The Annals of Applied Probability
2023, Vol. 33, No. 6A, 4272–4321
https://doi.org/10.1214/22-AAP1917
© Institute of Mathematical Statistics, 2023

CONTIGUITY UNDER HIGH-DIMENSIONAL GAUSSIANITY WITH APPLICATIONS TO COVARIANCE TESTING

BY QIYANG HAN[1,a], TIEFENG JIANG[2,b] AND YANDI SHEN[3,c]

[1]*Department of Statistics, Rutgers University,* [a]*qh85@stat.rutgers.edu*
[2]*School of Statistics, University of Minnesota,* [b]*jiang040@umn.edu*
[3]*Department of Statistics, University of Chicago,* [c]*ydshen@uchicago.edu*

Le Cam's third/contiguity lemma is a fundamental probabilistic tool to compute the limiting distribution of a given statistic T_n under a non-null sequence of probability measures $\{Q_n\}$, provided its limiting distribution under a null sequence $\{P_n\}$ is available, and the log likelihood ratio $\{\log(\mathrm{d}Q_n/\mathrm{d}P_n)\}$ has a distributional limit. Despite its wide-spread applications to low-dimensional statistical problems, the stringent requirement of Le Cam's third/contiguity lemma on the distributional limit of the log likelihood ratio makes it challenging, or even impossible to use in many modern high-dimensional statistical problems.

This paper provides a nonasymptotic analogue of Le Cam's third/contiguity lemma under high-dimensional normal populations. Our contiguity method is particularly compatible with sufficiently regular statistics T_n: the regularity of T_n effectively reduces both the problems of (i) obtaining a null (Gaussian) limit distribution and of (ii) verifying our new quantitative contiguity condition, to those of derivative calculations and moment bounding exercises. More important, our method bypasses the need to understand the precise behavior of the log likelihood ratio, and therefore possibly works even when it necessarily fails to stabilize—a regime beyond the reach of classical contiguity methods.

As a demonstration of the scope of our new contiguity method, we obtain asymptotically exact power formulae for a number of widely used high-dimensional covariance tests, including the likelihood ratio tests and trace tests, that hold uniformly over all possible alternative covariance under mild growth conditions on the dimension-to-sample ratio. These new results go much beyond the scope of previous available case-specific techniques, and exhibit new phenomenon regarding the behavior of these important class of covariance tests.

CONTENTS

1. Introduction 4273
   1.1. Le Cam's third/contiguity lemma: A review 4273
   1.2. An analogue to Le Cam's third/contiguity lemma under high-dimensional Gaussianity 4275
   1.3. Power formula of tests with high-dimensional normal population 4276
   1.4. Two concrete applications of (1.13) 4277
   1.5. Organization 4278
   1.6. Notation 4279
2. Contiguity under high-dimensional Gaussianity 4279
   2.1. The formal description of (1.7)–(1.8) 4279
   2.2. Power formula for tests with high-dimensional normal population 4281
   2.3. Proof of Theorem 2.1 4282
   2.4. Proof of Corollary 2.2 4284
3. Applications to high-dimensional covariance testing 4284

Received July 2021; revised June 2022.
MSC2020 subject classifications. 60F17, 62E17.
Key words and phrases. Contiguity, covariance test, power analysis, Poincaré inequalities.

3.1. Testing identity $\Sigma = I$ . . . 4284
3.1.1. LRT . . . 4285
3.1.2. Ledoit–Nagao–Wolf's test . . . 4286
3.2. Testing sphericity $\Sigma = \lambda I$ . . . 4287
3.2.1. LRT . . . 4287
3.2.2. John's test . . . 4289
3.3. Case study: Spiked covariance models . . . 4290
4. Some spectral estimates . . . 4292
5. Proofs for Section 3.1 (testing identity) . . . 4296
5.1. Proofs for Section 3.1.1 (LRT) . . . 4296
5.1.1. Evaluation of derivatives . . . 4296
5.1.2. Normal approximation . . . 4297
5.1.3. Contiguity . . . 4299
5.2. Proofs for Section 3.1.2 (Ledoit–Nagao–Wolf's test) . . . 4300
5.2.1. Evaluation of derivatives . . . 4300
5.2.2. Normal approximation . . . 4301
5.2.3. Contiguity . . . 4303
5.2.4. Completing the proof for power expansion . . . 4305
6. Proofs for Section 3.2 (testing sphericity) . . . 4306
6.1. Proofs for Section 3.2.1 (LRT) . . . 4306
6.1.1. Evaluation of derivatives . . . 4306
6.1.2. Normal approximation . . . 4307
6.1.3. Contiguity . . . 4307
6.1.4. Completing the proof for power expansion . . . 4310
6.2. Proofs for Section 3.2.2 (John's test) . . . 4311
6.2.1. Evaluation of derivatives . . . 4311
6.2.2. Normal approximation . . . 4312
6.2.3. Contiguity . . . 4314
6.2.4. Completing of the proof for power expansion . . . 4320
Acknowledgments . . . 4320
Funding . . . 4320
Supplementary Material . . . 4320
References . . . 4320

1. Introduction.

1.1. *Le Cam's third/contiguity lemma*: *A review.* For each $n \in \mathbb{N}$, let $(\Omega_n, \mathcal{B}_n)$ be a measurable space, on which a real-valued random variable T_n, and a pair of probability measures (P_n, Q_n) are defined. Here n is a generic index for asymptotics, which is usually related to "sample size" in statistics literature. Le Cam's third/contiguity lemma [21], which is essentially an asymptotic change of variable formula, computes the limiting law of $\{T_n\}$ under the laws of $\{Q_n\}$, provided that its limiting law under the laws of $\{P_n\}$ can be computed, and the distributions of the log-likelihood ratio $\{\log(\mathrm{d}Q_n/\mathrm{d}P_n)\}$ under the laws of $\{P_n\}$ can be precisely evaluated. The most common form of Le Cam's third/contiguity lemma states the following: Suppose that $\{T_n\}$ is asymptotically normal under the laws of $\{P_n\}$, that is,

$$\frac{T_n - m_{P_n}}{\sigma_{P_n}} \overset{P_n}{\rightsquigarrow} \mathcal{N}(0, 1) \tag{1.1}$$

for some $\{m_{P_n} \in \mathbb{R}\}$ and $\{\sigma_{P_n} > 0\}$. Here and below we use $\rightsquigarrow$ to denote weak convergence. Then the limiting law of the normalized random variable $\{(T_n - m_{P_n})/\sigma_{P_n}\}$ under the laws of $\{Q_n\}$ can be computed as

$$\frac{T_n - m_{P_n}}{\sigma_{P_n}} \overset{Q_n}{\rightsquigarrow} \mathcal{N}(\tau, 1), \tag{1.2}$$

provided that $\sigma_{P_n} \to \sigma_P$ for some $\sigma_P > 0$ and Le Cam's contiguity condition

$$\begin{pmatrix} T_n - m_{P_n} \\ \log(\mathrm{d}Q_n/\mathrm{d}P_n) \end{pmatrix} \overset{P_n}{\rightsquigarrow} \mathcal{N}\left(\begin{pmatrix} 0 \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} \sigma_P^2 & \tau\sigma_P \\ \tau\sigma_P & \sigma^2 \end{pmatrix} \right) \tag{1.3}$$

holds for some $\sigma^2 > 0$. The condition that $\sigma_{P_n} \to \sigma_P$ can typically be ensured by rescaling T_n appropriately, so the real nontrivial condition is the asymptotic distributional expansions of the statistic T_n and the log likelihood ratio in (1.3). We refer the reader to [30], Chapter 6, for an in-depth treatment of Le Cam's contiguity theory.

Le Cam's third/contiguity lemma, as stated above, has played a fundamental role in several major developments of estimation and testing theory in mathematical statistics. For instance, convolution and asymptotic minimax theorems for parametric models, which quantify the fundamental information theoretic limits of any regular statistical estimators, are proved with an essential use of Le Cam's third/contiguity lemma. In testing theory, Le Cam's third lemma also facilitates the computation of exact "power function" (to be defined ahead) of any statistical tests. See, for example, [30], Chapters 7, 8, 15, for a textbook treatment on these by-now classical topics.

From (1.1)–(1.3), it is clear that a successful application of Le Cam's third/contiguity lemma relies heavily on two crucial ingredients: (A) a central limit theorem (CLT) for $\{T_n\}$ under the laws of $\{P_n\}$ in (1.1), and (B) an easy-to-handle log likelihood ratio $\log(\mathrm{d}Q_n/\mathrm{d}P_n)$. This is indeed fairly straightforward in classical models. For instance, a standard application is the study of maximum likelihood estimator (MLE) in parametric statistical models: Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables from a probability distribution P_{θ_0} in a parametric class $\mathcal{P} \equiv \{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$, where Θ is an open set in $\mathbb{R}$, and $T_n \equiv \sqrt{n}(\widehat{\theta} - \theta_0)$ where $\widehat{\theta} \equiv \widehat{\theta}(X_1, \ldots, X_n)$ is the MLE for θ_0 within $\mathcal{P}$. We wish to compute the limiting law of $\{T_n\}$ under the local laws $\{P_{\theta_0 + h/\sqrt{n}}\}$ for a fixed $h \in \mathbb{R}$. To apply Le Cam's third lemma, we take $\{P_n \equiv P_{\theta_0}\}$ and $\{Q_n \equiv P_{\theta_0 + h/\sqrt{n}}\}$. Now part A can be easily tackled—it is classical knowledge that when $\mathcal{P}$ is "smooth enough", then a CLT for $\{T_n\}$ holds with $m_{P_n} = 0$ and $\sigma_{P_n} = (I_{\theta_0})^{-1/2}$ where I_{θ_0} is the Fisher information of $\mathcal{P}$ at θ_0. Part B can also be handled easily, for instance, a direct Taylor expansion of the log likelihood for sufficiently smooth $\mathcal{P}$[1] concludes, after some calculations, that $\tau = h I_\theta^{1/2}$. The limiting law of $\{T_n = \sqrt{n}(\widehat{\theta} - \theta_0)\}$ under $\{Q_n = P_{\theta_0 + h/\sqrt{n}}\}$ now follows immediately from (1.2).

