matrix exists with probability one. This assumption is weaker than the condition $p/n_i \to \kappa \in (0, 1]$ in [18] for the validity of their central limit theorem approximation with large p. Moreover, although the number of groups k is considered here as fixed, simulation results show that the accuracy of the directional test is not affected by the value of k (see Supplementary Material S3).

3.2. Testing the equality of several multivariate normal distributions

Under the same framework introduced in Section 3.1, we are interested in testing whether the multivariate normal distributions in k independent groups are identical, meaning

$$H_\psi : \mu_1 = \cdots = \mu_k, \Lambda_1 = \cdots = \Lambda_k. \tag{3.2}$$

The empirical within-groups variance A/n and the empirical between-groups variance B/n depend on the quantities $A = \sum_{i=1}^k y_i^T y_i - n_i \bar{y}_i \bar{y}_i^T$ and $B = \sum_{i=1}^k n_i \bar{y}_i \bar{y}_i^T - n\bar{y}\bar{y}^T$, such that $A + B = \sum_{i=1}^k y_i^T y_i - n\bar{y}\bar{y}^T$, where $\bar{y} = \sum_{i=1}^k n_i \bar{y}_i / n$. The unconstrained maximum likelihood estimates for all $i \in \{1, \ldots, k\}$ are the same as in hypothesis (3.1), while the constrained maximum likelihood estimates are $\hat{\mu}_0 = \bar{y}$, $\hat{\Lambda}_0^{-1} = (A + B)/n$. In this case, the log-likelihood ratio statistic is

$$W = n \log |\hat{\Lambda}_0^{-1}| - \sum_{i=1}^k n_i \log |\hat{\Lambda}_i^{-1}| \,,$$

which has asymptotically a χ_d^2 null distribution with $d = \{p(p+1)/2 + p\}(k-1) = p(p+3)(k-1)/2$, provided that $p = o(n_i^{1/2})$ for all $i \in \{1, \ldots, k\}$. The analogous condition for the Bartlett correction is $p = o(n_i^{2/3})$ for all $i \in \{1, \ldots, k\}$ [14]. The expressions for Skovgaard's [29] modifications of the likelihood ratio statistic can be found in Supplementary Material S1.2.

In order to obtain the directional p-value, we find the components of s_ψ

$$-\left\{ n_i(\bar{y}_i - \bar{y})^T,\ \frac{n_i}{2} \text{vech}\left(\hat{\Lambda}_0^{-1} - \frac{y_i^T y_i}{n_i} + \bar{y}\bar{y}^T \right)^T \right\}^T, \quad i \in \{1, \ldots, k\},$$

and the i-th group contribution to the tilted log-likelihood function $\ell(\varphi; t) = \sum_{i=1}^k \ell_i(\varphi_i; t)$ with

$$\ell_i(\varphi_i; t) \quad = \quad n_i \xi_i^T \left\{ t\bar{y}_i + (1-t)\bar{y} \right\} - \frac{n_i}{2} \text{tr} \left[\Lambda_i \left\{ \frac{t y_i^T y_i}{n_i} + (1-t) \left(\hat{\Lambda}_0^{-1} + \bar{y}\bar{y}^T \right) \right\} \right]$$

$$+\frac{n_i}{2}\log|\Lambda_i| - \frac{n_i}{2}\xi_i^T\Lambda_i^{-1}\xi_i.$$

The resulting maximum likelihood estimates from $\ell(\varphi;t)$ are $\hat{\mu}_i(t) = (1-t)\bar{y}+t\bar{y}_i$ and $\hat{\Lambda}_i(t)^{-1} = (1-t)\hat{\Lambda}_0^{-1}+t\hat{\Lambda}_i^{-1}+t(1-t)(\bar{y}_i-\bar{y})(\bar{y}_i-\bar{y})^T$. Hence, the saddlepoint approximation (2.5) along the line $s(t)$ is

$$h\{s(t);\psi\} = c\exp\left\{\sum_{i=1}^{k}\frac{n_i-p-2}{2}\log|\hat{\Lambda}_i^{-1}(t)|\right\},$$

where c is a normalizing constant. The value t_{sup} in (2.7) is the largest t for which every $\hat{\Lambda}_i(t)^{-1}$ is positive definite and has to be found iteratively. The following theorem gives conditions for the exactness of the directional p-value.

Theorem 3.2. *Assume that $p = p_n$ such that $n_i \geq p+2$ for all $n_i \geq 3, i \in \{1,\ldots,k\}$, with k fixed. Then, under the null hypothesis H_ψ (3.2), the directional p-value (2.7) is exactly uniformly distributed.*

The proof of Theorem 3.2 is similar to the one of Theorem 3.1 and is given in Appendix A.2.

4. Computational aspects

4.1. Determination of t_{sup}

The upper bound t_{sup} of the integrals in formula (2.7) is the largest value of t such that the maximum likelihood estimate $\hat{\Lambda}^{-1}(t)$ or $\hat{\Lambda}_i^{-1}(t), i \in \{1,\ldots,k\}$, is positive definite. Depending on the case, t_{sup} can be found analytically or approximated numerically. For instance, we can derive Lemma 4.1 and Lemma 4.2 to compute t_{sup} analytically for hypotheses (III)–(V) and (I), respectively. In particular, for hypotheses (III)–(V), we have that $t_{\text{sup}} = \{1-\nu_{(1)}\}^{-1}$, where $\nu_{(1)}$ is the smallest eigenvalue of $\hat{\Lambda}_0\hat{\Lambda}^{-1}$, while for hypothesis (I) $t_{\text{sup}} = \{1-\min_{1\leq i\leq k}\nu_{(1)}^i\}^{-1}$, where $\nu_{(1)}^i$ is the smallest eigenvalue of $\hat{\Lambda}_0\hat{\Lambda}_i^{-1}$. On the contrary, there is no available closed form for t_{sup} when testing hypotheses (II) and (VI). In such cases, we need to find t_{sup} by searching iteratively values of $t > 1$ until matrices $\hat{\Lambda}^{-1}(t)$ for hypothesis (VI) or $\hat{\Lambda}_i^{-1}(t), i \in \{1,\ldots,k\}$ for hypothesis (II) are no longer positive definite.

Lemma 4.1. *The estimator $\hat{\Lambda}^{-1}(t)$ is positive definite if and only if all elements $1-t+t\nu_l, l \in \{1,\ldots,p\}$, are positive, where ν_l are the eigenvalues of the matrix $\hat{\Lambda}_0\hat{\Lambda}^{-1}$. Specifically, $\hat{\Lambda}^{-1}(t)$ is positive definite in $t \in [0,\{1-\nu_{(1)}\}^{-1}]$, where $\nu_{(1)}$ is the smallest eigenvalue of $\hat{\Lambda}_0\hat{\Lambda}^{-1}$.*

The proof of Lemma 4.1 is given in Appendix A.3.