Although applied in a wide range of contexts with great success in classical low-dimensional statistical problems, Le Cam's third/contiguity lemma faces a key challenge in its stringent requirement for the exact distributional behavior of the log likelihood $\{\log(\mathrm{d}Q_n/\mathrm{d}P_n)\}$ under $\{P_n\}$ as in (1.3). In many modern high-dimensional statistical applications, the distributional expansion of the log likelihood can sometimes be extremely difficult, or even impossible to handle. One leading example is given by high-dimensional normal populations with spiked covariance, that is, $P_n =$ (resp. Q_n) law of n i.i.d. observations of p-dimensional Gaussian vectors with covariance I_p (resp. $I_p + \Delta_p$), where $\Delta_p = \sum_{\ell=1}^r h_\ell v_\ell v_\ell^\top$ is a (fixed) rank r perturbation matrix with $h_\ell \geq 0$ and $\|v_\ell\| = 1$ for all $1 \leq \ell \leq r$. The log likelihood $\{\log(\mathrm{d}Q_n/\mathrm{d}P_n)\}$ can be computed both over the original data and over the maximal invariant (i.e., the eigenvalues of the sample covariance):

- For the log likelihood over the original data, fairly straightforward calculations show that, already in the simplest possible rank one case $r = 1$, $\{\log(\mathrm{d}Q_n/\mathrm{d}P_n)\}$ can stochastically stabilize under $\{P_n\}$ only if $h_1 = \mathcal{O}(n^{-1/2})$, a regime of almost no practical relevance in high-dimensional settings, say, $\lim(p/n) = y \in (0, \infty)$.

[1]The celebrated local asymptotic normality (LAN) (cf. [21]) condition can also be used for this purpose to weaken smoothness requirements.

- For the log likelihood over the maximal invariant, [25, 26] showed that a distributional limit of $\{\log(\mathrm{d}Q_n/\mathrm{d}P_n)\}$ exists in the subcritical regime below the Baik–Ben Arous–Péché (BBP) phase transition [3] $\max_{1\le\ell\le r} h_\ell < \sqrt{y}$, where $y \equiv \lim(p/n)$. The hard threshold $\sqrt{y}$ is necessary as the weak limit does not exist when $\max_{1\le\ell\le r} h_\ell > \sqrt{y}$.

Consequently, even for the fixed-rank spiked covariance alternatives, Le Cam's contiguity condition (1.3) already fails to obtain nonnull distributions of the type (1.2) except for a highly restrictive set of alternatives. Fundamentally, such restrictions arise as it is more than necessary and in fact far too strong to require a weak limit, or even only stochastic stabilization of the log likelihood ratio as in (1.3), for the purpose of computing nonnull distributions for a given statistics, in particular in high-dimensional settings.

1.2. *An analogue to Le Cam's third/contiguity lemma under high-dimensional Gaussianity.* In this paper, we establish a nonasymptotic analogue of Le Cam's third/contiguity lemma in the form of (1.1)–(1.2), without the requirement for an exact distributional evaluation or even stochastic stabilization of the limiting log likelihood ratio as in (1.3), in the setting where T_n is a sufficient "regular" function of n i.i.d. observations of a p-dimensional normal distribution.

Formally, let $X_1, \dots, X_n$ be i.i.d. samples from a p-dimensional normal distribution $\mathcal{N}_p(\mu, \Sigma)$, where $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$ with $\mathcal{M}_p$ denoting the set of all $p \times p$ covariance matrices. Let $X = [X_1, \dots, X_n]^\top \in \mathbb{R}^{n\times p}$ be the data matrix. We will be concerned with the statistic $T_n = T(X)$ for some $T : \mathbb{R}^{n\times p} \to \mathbb{R}$ living in the Sobolev space $W^{1,2}(\gamma_{n\times p})$, where $\gamma_{n\times p}$ is the standard Gaussian measure on $\mathbb{R}^{n\times p}$ (precise definitions can be found in Section 1.6). Let (throughout the paper we use the symbol $\equiv$ for definition)

$$m_{(\mu,\Sigma)} \equiv \mathbb{E}_{(\mu,\Sigma)} T(X), \qquad \sigma^2_{(\mu,\Sigma)} \equiv \mathrm{Var}_{(\mu,\Sigma)}\big(T(X)\big) \tag{1.4}$$

be the mean and variance of $T(X)$ under $\mathcal{N}_p(\mu, \Sigma)$, respectively. We always assume that the two quantities in (1.4) are finite. In a similar spirit, we use the subscript (μ, Σ) in $\mathbb{E}_{(\mu,\Sigma)}$ and other probabilistic notation to indicate that the evaluation is under measure $\mathcal{N}_p(\mu, \Sigma)$.

To motivate the formulation of our results, let us take a pause to see how one may interpret Le Cam's formulation (1.1)–(1.2) without going through explicitly the quantities appearing in the contiguity condition (1.3). The key observation is that, under mild additional integrability, the asymptotics in (1.1) and (1.2) necessarily entail that $\mathbb{E}_{P_n}(T_n - m_{P_n})/\sigma_{P_n} \approx 0$ and $\mathbb{E}_{Q_n}(T_n - m_{P_n})/\sigma_{P_n} \approx \tau$. This gives $\tau \approx (\mathbb{E}_{Q_n} T_n - \mathbb{E}_{P_n} T_n)/\sigma_{P_n}$, and therefore we may interpret (1.1)–(1.2) as

$$\frac{T_n - m_{P_n}}{\sigma_{P_n}} \overset{d}{\approx} \mathcal{N}(0,1) \quad \text{under } P_n \tag{1.5}$$

$$\overset{(1.3)}{\Rightarrow} \quad \frac{T_n - m_{P_n}}{\sigma_{P_n}} \overset{d}{\approx} \mathcal{N}\Big(\frac{\mathbb{E}_{Q_n} T_n - \mathbb{E}_{P_n} T_n}{\sigma_{P_n}}, 1\Big) \quad \text{under } Q_n. \tag{1.6}$$

Compared to (1.1)–(1.2), the above formulation does not involve parameters appearing in (1.3). Our first main result of this paper establishes an analogue of this formulation (1.5)–(1.6) for $T_n = T(X)$: Under mild regularity conditions on T, for any pair (μ_0, Σ_0), $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$,

$$\frac{T(X) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} \overset{d}{\approx} \mathcal{N}(0,1) \quad \text{under } (\mu_0, \Sigma_0) \tag{1.7}$$

$$\overset{(**)}{\Rightarrow} \quad \frac{T(X) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} \overset{d}{\approx} \mathcal{N}\Big(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}}, 1\Big) \quad \text{under } (\mu, \Sigma), \tag{1.8}$$

whereas our contiguity condition $(**)$ reads

$$\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)} \equiv \frac{V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{\max\{|m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}|, \sigma_{(\mu_0,\Sigma_0)}\}} \to 0. \tag{1.9}$$

See Theorem 2.1 for a formal statement. The "variance" parameter $V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}$, formally defined in (2.1) ahead, characterizes the order of stochastic dispersion of $T(X)$ under (μ, Σ) compared to that under (μ_0, Σ_0). Compared to Le Cam's contiguity condition (1.3) that requires an exact and strict distributional limit for the log likelihood ratio, (1.9) typically holds in a far broader regime than the prescribed regime in which a distributional limit of the log likelihood ratio exists. For example, in the special case of fixed-rank spiked covariance alternatives, (1.9) already holds for all possible $h_\ell \geq 0$ with all the covariance test statistics studied in this paper, as opposed to the highly restrictive regime imposed by the existence of the weak limit of the log likelihood ratio. In fact, a striking advantage of (1.9) in all the studied examples is its uniform validity over all possible covariance matrices without the need of specifying any particular structure (e.g., spiked alternatives).

In addition, for sufficiently regular $T_n = T(X)$ with $T(X) \in W^{1,2}(\gamma_{n\times p})$, verification of the contiguity condition (1.9) also has major operational advantages compared to the original Le Cam's contiguity condition (1.3). In particular, upper and lower bounds for the stochastic dispersion $V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}$, the mean difference $|m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}|$ and the null standard deviation $\sigma_{(\mu_0,\Sigma_0)}$ can usually be reduced to derivative calculations and their moment bounds via efficient applications of Poincaré inequalities (or other Fourier techniques in classical Gaussian analysis). If a bit further regularity persists in T_n in that $T(X) \in W^{2,4}(\gamma_{n\times p})$, which is the case for all examples considered in this paper, then a null CLT (1.7) can also be reduced to the same derivative calculations and moment bounding exercises, via the renowned second-order Poincaré inequality [8]. In essence, when the given statistic T_n possesses sufficient regularity, our contiguity method (1.7)–(1.8) can be used to derive its nonnull distributions in a rather "mechanical" way by evaluating derivatives and their moment (upper and lower) bounds.

It should be mentioned that while our approach (1.7)–(1.8) here appears to be particularly effective with the sufficient regularity of T_n that naturally postulates a null CLT, it seems less useful when such regularity fails and a different limit occurs under the null; see Remark 2.3 for some technical discussions. Whether a general, effective contiguity approach as (1.7)–(1.8) exists in the "low regularity" regime of T_n remains an interesting open question.

1.3. *Power formula of tests with high-dimensional normal population.* Similar to the wide applicability of Le Cam's third/contiguity lemma in (1.1)–(1.2) in classical statistical testing problems, our contiguity result in (1.7)–(1.8) can be applied to many modern high-dimensional statistical problems. Here is a general formulation of the testing problem with normal populations:

$$H_0 : (\mu, \Sigma) \in \mathscr{H}_0 \quad \text{versus} \quad H_1 : H_0 \text{ does not hold}, \tag{1.10}$$

where $\mathscr{H}_0$ is a subset of $\mathbb{R}^p \times \mathcal{M}_p$.

Let $T(X)$ be a generic test statistic whose distribution is invariant under H_0, that is, the law of $T(X)$ remains the same for any $(\mu, \Sigma) \in \mathscr{H}_0$ in (1.10). Due to the distributional invariance of $T(X)$, its mean and variance under the null

$$m_{H_0} \equiv m_{(\mu_0,\Sigma_0)}, \qquad \sigma^2_{H_0} \equiv \sigma^2_{(\mu_0,\Sigma_0)} \tag{1.11}$$

are well-defined for any specification of $(\mu_0, \Sigma_0) \in \mathscr{H}_0$. Suppose further that $T(X)$ verifies the CLT in (1.7); this indeed holds in all examples considered in this paper due to their sufficient regularity, and is also anticipated as many covariance tests statistics depend on

"sufficient average" of eigenvalues of the sample covariance matrix. Then an asymptotically exact test can be constructed immediately: for any prescribed $\alpha \in (0, 1)$,

$$\Psi(X) \equiv \Psi(X; m_{H_0}, \sigma_{H_0}) \equiv \mathbf{1}\Big(\frac{T(X) - m_{H_0}}{\sigma_{H_0}} > z_\alpha\Big), \tag{1.12}$$

where z_α is the normal quantile such that $\mathbb{P}(\mathcal{N}(0, 1) > z_\alpha) = \alpha$. Here, following the convention in statistics literature, $\Psi(X) = 1$ (resp. $\Psi(X) = 0$) indicates rejection (resp. acceptance) of the null hypothesis H_0 when X is observed. The quantities m_{H_0} and $\sigma^2_{H_0}$ are usually known in closed forms, at least asymptotically. Even not amenable to exact expression, these quantities can be simulated easily as well.

The quality of the test $\Psi(X)$ is measured by the *power function*, defined for each (μ, Σ) as

$$\text{Power of } \Psi(X) \text{ at } (\mu, \Sigma) \equiv \mathbb{P}_{(\mu,\Sigma)}\big(\Psi(X) \text{ rejects the null } H_0\big) = \mathbb{E}_{(\mu,\Sigma)}\Psi(X).$$

Applying our contiguity method (1.7)–(1.8), we get the following power formula for the test $\Psi(X)$ associated with the test statistic $T(X)$: Under the assumed CLT condition (1.7) and (a slight variation of) the contiguity condition (1.9),

$$\mathbb{E}_{(\mu,\Sigma)}\Psi(X) \approx 1 - \Phi\Big(z_\alpha - \frac{m_{(\mu,\Sigma)} - m_{H_0}}{\sigma_{H_0}}\Big). \tag{1.13}$$

Interestingly, the contiguity condition (1.9) is usually verified both for the case where (μ, Σ) is away from null set $\mathscr{H}_0$ in which the mean difference $|m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}|$ dominates the stochastic dispersion $V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}$, and for the case where (μ, Σ) is very close to the null in which the null standard deviation $\sigma_{(\mu_0,\Sigma_0)}$ dominates $V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}$. As such, the power formula (1.13) can usually be strengthened uniformly over all $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$.

1.4. *Two concrete applications of* (1.13). We give two concrete applications of (1.13) in the context of covariance testing, as a demonstration of the power of our contiguity result (1.7)–(1.8).

The first application of (1.13) is the test for identity $\Sigma = I$. In the growing p setting, this problem has been extensively studied in the literature; see,, for example, [1, 7, 10, 11, 14, 17, 22, 29, 32]. Among the tests studied in the above works, we apply our general theory (1.13) to the following two tests: likelihood ratio test (LRT) (see Section 3.1.1) and Ledoit–Nagao–Wolf's test [22, 24] (see Section 3.1.2). As an example, the LRT, denoted by $\Psi_{\mathrm{LRT}}(X)$, is shown to admit the following asymptotic power formula (see Theorem 3.3): under $\min\{n, p\} \to \infty$ with $\overline{\lim}(p/n) < 1$,

$$\mathbb{E}_{(\mu,\Sigma)}\Psi_{\mathrm{LRT}}(X) \sim 1 - \Phi\Big(z_\alpha - \frac{\mathcal{L}_S(\Sigma, I)}{\sqrt{2(-\frac{p}{n-1} - \log(1 - \frac{p}{n-1}))}}\Big). \tag{1.14}$$

Here $a \sim b$ stands for $a/b \to 1$ under the prescribed asymptotics, and $\mathcal{L}_S(\cdot, \cdot)$ is the matrix Stein loss to be defined in (3.5) ahead.

To give a flavor of how (1.14) follows from (1.13), recall that the key step in applying (1.13) is to establish that the contiguity condition $\overline{\mathrm{err}}_{(\mu,\Sigma)} \to 0$ in (1.9) (or the current variation defined in (2.6) ahead). In the LRT setting, a much stronger estimate can be proved in that $\overline{\mathrm{err}}_{(\mu,\Sigma)} \le Cp^{-1/3}$ holds for some absolute constant $C > 0$. This key estimate follows from a series of algebraic manipulations, upon calculating that $V^2_{(\mu,\Sigma)} = (n-1)\|\Sigma - I\|^2_F$, $m_{(\mu,\Sigma)} - m_{H_0} = [(n-1)/2]\mathcal{L}_S(\Sigma, I)$, and $\sigma^2_{H_0} \ge cp^2$ for some absolute constant $c > 0$. See Proposition 3.2 and its proof for more details.

The second application is the sphericity test $\Sigma = \lambda I$ for some unspecified $\lambda > 0$. In the growing p setting, this problem has previously been studied in [11, 14, 16, 17, 22, 29]. We

study in this paper the following two widely used tests: LRT for sphericity (Section 3.2.1), John's test [18] (Section 3.2.2), both invariant under H_0. Similar to the previous case, our results on the power behavior of these tests do not pose any assumption on the alternative Σ. As an example, the LRT for sphericity, denoted by $\Psi_{\mathrm{LRT},s}(X)$, is shown to admit the following asymptotic power formula (see Theorem 3.9): under $\min\{n, p\} \to \infty$ with $\overline{\lim}(p/n) < 1$,

$$\mathbb{E}_{(\mu,\Sigma)}\Psi_{\mathrm{LRT},s}(X) \sim 1 - \Phi\Bigg(z_\alpha - \frac{-\log\det(\Sigma \cdot b^{-1}(\Sigma))}{\sqrt{2(-\frac{p}{n-1} - \log(1 - \frac{p}{n-1}))}}\Bigg). \tag{1.15}$$

Here $\det(\cdot)$ is the matrix determinant and $b(\Sigma) \equiv \mathrm{tr}(\Sigma)/p$ with $\mathrm{tr}(\cdot)$ denoting the trace. To the best of our knowledge, the above power formula for the LRT in the sphericity is new in the literature.

The common feature of the power formulae obtained in this paper is that they require no assumptions on the alternative Σ and only mild conditions on the growth of (n, p), which goes much beyond the realm of previous available techniques. Roughly speaking, these available techniques either (i) directly establish a CLT under the alternative that crucially exploits the exact form of the test (cf. [7, 10, 15, 31]) and usually requires additional restrictions on the growth of (n, p) and the alternative covariance, or (ii) resort to the classical Le Cam's third/contiguity lemma in (1.1)–(1.2) (cf. [25, 26]) which, as mentioned above, necessarily fails in a broad regime of alternative covariance for which the log likelihood ratio cannot stabilize.

It is also worth mentioning that the precise power formulae we obtain for the aforementioned tests also have interesting implications compared to previous results in the literature that target at spiked covariance alternatives [25, 26, 31]. In particular, as will be clear in Section 3.3, although [25, 26, 31] showed that some of the aforementioned tests have asymptotically equivalent power behavior under the spiked covariance alternative with a fixed number of spikes, our new power characterizations indicate that such equivalence in general fails when many spikes exist.