Lemma 4.2. *The estimator $\hat{\Lambda}_i^{-1}(t), i \in \{1,\ldots,k\}$, is positive definite if and only if all elements $1-t+t\nu_l^i, l \in \{1,\ldots,p\}$, are positive, where ν_l^i are the*

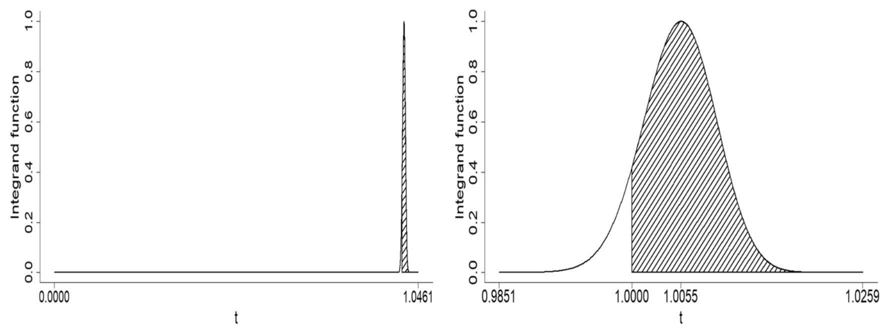

FIG 1. *Integrand function* $\exp\{\bar{g}(t;\psi)-\bar{g}(\hat{t};\psi)\}$ *in an example of the directional p-value to test the hypothesis (3.1). The* $n = 100$ *observations are sampled from a* $N_p(0_p, I_p)$ *distribution with* $p = 70$*. The left panel refers to the interval* $[0, t_{\text{sup}}] = [0, 1.046135]$*, the right panel to the interval* $[t_{\min}, t_{\max}] = [0.985117, 1.025937]$.

eigenvalues of the matrix $\hat{\Lambda}_0\hat{\Lambda}_i^{-1}$*. Specifically,* $\hat{\Lambda}_i^{-1}(t), i \in \{1,\ldots,k\}$*, are all positive definite in* $t \in \left[0, \{1-\min_{1\le i\le k}\nu_{(1)}^i\}^{-1}\right]$*, where* $\nu_{(1)}^i$ *is the smallest eigenvalue of* $\hat{\Lambda}_0\hat{\Lambda}_i^{-1}$.

The proof of Lemma 4.2 is given in Appendix A.4.

4.2. Numerical integration for the directional p-value

Let $g(t;\psi) = t^{d-1}h\{s(t);\psi\} = \exp[(d-1)\log t + \log h\{s(t);\psi\}] = \exp\{\bar{g}(t;\psi)\}$ be the integrand function in formula (2.7). In order to account for situations in which $g(t;\psi)$ is numerically too small or too large, we consider rescaling $\bar{g}(t;\psi)$ in the interval $[0, t_{\text{sup}}]$ using $\bar{g}(\hat{t};\psi) = \sup_{t\in[0,t_{\text{sup}}]}\bar{g}(t;\psi)$. The directional p-value can then be computed as

$$p(\psi) = \frac{\int_1^{t_{\text{sup}}}\exp\{\bar{g}(t;\psi)-\bar{g}(\hat{t};\psi)\}dt}{\int_0^{t_{\text{sup}}}\exp\{\bar{g}(t;\psi)-\bar{g}(\hat{t};\psi)\}dt}.$$

Moreover, when the dimension p is large, the integrand function often concentrates on a very small range, meaning that it is significantly different from zero in a very small interval around $\hat{t}$. Using an example of the hypothesis problem (3.1) as an illustration, in the left hand panel of Figure 1 the integrand function is plotted over the interval $[0, t_{\text{sup}}]$. We can observe that only for very few t values the function is appreciably different from zero. For a more accurate and efficient numerical integration, we can apply the Gauss-Hermite quadrature [21], and focus on a narrower integration interval $[t_{\min}, t_{\max}]$. The integrand function curve in such an interval is displayed in the right hand panel of Figure 1. Hence, the directional p-value can be well approximated by

$$p(\psi) \doteq \frac{\int_1^{t_{\max}}\exp\{\bar{g}(t;\psi)-\bar{g}(\hat{t};\psi)\}dt}{\int_{t_{\min}}^{t_{\max}}\exp\{\bar{g}(t;\psi)-\bar{g}(\hat{t};\psi)\}dt}. \tag{4.1}$$

Details on the implemention of the Gauss-Hermite quadrature (4.1) for the hypotheses considered in Section 3 and Supplementary Material S2 are described in Supplementary Material S1.3.

5. Simulation studies

5.1. Setup

The performance of the directional test for the hypotheses of Section 3 in the high dimensional multivariate normal framework is here assessed via Monte Carlo simulations based on 100 000 replications. The exact directional test is compared with the χ^2_d approximation for the log-likelihood ratio test, its Bartlett correction, two Skovgaard's modifications [29], and with the normal approximation for the test proposed by Jiang and Yang [18]. The six tests are evaluated in terms of empirical distribution, empirical distribution of the corresponding p-values, estimated size and power. Simulation results for the hypotheses (III)–(VI) are reported in Supplementary Material S4.

Samples of size $n_i, i \in \{1, \dots, k\}$, are generated from the p-variate standard normal distribution $N_p(0_p, \mathrm{I}_p)$ under the null hypothesis. For each simulation experiment, we show results for $k = 3$, $n_i = 100$ for all $i = 1, 2, 3$, and $p/n_i \in \{0.05, 0.1, 0.3, 0.5, 0.7, 0.9\}$. Additional results for different values of n_i and p/n_i are reported in Supplementary Material S5–S6. The various simulation setups are detailed below, partly taken from Jiang and Yang [18].

Hypothesis (I): testing the equality of covariance matrices in k normal distributions. When evaluating power, four settings are considered for the alternative hypothesis: (1) $\Lambda_1^{-1} = \mathrm{I}_p$, $\Lambda_2^{-1} = 1.21\mathrm{I}_p$ and $\Lambda_3^{-1} = 0.81\mathrm{I}_p$; (2) $\Lambda_1^{-1} = \mathrm{I}_p$, $\Lambda_2^{-1} = \Lambda_3^{-1} = \Lambda_1^{-1} + \delta(pn_i)^{-1/2}\mathrm{I}_p$; (3) $\Lambda_1^{-1} = \mathrm{I}_p$, $\Lambda_2^{-1} = \Lambda_3^{-1} = (1-\rho)\mathrm{I}_p + \rho 1_p 1_p^T$ with $\rho = \delta(pn_i)^{-1/2}$; (4) $\Lambda_1^{-1} = \mathrm{I}_p$, $\Lambda_2^{-1} = \Lambda_3^{-1} = \mathrm{diag}(\eta, 1_{p-1}^T)$ where $\eta \in \mathbb{R}^+$.