An interesting question untouched in this paper concerns what the information-theoretic optimal power curve for (1.10) looks like, and whether the power formulae (1.14)–(1.15) (or power formulae for other tests) achieve such curves. In general, a successful study of this optimality problem requires two necessary elements: (i) an identification of the limits of the statistical experiments, and (ii) the solvability of the optimality problem in the identified limiting experiments. In classical low-dimensional statistical models (cf. [30]), (i) is achieved by the LAN property of the log likelihood ratio in these models with the help of the classical Le Cam's contiguity/third lemma (1.1)–(1.2), and (ii) is a consequence of the classical decision-theoretic optimality properties of the limiting Gaussian location shift model. In the context of covariance testing, significant progress has been made in [25, 26], in which the "limiting Gaussian experiment" is obtained for fixed-rank covariance alternatives. However, as the limiting experiment is not of the LAN type and its optimality properties remain unclear, the information-theoretic optimal power curve remains unknown. While we believe our contiguity method (1.7)–(1.8) is highly relevant to the optimality problem, achieving fully this goal is far beyond the content and scope of the current paper, and will therefore be deferred to a future study elsewhere.

1.5. *Organization*. The rest of the paper is organized as follows. We formalize our contiguity results (1.7)–(1.8) in Section 2. Section 3 is devoted to the application of our contiguity result to the problem of high-dimensional covariance testing. Some key spectral estimates that will be used in the proofs for the results in Section 3 are presented in Section 4 and may be of independent interest. Sections 5–6 contain the main proofs of results in Sections 3.1 and 3.2, with the rest of technical/auxiliary details deferred to the appendices [12].

1.6. *Notation.* For any positive integer n, let $[n]$ denote the set $\{1, \dots, n\}$. For $a, b \in \mathbb{R}$, $a \vee b \equiv \max\{a, b\}$ and $a \wedge b \equiv \min\{a, b\}$. For $a \in \mathbb{R}$, let $a_+ \equiv a \vee 0$ and $a_- \equiv (-a) \vee 0$. For $x \in \mathbb{R}^n$, let $\|x\|_p = \|x\|_{\ell_p(\mathbb{R}^n)}$ denote its p-norm ($0 \le p \le \infty$) with $\|x\|_2$ abbreviated as $\|x\|$. Let $B_p(r; x) \equiv \{z \in \mathbb{R}^p : \|z - x\| \le r\}$ be the unit ℓ_2 ball in $\mathbb{R}^p$. By $\mathbf{1}_n$ we denote the vector of all ones in $\mathbb{R}^n$. For a matrix $M \in \mathbb{R}^{n\times n}$, let $\|M\|_{\mathrm{op}}$ and $\|M\|_F$ denote the spectral and Frobenius norms of M respectively. We use $\{e_j\}$ to denote the canonical basis, whose dimension should be self-clear from the context.

We use C_x to denote a generic constant that depends only on x, whose numeric value may change from line to line unless otherwise specified. Notation $a \lesssim_x b$ and $a \gtrsim_x b$ mean $a \le C_x b$ and $a \ge C_x b$ respectively, and $a \asymp_x b$ means $a \lesssim_x b$ and $a \gtrsim_x b$. The symbol $a \lesssim b$ means $a \le Cb$ for some absolute constant C. For two nonnegative sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \ll b_n$ (respectively $a_n \gg b_n$) if $\lim_{n\to\infty}(a_n/b_n) = 0$ (respectively $\lim_{n\to\infty}(a_n/b_n) = \infty$). We write $a_n \sim b_n$ if $\lim_{n\to\infty}(a_n/b_n) = 1$. We follow the convention that $0/0 = 0$.

Let φ, Φ be the density and the cumulative distribution function of a standard normal random variable. For any $\alpha \in (0, 1)$, let z_α be the normal quantile defined by $\mathbb{P}(\mathcal{N}(0, 1) > z_\alpha) = \alpha$. For two random variables X, Y on $\mathbb{R}$, we use $d_{\mathrm{TV}}(X, Y)$ and $d_{\mathrm{Kol}}(X, Y)$ to denote their total variation distance and Kolmogorov distance defined respectively by

$$\begin{aligned} d_{\mathrm{TV}}(X, Y) &\equiv \sup_{B\in\mathcal{B}(\mathbb{R})} \big|\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)\big|, \\ d_{\mathrm{Kol}}(X, Y) &\equiv \sup_{t\in\mathbb{R}} \big|\mathbb{P}(X \le t) - \mathbb{P}(Y \le t)\big|. \end{aligned} \tag{1.16}$$

Here $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra of $\mathbb{R}$.

Let γ_d be the standard Gaussian measure on $\mathbb{R}^d$, and for $r, p \ge 1$ let $W^{r,p}(\gamma_d)$ be the completion of $C_0^\infty(\mathbb{R}^d)$, the space of smooth and compactly supported functions in $\mathbb{R}^d$, with respect to the norm

$$\|f\|_{r,p} \equiv \Big[\sum_{|\boldsymbol{\alpha}|\le r} \int |\partial^{\boldsymbol{\alpha}} f(x)|^p \gamma_d(\mathrm{d}x)\Big]^{1/p}. \tag{1.17}$$

In other words, $W^{r,p}(\gamma_d)$ is the Sobolev space with respect to the Gaussian measure γ_d.

2. Contiguity under high-dimensional Gaussianity.

2.1. *The formal description of* (1.7)–(1.8). Let $T : \mathbb{R}^{n\times p} \to \mathbb{R}$ be a measurable map. For any $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$, let $\mathscr{T}_{(\mu,\Sigma)} : \mathbb{R}^{n\times p} \to \mathbb{R}^{n\times p}$ be defined by

$$\mathscr{T}_{(\mu,\Sigma)}(z) \equiv \nabla T\big(z\Sigma^{1/2} + \mathbf{1}_n\mu^\top\big)\Sigma^{1/2}, \quad z \in \mathbb{R}^{n\times p}.$$

Here $\mathbf{1}_n$ is the n-vector of all ones, and $\nabla T : \mathbb{R}^{n\times p} \to \mathbb{R}^{n\times p}$ is the map with $(\nabla T(z))_{ij} = \partial T(z)/\partial z_{ij}$. Let $Z_1, \dots, Z_n$ be i.i.d. random variables with a standard p-variate normal distribution $\mathcal{N}(0, I_p)$. For any $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$, define the quantity

$$V^2_{(\mu,\Sigma);(\mu_0,\Sigma_0)} \equiv \mathbb{E}\big\|\mathscr{T}_{(\mu,\Sigma)}(Z) - \mathscr{T}_{(\mu_0,\Sigma_0)}(Z)\big\|_F^2. \tag{2.1}$$

Now we are in a position to give a formal description of our contiguity result (1.7)–(1.8) under the condition (1.9). Recall the quantities $m_{(\mu,\Sigma)}$, m_{H_0}, $\sigma^2_{(\mu,\Sigma)}$, $\sigma^2_{H_0}$ defined in (1.4) and (1.11) and that $\gamma_{n\times p}$ denotes the standard Gaussian measure in $\mathbb{R}^{n\times p}$.

THEOREM 2.1. *Suppose that $T : \mathbb{R}^{n\times p} \to \mathbb{R}$ is an element of $W^{1,2}(\gamma_{n\times p})$. Then for any pair $(\mu_0, \Sigma_0), (\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$ and $t \in \mathbb{R}$,*

$$\left|\mathbb{P}_{(\mu,\Sigma)}\left(\frac{T(X)-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\right) - \mathbb{P}\left(\mathcal{N}\left(\frac{m_{(\mu,\Sigma)}-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}}, 1\right) > t\right)\right|$$
$$\le \mathrm{err}_{(\mu_0,\Sigma_0)} + C\cdot\big((1+|t|)\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)}\big)^{2/3}.$$

Here $C > 0$ is a universal constant, $\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)}$ is defined in (1.9), and

$$\mathrm{err}_{(\mu_0,\Sigma_0)} \equiv d_{\mathrm{Kol}}\left(\frac{T(X)-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}}, \mathcal{N}(0,1)\right) \quad \textit{under } (\mu_0,\Sigma_0) \tag{2.2}$$

is the normal approximation error of $T(X)$ under (μ_0, Σ_0) in Kolmogorov distance as defined in (1.16).

At this point, Theorem 2.1 does not yet exactly guarantee the closeness of the distribution of the random variable $(T(X)-m_{(\mu_0,\Sigma_0)})/\sigma_{(\mu_0,\Sigma_0)}$ to a standard normal shifted by $(m_{(\mu,\Sigma)}-\mu_{(\mu_0,\Sigma_0)})/\sigma_{(\mu_0,\Sigma_0)}$ under (μ, Σ), as $t \in \mathbb{R}$ in the above theorem cannot be chosen arbitrarily large to yield an informative bound. This is however not a deficiency of our formulation in (1.7)–(1.8). In fact, the range of admissible $t \in \mathbb{R}$ depends on the magnitude of the mean shift parameter $(m_{(\mu,\Sigma)}-\mu_{(\mu_0,\Sigma_0)})/\sigma_{(\mu_0,\Sigma_0)}$. As shown in the following corollary, when the mean shift parameter $(m_{(\mu,\Sigma)}-\mu_{(\mu_0,\Sigma_0)})/\sigma_{(\mu_0,\Sigma_0)}$ is bounded, the conclusion of Theorem 2.1 can indeed be strengthened to be uniform in $t \in \mathbb{R}$. This is in similar spirit to the classical Le Cam's formulation (1.1)–(1.2), in which the mean shift parameter $\tau \in \mathbb{R}$ is treated as a fixed, finite real number in the asymptotics.