Hypothesis (II): testing the equality of k multivariate normal distributions. When evaluating power, four settings are considered for the alternative hypothesis: (1) $\mu_1 = 0_p$, $\mu_2 = \mu_3 = 0.1 \cdot 1_p$ and $\Lambda_1^{-1} = 0.51_p 1_p^T + 0.5\mathrm{I}_p$, $\Lambda_2^{-1} = 0.61_p 1_p^T + 0.4\mathrm{I}_p$, $\Lambda_3^{-1} = 0.51_p 1_p^T + 0.31\mathrm{I}_p$; (2) and (3) $\mu_1 = 0_p$, $\mu_2 = \mu_3 = \delta(pn_i)^{-1/2} 1_p$; (4) $\mu_1 = 0_p$, $\mu_2 = \mu_3 = \{10(pn_i)^{-1/2}, 0_{p-1}^T\}^T$, and the setup of covariance matrices of (2)–(4) as in Hypothesis (I).

In Supplementary Material S3 we report the empirical results for hypotheses (I)–(II) with large group values of $k \in \{30, 300\}$, which show that the accuracy of the directional p-value does not change.

5.2. Null distribution

The Monte Carlo simulations for the hypotheses (I) and (II) described in Section 3 are here illustrated. The Type I error at level $\alpha = 0.05$ based on the approximate null distribution of the various statistics is evaluated here. The empirical distribution of p-values for the six tests is examined by comparison

TABLE 1

Empirical probability of Type I error for hypotheses (I) and (II) for the directional test (DT), central limit theorem test (CLT), log-likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications [29] (Sko1 and Sko2, respectively) at the nominal level $\alpha = 0.05$, $n_1 = n_2 = n_3 = 100$

Hypothesis	p/n_i	DT	CLT	LRT	BC	Sko1	Sko2
(I)	0.05	0.050	0.078	0.062	0.050	0.048	0.048
	0.1	0.049	0.064	0.102	0.049	0.041	0.040
	0.3	0.051	0.057	0.950	0.067	0.010	0.006
	0.5	0.050	0.054	1.000	0.183	0.000	0.000
	0.7	0.050	0.054	1.000	0.865	0.000	0.000
	0.9	0.049	0.054	1.000	1.000	0.065	0.000
(II)	0.05	0.049	0.061	0.068	0.049	0.045	0.045
	0.1	0.048	0.055	0.115	0.049	0.037	0.036
	0.3	0.051	0.055	0.967	0.068	0.007	0.003
	0.5	0.050	0.053	1.000	0.192	0.000	0.000
	0.7	0.050	0.053	1.000	0.880	0.000	0.000
	0.9	0.049	0.053	1.000	1.000	0.032	0.000

with the Uniform$(0, 1)$ distribution in Supplementary Material S3. The limiting null distribution of the statistics is also compared with their corresponding chi-square or standard normal distribution in Supplementary Material S3.

Table 1 reports the empirical Type I error at the nominal level $\alpha = 0.05$ under the null hypotheses (I) and (II), from top to bottom respectively. The directional test exhibits an excellent performance in terms of empirical Type I error, indeed the directional p-value is exact up to simulation error for all different choices of p, as suggested by the theory in Section 3. Specifically, it is significantly better than that of the central limit theorem test of Jiang and Yang [18] which has a slightly liberal empirical Type I error. In addition, the four statistics with chi-square approximate distributions are not very accurate, and even remarkably unreliable with increasing p/n_i. This behavior confirms results in [14]: the chi-square approximation to the log-likelihood ratio statistic's distribution applies if and only if $p = o(n_i^{1/2})$ and that to its Bartlett correction if and only if $p = o(n_i^{2/3})$, $i \in \{1, \ldots, k\}$, which are both instances of low dimensional asymptotic regimes. There is no analogous theoretical result for Skovgaard's statistics [29], yet the numerical evidence suggests an intermediate condition between those of the log-likelihood ratio statistic and its Bartlett corrected version. We note that, somehow surprisingly, the empirical Type I error for one of the two Skovgaard's modifications improves for the largest value $p/n_i = 0.9$.

5.3. Empirical corrected power

The power of the tests considered for the hypotheses problems in the previous section are here investigated empirically for some alternative settings. In particular, four possible choices for μ_i and Λ_i^{-1} under the alternative hypotheses detailed in Section 5.1 are studied. The first alternative setting (1) is taken from Jiang and Qi [17], who extended the use of the central limit theorem test devel-

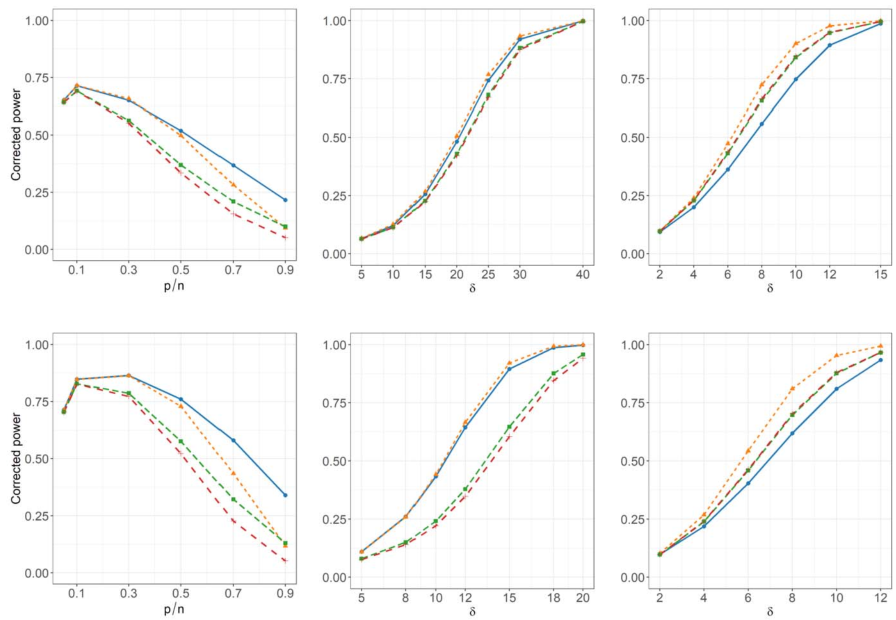

Fig 2. *Empirical corrected powers of four tests. The solid, dashed, long-dashed, and dot-dashed curves are the empirical power functions of the central limit theorem test, directional test and two Skovgaard's modifications [29], respectively. The top and bottom rows correspond to hypotheses (I) and (II), respectively; the left, middle and right columns correspond to alternative hypothesis settings (1), (2) and (3), respectively. The two right-most columns refer to the scenario with* $p/n_i = 0.3$.

oped in Jiang and Yang [18] to cases where p is very close to n (see Section 6 for further details). The second alternative setting (2) deals with situations where the Frobenius norm between the null and alternative parameters converges to zero as n_i goes to infinity. The third alternative setting (3) is based on the compound symmetry structure of the covariance matrix with correlation going to zero as n_i diverges, while only one group has the identity structure. The last alternative setting (4) is motivated by Jensen [16] and considers a situation where only one or two elements of the vector parameter differ between the null and alternative hypotheses. Due to space constraints, we report here results referred to the corrected power only. Corrected power is based on the corrected Type I error, which is the 5% quantile of the empirical p-values obtained under the null hypothesis, and is reported in Supplementary Material S3. This allows a fair comparison among the tests, since power is intended with a given significance level. However, it is important to remark that the directional p-value is the only approach that does not need a correction for the Type I error, being exact under the null hypothesis. The central limit theorem, log-likelihood ratio test and Bartlett correction have the same corrected power as they use the same test statistic W and result in different cutoff values for the corrected Type I error.

The left-most column of Figure 2 summarizes simulation results for the hypotheses (I) and (II) and $p/n_i \in \{0.05, 0.1, 0.3, 0.5, 0.7, 0.9\}$. The alternative setting for each hypothesis is the same as in Jiang and Qi [17], where the use of the central limit theorem test was recommended. The power of the directional test across the different ratios p/n_i is always greater than the nominal level 0.05; it is comparable with the corrected power of the central limit theorem test, log-likelihood ratio test and Bartlett correction when p is moderate, but it is lower otherwise. However, it must be taken into account that the log-likelihood ratio test and Bartlett correction do not control the Type I error when p is large, therefore their power is meaningless in such scenarios. Finally, Skovgaard's tests have uniformly the lowest corrected power.

We also investigate the local power, i.e. how large δ in the alternative settings of Section 5.1 needs to be so that the power can tend to 1. The middle and right columns of Figure 2 display the empirical local corrected power of the tests for various values of δ and ratio $p/n_i = 0.3$. Under the alternative setting (2), shown in the middle column, the power of the directional test is comparable or slightly superior to the corrected power of the central limit theorem test, and clearly higher than the corrected power of the Skovgaard's modifications. Under the alternative setting (3), shown in the right column, the directional test is the most powerful while the central limit theorem test has the worst power performance even after correction for Type I error.

Finally, Figures 3 and 4 analyse the empirical corrected power of the tests for various ratios p/n_i as η in Section 5.1 varies under the alternative setting (4) [16] of hypotheses (I) and (II), respectively. The directional test enjoys the best properties, proving to be particularly powerful with respect to its competitors when $p/n_i \geq 0.5$. Even in this case, the corrected power of the central limit theorem test is uniformly lowest.

6. Discussion

This work examines directional testing for hypotheses on a vector parameter of interest in p-variate normal distributions when n_i independent observations are available for the i-th group ($i = 1, \dots, k$) in the high dimensional regime with $p/n_i \to \kappa \in (0, 1]$ [4]. The construction of the directional test is based on the saddlepoint approximation to the density of the canonical sufficient statistic, which is found to be exact provided that each $n_i \geq p + 2$. The numerical results support the theoretical findings on the exact control of Type I error of the directional approach under these mild conditions. The simulation outcomes show that the directional test outperforms the omnibus tests which look in all directions of the parameter space for alternatives both when p is large and small relatively to n_i. Our formal derivations of the exactness of the underlying saddlepoint-type expansions provide also a theoretical ground to previous numerical findings obtained in high dimensional simulation settings [9, Example 5.3].

The six hypotheses testing problems considered here and in the Supplementary Material mainly come from [18] and [17]. Jiang and Qi [17] showed that

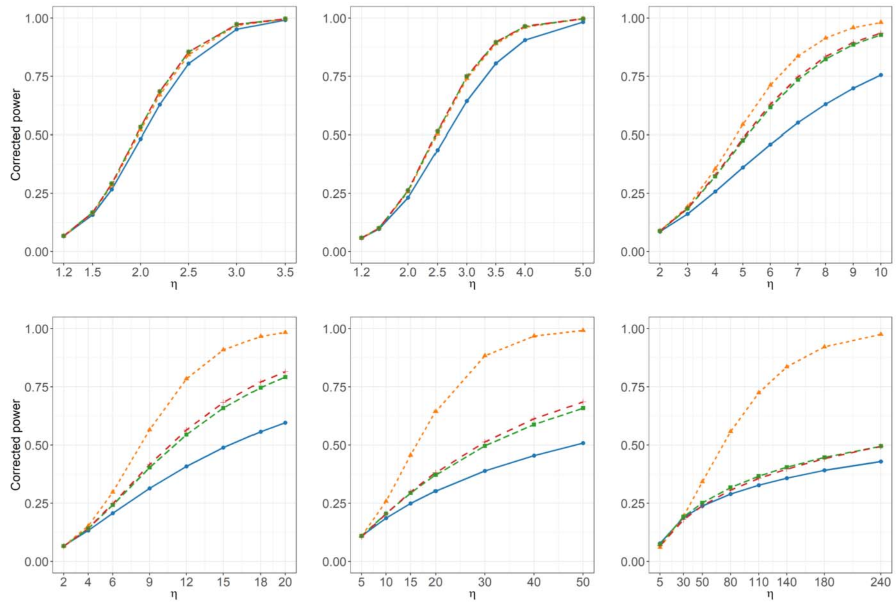

FIG 3. *Empirical corrected power of four tests for hypothesis (I) with different values of* η *and* p/n_i*. The solid, dashed, long-dashed, and dot-dashed curves are the empirical power functions of the central limit theorem test, directional test and two Skovgaard's modifications [29], respectively. The alternative setting (4) is given in Section 5.1. The six plots correspond to* $p/n_i \in \{0.05, 0.1, 0.3, 0.5, 0.7, 0.9\}$*, starting from top left and proceeding by row.*

the central limit theorem test works well when p is very close to n_i, assuming that $n_i > p + a$ for some constant $1 \leq a \leq 4$. In our Monte Carlo experiments, the central limit theorem result seems inaccurate when the dimension p is small, while the directional test is able to control exactly the Type I error for every value of p, provided that $n_i \geq p + 2$. The two tests have been compared empirically also in terms of corrected power for some alternative hypotheses. Similarly to the log-likelihood ratio test, the central limit theorem approach is an omnibus test, whereas the directional test measures the departure from the null hypothesis along the direction determined by the observed data point. In this respect, the latter is not constructed based on any kind of optimality [28] and its marginal power may change according to the specific alternative setting [16]. Nevertheless, our empirical results found not only that the power of the directional test does not need any correction for Type I error, but also that it is overall comparable with the corrected power of its main competitor.