COROLLARY 2.2. *Consider the same setting as in Theorem 2.1. Suppose further that*

$$\frac{|m_{(\mu,\Sigma)}-m_{(\mu_0,\Sigma_0)}|}{\sigma_{(\mu_0,\Sigma_0)}} \le K \tag{2.3}$$

for some $K > 0$. Then there exists some constant $C_K > 0$,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}_{(\mu,\Sigma)}\left(\frac{T(X)-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\right) - \mathbb{P}\left(\mathcal{N}\left(\frac{m_{(\mu,\Sigma)}-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}}, 1\right) > t\right)\right|$$
$$\le \mathrm{err}_{(\mu_0,\Sigma_0)} + C_K\cdot\overline{\mathrm{err}}^{4/9}_{(\mu,\Sigma);(\mu_0,\Sigma_0)}.$$

There is no a priori reason to believe that the exponent 4/9 is optimal, but this will have no impact on the qualitative distributional approximation under the contiguity condition $\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)} \to 0$.

REMARK 2.3. It is possible to formulate a version of Theorem 2.1 without assuming the regularity/integrability $T(X) \in W^{1,2}(\gamma_{n\times p})$ and a null CLT as follows. Suppose that sequences of $\{m_{(\mu,\Sigma)}\}, \{m_{(\mu_0,\Sigma_0)}\} \subset \mathbb{R}$, $\{\sigma_{(\mu_0,\Sigma_0)}\} \subset \mathbb{R}_{>0}$ are chosen such that the following hold:

- (Null distribution) There exists some random variable Y such that $(T(X)-m_{(\mu_0,\Sigma_0)})/\sigma_{(\mu_0,\Sigma_0)} \rightsquigarrow Y$ holds under the sequence of $\{(\mu_0,\Sigma_0)\}$.
- (Finite mean shift) $(m_{(\mu,\Sigma)}-m_{(\mu_0,\Sigma_0)})/\sigma_{(\mu_0,\Sigma_0)} \to \tau$ for some $\tau \in \mathbb{R}$.
- (Generalized contiguity) With $X^{(\mu,\Sigma)} \equiv Z\Sigma^{1/2} + \mathbf{1}_n\mu^\top$,

$$\frac{|(T(X^{(\mu,\Sigma)})-m_{(\mu,\Sigma)}) - (T(X^{(\mu_0,\Sigma_0)})-m_{(\mu_0,\Sigma_0)})|}{|m_{(\mu,\Sigma)}-m_{(\mu_0,\Sigma_0)}| \vee \sigma_{(\mu_0,\Sigma_0)}} = \mathfrak{o}_{\mathbf{P}}(1). \tag{2.4}$$

Then

$$\frac{T(X) - m_{(\mu_0,\Sigma_0)}}{\sigma_{\mu_0,\Sigma_0}} \rightsquigarrow Y + \tau \quad \text{under the sequence of } \{(\mu, \Sigma)\}.$$

The proof of this result is simpler than that of Theorem 2.1 so we omit the details.

While the above result may be of some interest at an abstract level, currently we are not aware of a concrete example in which the above result leads to genuine reductions in the problem complexity beyond the scope of Theorem 2.1. A leading test case here is to take $T(X) = \lambda_1(S)$ as the top eigenvalue of the sample covariance S defined in (3.2) below, under the rank-one spiked alternative $\Sigma = I + hvv^\top$, $\|v\| = 1$. Then in the subcritical regime $h \in (0, \sqrt{c})$ where $c \equiv \lim p/n$, it is not hard to see that verification of the generalized contiguity condition (2.4) is equivalent to proving a (Type I) Tracy–Widom limit for λ_1 under the prescribed alternative. It remains an interesting open question, as to whether a general contiguity method without assuming $T(X) \in W^{1,2}(\gamma_{n\times p})$ and a null CLT can be formulated that leads a genuine reduction in deriving, say, at least the Tracy–Widom limit of λ_1 under the rank-one alternative in the prescribed subcritical regime.

2.2. *Power formula for tests with high-dimensional normal population.* Consider the general testing problem (1.10). Recall that $T(X)$ is a generic test statistic whose distribution remains invariant under H_0, and the generic test $\Psi(X)$ defined in (1.12). To formalize the power formula in (1.13), we need a slight variation of the quantity $V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}$ for nonsingleton $\mathscr{H}_0$. Let

$$V^2_{(\mu,\Sigma)} \equiv \inf_{(\mu_0,\Sigma_0)\in\mathscr{H}_0} \mathbb{E}\big\|\mathscr{T}_{(\mu,\Sigma)}(Z) - \mathscr{T}_{(\mu_0,\Sigma_0)}(Z)\big\|_F^2. \tag{2.5}$$

Now we formalize (1.13) which is an immediate consequence of Theorem 2.1.

COROLLARY 2.4. *Suppose that $T : \mathbb{R}^{n\times p} \to \mathbb{R}$ is an element of $W^{1,2}(\gamma_{n\times p})$ and the law of $T(X)$ is invariant under H_0. For any $\alpha \in (0,1)$, there exists some $C_\alpha > 0$ such that*

$$\left|\mathbb{E}_{(\mu,\Sigma)}\Psi(X) - \left[1 - \Phi\left(z_\alpha - \frac{m_{(\mu,\Sigma)} - m_{H_0}}{\sigma_{H_0}}\right)\right]\right| \le \mathrm{err}_{H_0} + C_\alpha\left(\frac{V_{(\mu,\Sigma)}}{|m_{(\mu,\Sigma)} - m_{H_0}| \vee \sigma_{H_0}}\right)^{2/3}$$

holds for any $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$. Here err_{H_0} is defined in (2.2) with $m_{(\mu_0,\Sigma_0)}$, $\sigma_{(\mu_0,\Sigma_0)}$ replaced by m_{H_0}, σ_{H_0}.

The above result reduces the analysis of the power behavior of $\Psi(X)$ into essentially the following two steps:

1. (*Normal approximation under H_0*) Show that

$$\mathrm{err}_{H_0} = d_{\mathrm{Kol}}\left(\frac{T(X) - m_{H_0}}{\sigma_{H_0}}, \mathcal{N}(0,1)\right) \to 0 \quad \text{under } H_0.$$

2. (*Contiguity condition*) Show that

$$\overline{\mathrm{err}}_{(\mu,\Sigma)} \equiv \frac{V_{(\mu,\Sigma)}}{|m_{(\mu,\Sigma)} - m_{H_0}| \vee \sigma_{H_0}} \to 0. \tag{2.6}$$

Normal approximation of $T(X)$ under H_0 can be established in different ways. When $T(X)$ possesses further regularity, say $T(X) \in W^{2,4}(\gamma_{n\times p})$, a null CLT can usually be established efficiently via Chatterjee's second-order Poincaré inequality [8]. This approach is particularly compatible with the contiguity condition (2.6), as we only need to calculate derivatives of $T(X)$ and obtain good enough moment upper and lower bounds for these derivatives. In Section 3 ahead, we implement this method to a variety of statistics in two concrete problems of high-dimensional covariance testing. Some of the resulting exact power results are new, and some improve significantly over earlier results in the literature, both in terms of (n, p)-conditions and applicable alternatives.

2.3. *Proof of Theorem 2.1.*

LEMMA 2.5. *For any $t, u \in \mathbb{R}$ and $\eta \in \mathbb{R}$,*

$$\bigl|\mathbb{P}\bigl(\mathcal{N}(u,1) \le t\bigr) - \mathbb{P}\bigl(\mathcal{N}\bigl((1+\eta)u, 1\bigr) \le t\bigr)\bigr| \le 2\bigl(1+|t|\bigr)\cdot|\eta|.$$

PROOF. This result strengthens [13], Lemma 5.4. We assume without loss of generality $\eta \in [-1/2, 1/2]$ because otherwise the right-hand side of the desired display is greater than or equal to 1. Note that the left-hand side is bounded by

$$\left|\int_{t-u}^{t-(1+\eta)u} \varphi(z)\,\mathrm{d}z\right| \le |\eta| \cdot \Bigl[\sup_{v\in[(t-u)-|\eta u|,(t-u)+|\eta u|]} \varphi(v)|u|\Bigr] \equiv |\eta| \cdot M_t(u).$$

Here $\varphi(\cdot)$ is the normal density. First consider the case $u \ge 0$. Then $M_t(u) \le \sup_{v\in[t-3u/2,t-u/2]} \varphi(v)u$, which can be bounded further in different situations:

- If $t - u/2 \le 0$, then

$$\begin{aligned} M_t(u) &\le \varphi\Bigl(t-\frac{u}{2}\Bigr)u = \varphi\Bigl(t-\frac{u}{2}\Bigr)(u-2t) + 2t\varphi\Bigl(t-\frac{u}{2}\Bigr) \\ &\le 2\sup_{x\in\mathbb{R}}|x|\varphi(x) + \frac{2}{\sqrt{2\pi}}|t| = \frac{2}{\sqrt{2\pi e}} + \frac{2}{\sqrt{2\pi}}|t|. \end{aligned}$$

  Here we used the readily verified fact that $\sup_{x\in\mathbb{R}}|x|\varphi(x) = 1/\sqrt{2\pi e}$.
- If $t - 3u/2 \ge 0$, then

$$\begin{aligned} M_t(u) &\le \varphi\Bigl(t-\frac{3u}{2}\Bigr)u = \varphi\Bigl(t-\frac{3u}{2}\Bigr)\Bigl(u-\frac{2t}{3}\Bigr) + \frac{2}{3}t\varphi\Bigl(t-\frac{3u}{2}\Bigr) \\ &\le \frac{2}{3}\Bigl(\frac{1}{\sqrt{e}} + \frac{1}{\sqrt{2\pi}}|t|\Bigr). \end{aligned}$$

- Otherwise $(2/3)t \le u \le 2t$, so $M_t(u) \le |u| \le 2|t|$.