The asymptotic theory for the directional test derived in this paper applies to linear exponential family models with hypotheses regarding linear functions of the canonical parameter, as in [9]. Similar results for tests regarding the mean vector and/or covariance matrix that cannot be expressed as hypotheses on linear function of the canonical parameter could be obtained under the more general framework in [13]. Further research might focus on deriving the proper-

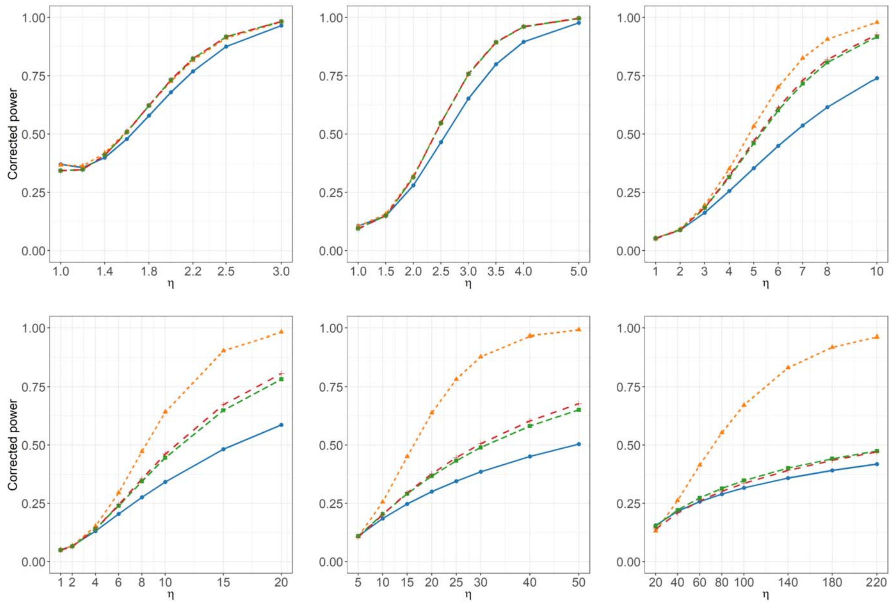

FIG 4. *Empirical corrected power of four tests for hypothesis (II) with different values of* η *and* p/n_i*. The solid, dashed, long-dashed, and dot-dashed curves are the empirical power functions of the central limit theorem test, directional test and two Skovgaard's modifications [29], respectively. The alternative setting (4) is given in Section 5.1. The six plots correspond to* $p/n_i \in \{0.05, 0.1, 0.3, 0.5, 0.7, 0.9\}$*, starting from top left and proceeding by row.*

ties of directional inference when p increases with n under the models considered previously by [13] and [23] for fixed p only.

In general, the accuracy of the directional p-value stems from the accuracy of the underlying saddlepoint approximation to the conditional density of the canonical sufficient statistic. For instance, in the high dimensional regime the directional test for the one-sample hypothesis on the normal mean vector $H_\psi : \mu = \mu_0$ is expected to behave as those seen here, since it was shown equivalent to the Hotelling's T^2 statistic [23]. In the multiple-sample case, preliminary results reveal that the high dimensional accuracy determined by the exactness of the directional p-value for testing the equality of the mean vectors is preserved only when assuming an identical covariance matrix for the k independent groups.

For multivariate continuous distributions there are other instances of exactness of the saddlepoint approximation [26, Section 10.9] where we can expect accuracy comparable with the high dimensional normal case. On the other hand, saddlepoint methods cannot be exact with discrete probability functions. In settings where the hypotheses are not linear in the canonical parameter or the saddlepoint approximation is not exact, ongoing simulation results and previous works [13, 9, Section 4.2] suggest that a low dimensional asymptotic regime where $p/n \to 0$, typically with $p = O(n^\alpha)$, $0 \leq \alpha < 1$, might be required for observing the same accuracy of the directional p-value found in this paper.

Our interest in this work lies exclusively in the high dimensional asymptotic regime because the maximum likelihood estimation is generally feasible in such situation, and so is the computation of the directional p-value. That being said, under particular sparsity assumptions [4, Section 4.4] it is possible that the maximum likelihood estimator exists even if $p > n$, thus in principle also the directional approach can be adopted in the ultra-high dimensional regime. As an example, consider hypothesis (III) in the Supplementary Material S2, testing the sphericity of the concentration matrix. The maximum likelihood estimate exists as long as n is larger than the maximal clique size of the corresponding graph [5]; hence, if the concentration matrix is assumed sparse enough directional inference can still be applied [10].

Appendix

A.1. Proof of Theorem 3.1

Proof. Suppose $y_{ij} \sim N_p(\mu_i, \Lambda_i^{-1})$, $i \in \{1, \dots, k\}$, $j \in \{1, \dots, n_i\}$. For each i-th group, the random variables y_{ij} are independent. Let $\bar{y}_i = n_i^{-1} 1_{n_i}^T y_i$, $\hat{\Lambda}_i^{-1} = n_i^{-1} y_i^T y_i - \bar{y}_i \bar{y}_i^T$. In this case, $\hat{\Lambda}_i^{-1} \sim W_p(n_i - 1, n_i^{-1}\Lambda_i^{-1})$, for $i \in \{1, \dots, k\}$. Due to the groups independence, the joint distribution of $\hat{\Lambda}_i$, $i \in \{1, \dots, k\}$, is the product of Wishart densities $\prod_{i=1}^k f(\hat{\Lambda}_i^{-1}; \Lambda_i^{-1})$ with

$$
\begin{aligned}
f(\hat{\Lambda}_i^{-1}; \Lambda_i^{-1}) \quad &= \quad \left(\frac{n_i}{2}\right)^{\frac{p(n_i-1)}{2}} \Gamma_p \left(\frac{n_i - 1}{2}\right)^{-1} \\
&\quad \times |\Lambda_i^{-1}|^{-\frac{n_i-1}{2}} \operatorname{etr}\left(-\frac{n_i}{2}\Lambda\hat{\Lambda}_i^{-1}\right) |\hat{\Lambda}_i^{-1}|^{\frac{n_i-p-2}{2}}. \qquad \text{(A.1)}
\end{aligned}
$$

The log-likelihood for the canonical parameter φ under the multivariate normal distribution is

$$
\ell(\varphi; s) = \sum_{i=1}^k \frac{n_i}{2} \log|\Lambda_i| - \frac{1}{2}\operatorname{tr}(\Lambda_i y_i^T y_i) + n\bar{y}_i^T \xi_i - \frac{n_i}{2}\xi_i^T \Lambda_i^{-1}\xi_i.
$$

In order to assess the exactness of the saddlepoint approximation, it is convenient to express the log-likelihood function for φ as

$$
\ell(\varphi; s) = \sum_{i=1}^k -\frac{n_i}{2}\log|\Lambda_i^{-1}| - \frac{n_i}{2}\operatorname{tr}(\Lambda_i\hat{\Lambda}_i^{-1}) - \frac{n_i}{2}(\bar{y}_i - \Lambda_i^{-1}\xi_i)^T\Lambda_i(\bar{y}_i - \Lambda_i^{-1}\xi_i). \tag{A.2}
$$

The maximum likelihood estimate $\hat{\varphi}$ has components $\{\hat{\xi}_i^T, \operatorname{vech}(\hat{\Lambda}_i)^T\}^T = \{\bar{y}_i^T\hat{\Lambda}_i, \operatorname{vech}(\hat{\Lambda}_i)^T\}^T$ and the constrained maximum likelihood estimate $\hat{\varphi}_\psi$ has components $\{\bar{y}_i^T\hat{\Lambda}_0, \operatorname{vech}(\hat{\Lambda}_0)^T\}^T$, $i \in \{1, \dots, k\}$. Evaluating (A.2) at the unconstrained and constrained maximum likelihood estimates for φ, the corresponding log-likelihood at $\hat{\varphi}$ and $\hat{\varphi}_\psi$ are $\ell(\hat{\varphi}; s) = 2^{-1}\sum_{i=1}^k -n_i \log|\hat{\Lambda}_i^{-1}| - n_i p$ and $\ell(\hat{\varphi}_\psi; s) = 2^{-1}\sum_{i=1}^k -n_i \ \log|\hat{\Lambda}_0^{-1}| - n_i \operatorname{tr}(\hat{\Lambda}_0\hat{\Lambda}_i^{-1})$, respectively. Then, under the null hypothesis H_ψ, and using the fact that $|J_{\varphi\varphi}(\hat{\varphi})|$ is proportional to $\prod_{i=1}^k |\hat{\Lambda}_i^{-1}|^{p+2}$ (see Supplementary Material S1.1), the saddlepoint approximation (2.5) is

$$h(s;\psi) = \prod_{i=1}^k c_i(\psi)\, |\hat{\Lambda}_0^{-1}|^{-\frac{n_i-1}{2}} \exp\left\{-\frac{n_i}{2}\mathrm{tr}(\hat{\Lambda}_0\hat{\Lambda}_i^{-1})\right\} |\hat{\Lambda}_i^{-1}|^{\frac{n_i-p-2}{2}}. \tag{A.3}$$

Formula (A.3) is the exact joint distribution of $\hat{\Lambda}_1^{-1}, \dots, \hat{\Lambda}_k^{-1}$, i.e. a product of Wishart densities with parameters $(n_i - 1, \hat{\Lambda}_0^{-1})$ given in (A.1), if $\hat{\Lambda}_0^{-1}$ is considered as fixed. In particular, we have $c_i(\psi) = c_i = (n_i/2)^{p(n_i-1)/2}\, \Gamma_p\left\{(n_i-1)/2\right\}^{-1}$. It is indeed correct to fix $\hat{\Lambda}_0^{-1}$ because when considering the saddlepoint approximation density along the line $s(t)$, by construction the constrained maximum likelihood estimates of Λ_i^{-1} is fixed and equal to the observed value $\hat{\Lambda}_0^{-1}$.

When we consider the density of $s(t)$, we just need to replace $\hat{\Lambda}_i^{-1}$ in (A.3) with $\hat{\Lambda}_i^{-1}(t)$, i.e. the value which maximizes $\ell\left\{\varphi; s(t)\right\}$. Then, given that $\hat{\Lambda}_i^{-1}(t) = (1-t)\hat{\Lambda}_0^{-1} + t\hat{\Lambda}_i^{-1}$ and the groups are independent, under H_ψ we have

$$\begin{aligned} h\{s(t);\psi\} &= \prod_{i=1}^k c_i\, |\hat{\Lambda}_0^{-1}|^{-\frac{n_i-1}{2}} \exp\left[-\frac{n_i}{2}\mathrm{tr}\{\hat{\Lambda}_0\hat{\Lambda}_i(t)^{-1}\}\right] |\hat{\Lambda}_i(t)^{-1}|^{\frac{n_i-p-2}{2}} \\ &\propto \exp\left\{\sum_{i=1}^k \frac{n_i-p-2}{2}\log|\hat{\Lambda}_i^{-1}(t)|\right\}, \end{aligned}$$

where we have used the equality $n^{-1}\mathrm{tr}(\sum_{i=1}^k n_i\hat{\Lambda}_0\hat{\Lambda}_i^{-1}) = p$ with $n = \sum_{i=1}^k n_i$. Since the saddlepoint approximation $h\{s(t);\psi\}$ is exact, apart from the normalizing constant, the integral in the denominator of the directional p-value (2.7) is just the normalizing constant of the conditional distribution of $||s||$ given the direction $s/||s||$. Therefore, the directional p-value is the exact probability of $||s|| > ||s^0||$ given the direction $s/||s||$ under the null hypothesis, and is thus exactly uniformly distributed. □

A.2. Proof of Theorem 3.2

Proof. We know that $\bar{y}_i \sim N_p(\mu_i, n_i^{-1}\Lambda_i^{-1})$ and $\hat{\Lambda}_i^{-1} \sim W_p(n_i - 1, n_i^{-1}\Lambda_i^{-1})$, $i \in \{1, \dots, k\}$. In addition, $\bar{y}_i$ and $\hat{\Lambda}_i$ are independent [25, Section 10.8], thus the joint distribution of $\bar{y}_i$ and $\hat{\Lambda}_i$ takes the form $\prod_{i=1}^k f(\bar{y}_i;\mu_i,\Lambda_i^{-1}) f(\hat{\Lambda}_i^{-1};\Lambda_i^{-1})$ with

$$\begin{aligned} f(\bar{y}_i;\mu_i,\Lambda_i^{-1}) &= (2\pi)^{-\frac{p}{2}}|\Lambda_i^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{n_i}{2}(\bar{y}_i-\mu_i)^T\Lambda_i(\bar{y}_i-\mu_i)\right\}, \\ f(\hat{\Lambda}_i^{-1};\Lambda_i^{-1}) &= (n_i/2)^{\frac{p(n_i-1)}{2}}\Gamma_p\left(\frac{n_i-1}{2}\right)^{-1} \\ &\quad \times|\Lambda_i^{-1}|^{-\frac{n_i-1}{2}}\mathrm{etr}\left(-\frac{n_i}{2}\Lambda\hat{\Lambda}_i^{-1}\right)|\hat{\Lambda}_i^{-1}|^{\frac{n_i-p-2}{2}}. \end{aligned}$$