The case $u < 0$ can be handled similarly, so we have $\sup_u M_t(u) \le 2(1+|t|)$. □

PROOF OF THEOREM 2.1. Let $Z \in \mathbb{R}^{n\times p}$ be a matrix generated by n i.i.d. samples from $\mathcal{N}(0, I_p)$. Let $X^{(\mu,\Sigma)} \equiv Z\Sigma^{1/2} + \mathbf{1}_n\mu^\top$. Then,

$$\begin{aligned} \frac{T(X) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} &\overset{d}{=} \frac{T(X^{(\mu,\Sigma)}) - T(X^{(\mu_0,\Sigma_0)})}{\sigma_{(\mu_0,\Sigma_0)}} + \frac{T(X^{(\mu_0,\Sigma_0)}) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} \\ &= \frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \frac{W(Z)}{\sigma_{(\mu_0,\Sigma_0)}} + \frac{T(X^{(\mu_0,\Sigma_0)}) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}}. \end{aligned} \tag{2.7}$$

Here $W(Z)$ is the centered variable defined by

$$W(Z) \equiv T\bigl(Z\Sigma^{1/2} + \mathbf{1}_n\mu^\top\bigr) - T\bigl(Z\Sigma_0^{1/2} + \mathbf{1}_n\mu_0^\top\bigr) - (m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}).$$

Using the chain rule,

$$\begin{aligned} \partial_{(ij)}W(Z) &= \bigl(\nabla T\bigl(X^{(\mu,\Sigma)}\bigr)\Sigma^{1/2} - \nabla T\bigl(X^{(\mu_0,\Sigma_0)}\bigr)\Sigma_0^{1/2}\bigr)_{ij} \\ &= \bigl(\mathscr{T}_{(\mu,\Sigma)}(Z) - \mathscr{T}_{(\mu_0,\Sigma_0)}(Z)\bigr)_{ij}. \end{aligned}$$

By the Gaussian–Poincaré inequality ([5], Theorem 3.20),

$$\mathrm{Var}\bigl(W(Z)\bigr) \le \mathbb{E}\Bigl[\sum_{(ij)}\bigl(\partial_{(ij)}W(Z)\bigr)^2\Bigr] = \mathbb{E}\bigl\|\mathscr{T}_{(\mu,\Sigma)}(Z) - \mathscr{T}_{(\mu_0,\Sigma_0)}(Z)\bigr\|_F^2 = V^2_{(\mu,\Sigma);(\mu_0,\Sigma_0)}.$$

This means for any $u > 0$, on an event E with probability at least $1 - u^{-2}$,

$$\big|W(Z)\big| \le u \cdot V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}.$$

Hence, for any $t \in \mathbb{R}$, the decomposition (2.7) entails that (recall the definition of $\mathrm{err}_{(\mu_0,\Sigma_0)}$ in (2.2))

$$\begin{aligned}
&\mathbb{P}\bigg(\frac{T(X^{(\mu,\Sigma)}) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\bigg) \\
&\quad= \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)} + W(Z)}{\sigma_{(\mu_0,\Sigma_0)}} + \frac{T(X^{(\mu_0,\Sigma_0)}) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\bigg) \\
&\quad\le \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)} + u \cdot V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \frac{T(X^{(\mu_0,\Sigma_0)}) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\bigg) + \frac{1}{u^2} \\
&\quad\le \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)} + u \cdot V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \mathcal{N}(0,1) > t\bigg) + \frac{1}{u^2} + \mathrm{err}_{(\mu_0,\Sigma_0)} \\
&\quad\equiv \mathfrak{p}(u) + \mathrm{err}_{(\mu_0,\Sigma_0)}.
\end{aligned}$$

Next we bound $\mathfrak{p}(\cdot)$ using two different ways. First, by Lemma 2.5, we have

$$\begin{aligned}
\inf_{u>0} \mathfrak{p}(u) &\le \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \mathcal{N}(0,1) > t\bigg) \\
&\qquad + \inf_{u>0}\bigg[2\big(1+|t|\big)u \cdot \frac{V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{|m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}|} + \frac{1}{u^2}\bigg] \\
&\le \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \mathcal{N}(0,1) > t\bigg) + C\bigg(\frac{(1+|t|)V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{|m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}|}\bigg)^{2/3}.
\end{aligned}$$

On the other hand, by anticoncentration of the standard normal distribution, that is, $|\mathbb{P}(\mathcal{N}(0,1) \le a) - \mathbb{P}(\mathcal{N}(0,1) \le b)| \le |a-b|$ for any $a, b \in \mathbb{R}$,

$$\begin{aligned}
\inf_{u>0} \mathfrak{p}(u) &\le \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \mathcal{N}(0,1) > t\bigg) + \inf_{u>0}\bigg[\frac{u \cdot V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \frac{1}{u^2}\bigg] \\
&\le \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \mathcal{N}(0,1) > t\bigg) + C\bigg(\frac{V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}}\bigg)^{2/3}.
\end{aligned}$$

Collecting the bounds completes the proof for one direction. For the other direction, we have

$$\begin{aligned}
&\mathbb{P}\bigg(\frac{T(X^{(\mu,\Sigma)}) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\bigg) \\
&\quad= \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)} + W(Z)}{\sigma_{(\mu_0,\Sigma_0)}} + \frac{T(X^{(\mu_0,\Sigma_0)}) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\bigg) \\
&\quad\ge \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)} - u \cdot V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \frac{T(X^{(\mu_0,\Sigma_0)}) - m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\bigg) - \frac{1}{u^2} \\
&\quad\ge \mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)} - u \cdot V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \mathcal{N}(0,1) > t\bigg) - \frac{1}{u^2} - \mathrm{err}_{(\mu_0,\Sigma_0)}.
\end{aligned}$$

Using similar arguments as in the previous direction by invoking the two different bounds concludes the inequality. □

2.4. *Proof of Corollary 2.2.* Using the decomposition (2.7), and the assumed boundedness condition $|(m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)})/\sigma_{(\mu_0,\Sigma_0)}| \le K$ (which entails that $\sigma_{(\mu_0,\Sigma_0)} \ge (K+1)^{-1}(|m_{(\mu,\Sigma)} - m_{(\mu_0,\Sigma_0)}| + \sigma_{(\mu_0,\Sigma_0)})$), we have

$$\begin{aligned}\operatorname{Var}_{(\mu,\Sigma)}\biggl(\frac{T(X)-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}}\biggr) &\lesssim K^2 + \biggl(\frac{V_{(\mu,\Sigma);(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}}\biggr)^2 + 1\\ &\lesssim (1\vee K)^2(1+\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)})^2.\end{aligned}$$

Fix $t>0$. Then the above variance bound leads to

$$\mathbb{P}\biggl(\frac{T(X^{(\mu,\Sigma)})-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\biggr) \lesssim (t-K)_+^{-2}(1\vee K)^2(1+\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)})^2.$$

On the other hand,

$$\mathbb{P}\biggl(\frac{m_{(\mu,\Sigma)}-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} + \mathcal{N}(0,1) > t\biggr) \le e^{-(t-K)_+^2/2}.$$

Combined with Theorem 2.1, we have, for any $t>0$,

$$\begin{aligned}&\biggl|\mathbb{P}_{(\mu,\Sigma)}\biggl(\frac{T(X)-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}} > t\biggr) - \mathbb{P}\biggl(\mathcal{N}\biggl(\frac{m_{(\mu,\Sigma)}-m_{(\mu_0,\Sigma_0)}}{\sigma_{(\mu_0,\Sigma_0)}},1\biggr) > t\biggr)\biggr|\\ &\quad\le \mathrm{err}_{(\mu_0,\Sigma_0)} + C_K\cdot\min\bigl\{\bigl((1+|t|)\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)}\bigr)^{2/3}, t^{-2}(1+\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)})^2\bigr\}.\end{aligned}$$

If $\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)} \ge 1$, then a trivial bound works; otherwise if $\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)} \le 1$, then the right-hand side of the above display can be further bounded by

$$\begin{aligned}&\mathrm{err}_{(\mu_0,\Sigma_0)} + C_K\cdot\min\bigl\{\bigl((1+|t|)\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)}\bigr)^{2/3}, t^{-2}\bigr\}\\ &\quad\le \mathrm{err}_{(\mu_0,\Sigma_0)} + C_K'\cdot\overline{\mathrm{err}}_{(\mu,\Sigma);(\mu_0,\Sigma_0)}^{4/9}.\end{aligned}$$

A similar bound holds when $t\le 0$.