Similarly to the proof of Theorem 3.1, we can easily obtain the saddlepoint approximation to the density of the sufficient statistic s as

$$\begin{aligned} h(s;\psi) &= \prod_{i=1}^{k} c_{i1}\,|\hat{\Lambda}_0^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{n_i}{2}(\bar{y}_i-\hat{\mu}_0)^T\hat{\Lambda}_0(\bar{y}_i-\hat{\mu}_0)\right\} \\ &\quad \times c_{i2}\,|\hat{\Lambda}_0^{-1}|^{-\frac{n_i-1}{2}} \exp\left\{-\frac{n_i}{2}\mathrm{tr}(\hat{\Lambda}_0\hat{\Lambda}_i^{-1})\right\}|\hat{\Lambda}_i^{-1}|^{\frac{n_i-p-2}{2}}. \end{aligned} \tag{A.4}$$

Expression (A.4) equals the exact joint distribution of $\bar{y}_1,\dots,\bar{y}_k$ and $\hat{\Lambda}_1^{-1},\dots,\hat{\Lambda}_k^{-1}$ with $c_{i1}=(2\pi)^{-p/2}$, $c_{i2}=\left(\frac{n_i}{2}\right)^{p(n_i-1)/2}\Gamma_p\left(\frac{n_i-1}{2}\right)^{-1}$ and with fixed $\hat{\mu}_0$ and $\hat{\Lambda}_0^{-1}$. It is indeed correct to consider $\hat{\mu}_0$ and $\hat{\Lambda}_0^{-1}$ as fixed since the constrained maximum likelihood estimate is fixed and equal to the observed value when considering the saddlepoint approximation along the line $s(t)$ under H_ψ. In such case we have $\hat{\mu}_i(t)=(1-t)\hat{\mu}_0+t\bar{y}_i$ and $\hat{\Lambda}_i(t)^{-1}=(1-t)\hat{\Lambda}_0^{-1}+t\hat{\Lambda}_i^{-1}+t(1-t)(\bar{y}-\bar{y}_i)(\bar{y}-\bar{y}_i)^T$ where $\hat{\mu}_0=\bar{y}$ and $\hat{\Lambda}_0^{-1}=n^{-1}(A+B)$ (see Section 3 for more details). Then, the saddlepoint approximation for the distribution of $s(t)$ under H_ψ follows from (A.4), and is equal to

$$\begin{aligned} h\{s(t);\psi\} &= \prod_{i=1}^{k} c_{1i}\,|\hat{\Lambda}_0^{-1}|^{-\frac{1}{2}} \exp\left[-\frac{n_i}{2}\{\hat{\mu}_i(t)-\hat{\mu}_0\}^T\hat{\Lambda}_0\{\hat{\mu}_i(t)-\hat{\mu}_0\}\right] \\ &\quad \times c_{i2}\,|\hat{\Lambda}_0^{-1}|^{-\frac{n_i-1}{2}} \exp\left[-\frac{n_i}{2}\mathrm{tr}\{\hat{\Lambda}_0\hat{\Lambda}_i(t)^{-1}\}\right]|\hat{\Lambda}_i(t)^{-1}|^{\frac{n_i-p-2}{2}} \\ &\propto \exp\left\{\sum_{i=1}^{k}\frac{n_i-p-2}{2}\log|\hat{\Lambda}_i(t)^{-1}|\right\}. \end{aligned}$$

The remaining part of the proof is similar to that for hypothesis (I). It follows then that the directional p-value is exactly uniformly distributed under the null hypothesis H_ψ. □

A.3. Proof of Lemma 4.1

Proof. If $t\in[0,1]$ the result is straightforward, because a convex combination of positive definite matrices is positive definite. Indeed, for all $x\in\mathbb{R}^p$, $x\neq 0$, $x^T\hat{\Lambda}^{-1}(t)x=(1-t)x^T\hat{\Lambda}_0^{-1}x+tx^T\hat{\Lambda}^{-1}x>0$ since $1-t\geq 0$ and $t\geq 0$. Let us focus on the case $t>1$. Consider a square root B_0 of $\hat{\Lambda}_0^{-1}$ such that $\hat{\Lambda}_0^{-1}=B_0B_0^T=B_0^TB_0$, which always exists if $\hat{\Lambda}_0^{-1}$ is positive definite. Hence, the estimator $\hat{\Lambda}^{-1}(t)=(1-t)\hat{\Lambda}_0^{-1}+t\hat{\Lambda}^{-1}$ can be rewritten as

$$\hat{\Lambda}^{-1}(t)=B_0^T\left\{(1-t)\mathrm{I}_p+t(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}\right\}B_0.$$

The matrix $(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}$ is symmetric since $\hat{\Lambda}^{-1}$ is symmetric. Moreover, according to the eigen decomposition [22, Theorem 1.13], there exists an orthogonal $p\times p$ matrix P whose columns are eigenvectors of $(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}$ and a diagonal matrix Q whose diagonal elements are the eigenvalues of $(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}$, such that $(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1} = PQP^T$. Therefore, we have $\hat{\Lambda}^{-1}(t) = B_0^T P\{(1-t)\mathrm{I}_p + tQ\} P^T B_0$. Lemma 4.1 can then be proved through the following three steps.

Step 1: checking that $\hat{\Lambda}^{-1}(t)$ is positive definite is equivalent to checking that $(1-t)\mathrm{I}_p + tQ$ is positive definite. Indeed, for all $x \in \mathbb{R}^p$, $x \neq 0$, then

$$\begin{aligned} x^T\hat{\Lambda}^{-1}(t)x &= x^T B_0^T P\{(1-t)\mathrm{I}_p + tQ\} P^T B_0 x \\ &= \tilde{x}^T\{(1-t)\mathrm{I}_p + tQ\}\tilde{x} > 0, \end{aligned}$$

where $\tilde{x} = P^T B_0 x$, with $\tilde{x} \neq 0$ if $x \neq 0$.

Step 2: checking that $(1-t)\mathrm{I}_p + tQ$ is positive definite is equivalent to checking that all diagonal elements of the diagonal matrix $(1-t)\mathrm{I}_p + tQ = \mathrm{diag}(1 - t + t\nu_l)$ are positive, where ν_l, $l \in \{1,\dots,p\}$, are the eigenvalues of the matrix $(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}$. We now need to find out the largest t such that $1 - t + t\nu_l > 0, l \in \{1,\dots,p\}$:

- if $1 - \nu_{(1)} > 0$, where $\nu_{(1)}$ is the smallest eigenvalue of $(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}$, then $t < \frac{1}{1-\nu_l} \leq \frac{1}{1-\nu_{(1)}}$;
- if $1 - \nu_{(1)} \leq 0$, then $t > \frac{1}{1-\nu_l}$ as $\frac{1}{1-\nu_l} < 0$, and this condition holds true $\forall\, t \in \mathbb{R}^+$.