3. Applications to high-dimensional covariance testing.

3.1. *Testing identity $\Sigma = I$.* Consider the testing problem

$$H_0: \Sigma = I \quad \text{versus} \quad H_1: H_0 \text{ does not hold.} \tag{3.1}$$

This is a special case of (1.10) by taking $\mathscr{H}_0 = \mathbb{R}^p \times \{I\}$, and has been extensively studied in the literature; see [1, 7, 10, 11, 14, 17, 22, 29, 32] for an incomplete list.

We introduce some additional notation. Based on i.i.d. samples $X_1,\dots,X_n$ from $\mathcal{N}(\mu,\Sigma)$, the sample covariance matrix and its unbiased modification are given by

$$\begin{aligned}S_* &\equiv n^{-1}\sum_{k=1}^n (X_k-\bar X)(X_k-\bar X)^\top \quad \text{with } \bar X \equiv n^{-1}\sum_{i=1}^n X_i,\\ S &\equiv \frac{n}{N}S_* \overset{d}{=} \frac{1}{N}\sum_{k=1}^N (X_k-\mu)^\top(X_k-\mu).\end{aligned} \tag{3.2}$$

Here

$$N = n-1 \tag{3.3}$$

and the equal in distribution in (3.2) follows from [23], Theorem 3.1.2. Throughout the rest of the paper, we will mainly work with S for mathematical simplicity (unless otherwise specified), and adopt the right most expression of (3.2) as its definition whenever no confusion could arise.

3.1.1. *LRT.* This subsection studies the behavior of the LRT for testing (3.1). The modified log likelihood ratio statistic $T_{\mathrm{LRT}} : \mathbb{R}^{N\times p} \to \mathbb{R}$ (cf. [23], Theorem 8.4.2) is defined as

$$T_{\mathrm{LRT}}(X) \equiv \frac{N}{2}\big[\mathrm{tr}(S) - \log\det S - p\big]. \tag{3.4}$$

Clearly the law of $T_{\mathrm{LRT}}(X)$ is invariant under H_0. Corollary 2.4 applies in view of the regularity of T_{LRT} (see Appendix B). We use $(m_{\Sigma;\mathrm{LRT}}, \sigma_{\Sigma;\mathrm{LRT}}, V_{\Sigma;\mathrm{LRT}})$ to represent their generic versions defined in (1.4) and (2.5).

Following the discussion after Corollary 2.4, we start by establishing a quantitative CLT for $T_{\mathrm{LRT}}(X)$ under H_0; its proof is presented in Section 5.1.2.

PROPOSITION 3.1. *Suppose $p/N \le 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Then there exists some constant $C = C(\varepsilon) > 0$, such that under H_0,*

$$d_{\mathrm{TV}}\left(\frac{T_{\mathrm{LRT}}(X) - m_{I;\mathrm{LRT}}}{\sigma_{I;\mathrm{LRT}}}, \mathcal{N}(0,1)\right) \le \frac{C}{p}.$$

The CLT for the log likelihood ratio statistic $T_{\mathrm{LRT}}(X)$ under H_0 was first derived in [1] using random matrix theory under the assumption that $p/n \to y$ for some $y \in (0,1)$. This result was then improved in [14] and [10] to hold under the condition $n > p+1$ and $p \to \infty$, and in [32] to relax the Gaussian assumption. The condition $p/N \le 1-\varepsilon$ in Proposition 3.1 is used to derive the stable estimate $\mathbb{E}\|S^{-1}\|_{\mathrm{op}} \le C$ for some constant $C = C(\varepsilon) > 0$; see Lemma 4.3 for details. To our best knowledge, the above result is the first quantitative CLT for $T_{\mathrm{LRT}}(X)$ under H_0 in the literature.

The following result establishes the contiguity condition (2.6) for the log likelihood ratio statistic $T_{\mathrm{LRT}}(X)$; its proof is presented in Section 5.1.3. For p.s.d. Σ_1 and p.d. Σ_2, let

$$\mathcal{L}_S(\Sigma_1, \Sigma_2) \equiv \mathrm{tr}\big(\Sigma_1\Sigma_2^{-1}\big) - \log\det\big(\Sigma_1\Sigma_2^{-1}\big) - p \tag{3.5}$$

be the Stein loss with the convention that $\mathcal{L}_S(\Sigma_1, \Sigma_2) \equiv \infty$ if Σ_1 is singular.

PROPOSITION 3.2. *Suppose Σ is nonsingular. The following hold*:

1. $V^2_{\Sigma;\mathrm{LRT}} = N\|\Sigma - I\|_F^2$.
2. $m_{\Sigma;\mathrm{LRT}} - m_{I;\mathrm{LRT}} = (N/2)\mathcal{L}_S(\Sigma, I)$.
3. *In the asymptotic regime $N \ge p+1$ with $p \to \infty$,*

$$\sigma^2_{I;\mathrm{LRT}} \sim \frac{N^2}{2}\left[-\frac{p}{N} - \log\left(1 - \frac{p}{N}\right)\right].$$

In particular, $\sigma^2_{I;\mathrm{LRT}} \ge cp^2$ for some universal constant $c > 0$.

4. *There exists some universal constant $C > 0$ such that*

$$\frac{V_{\Sigma;\mathrm{LRT}}}{|m_{\Sigma;\mathrm{LRT}} - m_{I;\mathrm{LRT}}| \vee \sigma_{I;\mathrm{LRT}}} \le \frac{C}{p^{1/2}}.$$

The above proposition gives a prototypical example of how to proceed with the contiguity condition (2.6). For the log likelihood ratio statistic $T_{\mathrm{LRT}}(X)$ defined in (3.4), both $V_{\Sigma;\mathrm{LRT}}$ and the mean difference $m_{\Sigma;\mathrm{LRT}} - m_{I;\mathrm{LRT}}$ admit easy-to-handle closed-form formulae. To give some insights for the bound obtained in Proposition 3.2-(4), let us consider the 'local regime' of alternatives in which $\mathcal{L}_S(\Sigma, I) \approx \|\Sigma - I\|_F^2$. Then (2.6) can be bounded, up to a constant, by

$$\frac{\sqrt{N\|\Sigma - I\|_F^2}}{N\|\Sigma - I\|_F^2 \vee \sigma_{I;\mathrm{LRT}}} \le \sup_{x\ge 0}\frac{x}{x^2 \vee \sigma_{I;\mathrm{LRT}}} = \frac{1}{\inf_{x\ge 0}(x \vee \frac{\sigma_{I;\mathrm{LRT}}}{x})} = \frac{1}{\sigma^{1/2}_{I;\mathrm{LRT}}},$$

in the prescribed local regime of alternatives. The above simple reasoning exemplifies the essential reason why (2.6) must be small in high dimensions: if Σ is sufficiently away from I, then the mean difference $m_{\Sigma;\mathrm{LRT}} - m_{I;\mathrm{LRT}}$ is substantially larger than $V_{\Sigma;\mathrm{LRT}}$, but would otherwise be compensated by a diverging $\sigma_{I;\mathrm{LRT}}$.

Let $\Psi_{\mathrm{LRT}}(X)$ be the LRT built from the generic test (1.10) and the log likelihood ratio statistic $T_{\mathrm{LRT}}(X)$. Now Corollary 2.4 yields the following.

THEOREM 3.3. *Suppose $p/N \le 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Then there exists some constant $C = C(\varepsilon,\alpha) > 0$ such that*

$$\left|\mathbb{E}_\Sigma \Psi_{\mathrm{LRT}}(X) - \mathbb{P}\left(\mathcal{N}\left(\frac{N\cdot \mathcal{L}_S(\Sigma, I)}{2\sigma_I}, 1\right) > z_\alpha\right)\right| \le C\cdot p^{-1/3}. \tag{3.6}$$

Consequently, in the asymptotic regime $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < 1$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{LRT}}(X) \sim 1 - \Phi\left(z_\alpha - \frac{\mathcal{L}_S(\Sigma, I)}{\sqrt{2(-\frac{p}{N} - \log(1-\frac{p}{N}))}}\right).$$

Compared to results in [10] on the power behavior of the LRT $\Psi_{\mathrm{LRT}}(X)$, we remove the condition $\sup_n \|\Sigma\|_{\mathrm{op}} < \infty$ completely. This unnecessary condition arises as a technical deficiency in the approach of [10] that attempts at directly establishing a CLT for $\Psi_{\mathrm{LRT}}(X)$ under general alternatives.

3.1.2. *Ledoit–Nagao–Wolf's test.* This subsection studies testing (3.1) using the (rescaled) modified Nagao's trace statistic [24] by Ledoit and Wolf [22]:

$$T_{\mathrm{LNW}}(X) \equiv \frac{N}{4}\left[\mathrm{tr}(S-I)^2 - \frac{1}{N}\mathrm{tr}^2(S)\right]. \tag{3.7}$$

An asymptotically equivalent statistic as an unbiased estimator of $\|\Sigma - I\|_F^2$ has also been studied in [29]. One advantage of using (3.7) is that it applies to the case $p > n$ where the LRT in Section 3.1.1 becomes degenerate.