Step 3: The last step consists of checking that the eigenvalues $\nu_1,\dots,\nu_p$ of $(B_0^T)^{-1}\hat{\Lambda}^{-1}\ B_0^{-1}$ are the same as those of $\hat{\Lambda}_0\hat{\Lambda}^{-1}$, which is equivalent to show that the matrices $(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}$ and $\hat{\Lambda}_0\hat{\Lambda}^{-1}$ are similar. In addition, $\hat{\Lambda}_0^{-1} = B_0^T B_0$, given the invertible matrix B_0 such that

$$B_0^{-1}(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}B_0 = B_0^{-1}(B_0^T)^{-1}\hat{\Lambda}^{-1} = \hat{\Lambda}_0\hat{\Lambda}^{-1}.$$

According to matrix similarity, $(B_0^T)^{-1}\hat{\Lambda}^{-1}B_0^{-1}$ and $\hat{\Lambda}_0\hat{\Lambda}^{-1}$ are similar and therefore have the same eigenvalues.

Finally, since $\hat{\Lambda}_0\hat{\Lambda}^{-1}$ is positive definite and $\mathrm{tr}(\hat{\Lambda}_0\hat{\Lambda}^{-1}) = p$, the smallest eigenvalue $\nu_{(1)}$ must be lower than 1. Therefore, $\hat{\Lambda}^{-1}(t)$ is positive definite in $t \in [0, \{1-\nu_{(1)}\}^{-1}]$. □

A.4. Proof of Lemma 4.2

Proof. Based on the proof of Lemma 4.1, it is easy to show that $\hat{\Lambda}_i^{-1}(t)$ for all $i \in \{1,\dots,k\}$, is positive definite if and only if all elements $1 - t + t\nu_l^i > 0$, where $\nu_l^i, l \in \{1,\dots,p\}$, are the eigenvalues of the matrix $\hat{\Lambda}_0\hat{\Lambda}_i^{-1}$, $i \in \{1,\dots,k\}$. Since $\hat{\Lambda}_0\hat{\Lambda}_i^{-1}$ are positive definite and $\mathrm{tr}(\hat{\Lambda}_0\hat{\Lambda}_i^{-1}) = p$ for all $i \in \{1,\dots,k\}$, there exists at least one of the $\nu_{(1)}^i$ lower than 1, where $\nu_{(1)}^i$ denotes the smallest eigenvalue of $\hat{\Lambda}_0\hat{\Lambda}_i^{-1}$. In this respect, $\hat{\Lambda}_i^{-1}(t)$, $\forall i \in \{1,\dots,k\}$, are positive definite in $t \in \left[0, \{1 - \min_{1\leq i\leq k}\nu_{(1)}^i\}^{-1}\right]$. □

Acknowledgments

We thank two anonymous Referees and the Associate Editor for their constructive comments which enabled us to improve our work.

Supplementary Material

Supplementary material to Directional testing for high dimensional multivariate normal distributions
(doi: 10.1214/22-EJS2089SUPP; .pdf). Supplementary material includes some auxiliary computational results, theoretical results for the hypotheses (III)-(VI) in Section 3 and extensive additional simulation studies.

References

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley.
[2] Barndorff-Nielsen, O. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322. MR0855891
[3] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. London Ser. A* **160** 268–282.
[4] Battey, H. and Cox, D. (2022). Some perspectives on inference in high dimensions. *Statistical Science* **37** 110–122.
[5] Buhl, S. L. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scandinavian Journal of Statistics* **20** 263–270.
[6] Cheah, P. K., Fraser, D. A. S. and Reid, N. (1994). Multiparameter testing in exponential models: Third order approximations from likelihood. *Biometrika* **81** 271–278.
[7] Cordeiro, G. and Cribari-Neto, F. (2014). *An Introduction to Bartlett Correction and Bias Reduction*. Springer-Verlag.
[8] Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.
[9] Davison, A. C., Fraser, D. A. S., Reid, N. and Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *J. Amer. Statist. Assoc.* **109** 302–314.
[10] Di Caterina, C., Reid, N. and Sartori, N. (2021). Directional tests in Gaussian graphical models. arXiv 2103.15394.
[11] Fan, Y., Demirkaya, E. and Lv, J. (2019). Nonuniformity of p-values can occur early in diverging dimensions. *J. Mach. Learn. Res.* **20** 1–33. MR3960931
[12] Fraser, D. A. S. and Massam, H. (1985). Conical tests: Observed levels of significance and confidence regions. *Statistische Hefte* **26** 1–17.
[13] Fraser, D. A. S., Reid, N. and Sartori, N. (2016). Accurate directional inference for vector parameters. *Biometrika* **103** 625–639.

[14] He, Y., Meng, B., Zeng, Z. and Xu, G. (2021). On the phase transition of Wilks' phenomenon. *Biometrika* **108** 741–748.
[15] Huang, C., Di Caterina, C. and Sartori, N. (2022). Supplement to "Directional testing for high dimensional multivariate normal distributions." DOI: 10.1214/22-EJS2089SUPP.
[16] Jensen, J. L. (2021). On the use of saddlepoint approximations in high dimensional inference. *Sankhya A* **83** 379–392. MR4227217
[17] Jiang, T. and Qi, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.* **42** 988–1009.
[18] Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074.
[19] Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.
[20] Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43** 295–303.
[21] Liu, Q. and Pierce, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika* **81** 624–629.
[22] Magnus, J. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. Wiley.
[23] Mccormack, A., Reid, N., Sartori, N. and Theivendran, S. A. (2019). A directional look at F-tests. *Canad. J. Statist.* **47** 619–627. MR4035792
[24] McCullagh, P. (2018). *Tensor Methods in Statistics*, 2nd ed. Dover Publications.
[25] Muirhead, R. J. (2009). *Aspects of Multivariate Statistical Theory*. Wiley.
[26] Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference from a Neo-Fisherian Perspective*. World Scientific Press.
[27] Severini, T. A. (2001). *Likelihood Methods in Statistics*. Oxford University Press.
[28] Skovgaard, I. M. (1988). Saddlepoint expansions for directional test probabilities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **50** 269–280.
[29] Skovgaard, I. M. (2001). Likelihood asymptotics. *Scand. J. Stat.* **28** 3–32.
[30] Sur, P. and Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci.* **116** 14516–14525.
[31] Sur, P., Chen, Y. and Candès, E. J. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probab. Theory Related Fields* **175** 487–558. MR4009715
[32] Tang, Y. and Reid, N. (2020). Modified likelihood root in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 1349–1369.
[33] Tang, Y. and Reid, N. (2021). Laplace and saddlepoint approximations in high dimensions. arXiv 2107.10885.