We will use $(m_{\Sigma;\mathrm{LNW}}, \sigma_{\Sigma;\mathrm{LNW}}, V_{\Sigma;\mathrm{LNW}})$ to represent their generic versions defined in (1.4) and (2.5).

PROPOSITION 3.4. *There exists an absolute constant $C > 0$ such that under H_0,*

$$d_{\mathrm{TV}}\left(\frac{T_{\mathrm{LNW}}(X) - m_{I;\mathrm{LNW}}}{\sigma_{I;\mathrm{LNW}}}, \mathcal{N}(0,1)\right) \le \frac{C}{N\wedge p}.$$

The proof is presented in Section 5.2.2. The CLT for $T_{\mathrm{LNW}}(X)$ was first derived in [22], Proposition 7, under the condition that $p/N \to y \in (0,\infty)$, which was later improved in [4], Theorem 3.6, to include the case $y \in \{0,\infty\}$. Here we give explicit error bounds in the normal approximation.

The following result establishes the contiguity condition (2.6) for T_{LNW}; its proof is presented in Section 5.2.3.

PROPOSITION 3.5. *Suppose $p/N \le M$ for some $M > 0$. Then the following hold*:

1. $V_{\Sigma;\mathrm{LNW}}^2 \le C_1 N(\|\Sigma\|_{\mathrm{op}}^2 \vee 1)\|\Sigma - I\|_F^2$ *for some constant $C_1 = C_1(M) > 0$.*
2. *With $Q_{\mathrm{LNW}}(\Sigma) \equiv (N^{-1} - 2N^{-2})\,\mathrm{tr}(\Sigma^2 - I)$,*

$$m_{\Sigma;\mathrm{LNW}} - m_{(0,I)} = \frac{N}{4}\left[\|\Sigma - I\|_F^2 + Q_{\mathrm{LNW}}(\Sigma)\right].$$

3. *In the asymptotic regime* $N \wedge p \to \infty$,

$$\sigma^2_{I;\mathrm{LNW}} \sim \frac{p^2}{4}.$$

4. *There exists some constant* $C_2 = C_2(M) > 0$ *such that*

$$\frac{V_{\Sigma;\mathrm{LNW}}}{|m_{\Sigma;\mathrm{LNW}} - m_{I;\mathrm{LNW}}| \vee \sigma_{I;\mathrm{LNW}}} \le \frac{C_2}{p^{1/2}}.$$

Compared to Proposition 3.2, although a closed-form formula is available for $m_{\Sigma;\mathrm{LNW}}$, a somewhat undesirable "residual term" $Q_{\mathrm{LNW}}(\Sigma)$ exists. Removing the effect of these terms in the final step (4) requires significant additional technicalities, as will be detailed in Section 5.2.3.

Let $\Psi_{\mathrm{LNW}}(X)$ be the test built from (1.10) and the statistic in (3.7). Combining the above results with Corollary 2.4 and some additional efforts to remove the residual term $Q_{\mathrm{LNW}}(\Sigma)$ in the mean difference formula (2) in the above proposition, we have the following asymptotic power formula for $\Psi_{\mathrm{LNW}}(X)$; see Section 5.2.4 for its proof.

THEOREM 3.6. *Suppose* $p/N \le M$ *for some* $M > 0$. *Then there exists some constant* $C = C(\alpha, M) > 0$ *such that*

$$\left|\mathbb{E}_\Sigma \Psi_{\mathrm{LNW}}(X) - \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot \|\Sigma - I\|_F^2}{4\sigma_{I;\mathrm{LNW}}}, 1\right) > z_\alpha\right)\right| \le C \cdot p^{-1/3}.$$

Consequently, in the asymptotic regime $N \wedge p \to \infty$ *with* $\overline{\lim}(p/N) < \infty$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{LNW}}(X) \sim 1 - \Phi\left(z_\alpha - \frac{\|\Sigma - I\|_F^2}{2(p/N)}\right).$$

The asymptotic behavior of T_{LNW} under the alternative is previously only known in [29], Theorem 4.1, under rather restrictive conditions on both Σ and growth of p. Theorem 3.6 only requires p/N to be bounded and makes no assumptions on Σ.

3.2. *Testing sphericity* $\Sigma = \lambda I$. Consider the testing problem

$$H_0 : \Sigma = \lambda I \quad \text{versus} \quad H_1 : H_0 \text{ does not hold} \tag{3.8}$$

for some un-specified $\lambda > 0$. This is a special case of (1.10) by taking $\mathscr{H}_0 = \mathbb{R}^p \times \{\lambda I : \lambda > 0\}$, and has been extensively studied previously in [11, 14, 17, 22, 29].

3.2.1. *LRT.* This subsection studies the LRT for (3.8). The (re-scaled) log-likelihood ratio statistic for (3.8) is defined by (cf. [23], Theorem 8.3.2)

$$T_{\mathrm{LRT},s}(X) \equiv \frac{N}{2}\big(p \log \mathrm{tr}(S) - \log\det S - p \log p\big). \tag{3.9}$$

Evidently, the law of $T_{\mathrm{LRT},s}(X)$ does not depend on the λ in (3.8) and hence is invariant under H_0. Thus the general principle in Theorem 2.1 applies due to regularity of $T_{\mathrm{LRT};s}$ (see Appendix B). We will use $(m_{\Sigma;\mathrm{LRT},s}, \sigma_{\Sigma;\mathrm{LRT},s}, V_{\Sigma;\mathrm{LRT},s})$ to represent their generic versions defined in (1.4) and (2.5).

For a symmetric $p \times p$ matrix M, let

$$b_\ell(M) \equiv p^{-1}\,\mathrm{tr}\big(M^\ell\big), \qquad b(M) \equiv b_1(M). \tag{3.10}$$

The next proposition establishes a quantitative CLT for $T_{\mathrm{LRT},s}(X)$; its proof is presented in Section 6.1.2. Recall that $T_{\mathrm{LRT},s}$ is nondegenerate only if $p \le n - 1 = N$.

PROPOSITION 3.7. *Suppose $p/N \le 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Then there exists some $C = C(\varepsilon) > 0$ such that under H_0,*

$$d_{\mathrm{TV}}\biggl(\frac{T_{\mathrm{LRT},s}(X) - m_{I;\mathrm{LRT},s}}{\sigma_{I;\mathrm{LRT},s}}, \mathcal{N}(0,1)\biggr) \le \frac{C}{p}.$$

The CLT for $T_{\mathrm{LRT},s}(X)$ was previously derived in [17], Theorem 1, under the asymptotics $y \in (0,1]$. The quantitative CLT above does not require p to grow proportionally to N but excludes the boundary case $y = 1$.

The following result establishes the contiguity condition (2.6) for $T_{\mathrm{LRT},s}$; see Section 6.1.3 for its proof.

PROPOSITION 3.8. *Suppose Σ is nonsingular. The following hold*:

1. *There exists some absolute constant $C_1 > 0$ such that*
$$V^2_{\Sigma;\mathrm{LRT},s} \le C_1 N \bigl\| \Sigma \cdot b^{-1}(\Sigma) - I \bigr\|_F^2$$
holds for N, p large enough.
2. *The mean difference is given by*
$$m_{\Sigma;\mathrm{LRT},s} - m_{I;\mathrm{LRT},s} = \frac{N}{2}\bigl[-\log\det\bigl(\Sigma \cdot b^{-1}(\Sigma)\bigr) + Q_{\mathrm{LRT},s}\bigl(\Sigma \cdot b^{-1}(\Sigma)\bigr)\bigr].$$
Here
$$\bigl|Q_{\mathrm{LRT},s}\bigl(\Sigma \cdot b^{-1}(\Sigma)\bigr)\bigr| \le C_2 N^{-1} b\bigl[\bigl(\Sigma \cdot b^{-1}(\Sigma)\bigr)^2\bigr] \tag{3.11}$$
for some absolute constant $C_2 > 0$.
3. *In the asymptotic regime $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < 1$,*
$$\sigma^2_{I;\mathrm{LRT},s} \sim \frac{N^2}{2}\biggl[-\frac{p}{N} - \log\Bigl(1 - \frac{p}{N}\Bigr)\biggr].$$
4. *There exists some absolute constant $C_3 > 0$ such that*
$$\frac{V_{\Sigma;\mathrm{LRT},s}}{|m_{\Sigma;\mathrm{LRT},s} - m_{I;\mathrm{LRT},s}| \vee \sigma_{I;\mathrm{LRT},s}} \le \frac{C_3}{(\sigma_{I;\mathrm{LRT},s} \wedge N)^{1/2}}.$$

There is a genuine difference between the above contiguity result and the previous ones studied in Section 3.1.1, in that a closed-form formula for the mean difference $m_{\Sigma;\mathrm{LRT},s} - m_{I;\mathrm{LRT},s}$ is no longer available. One therefore has to work with strong enough upper bounds for the "residual term" $Q_{\mathrm{LRT},s}(\Sigma \cdot b^{-1}(\Sigma))$, the removal of which constitutes the main technicalities in the proofs; see Section 6.1.3 for details.

Let $\Psi_{\mathrm{LRT},s}(X)$ be the test built from (1.10) and the statistic in (3.9). Combining the above results with Theorem 2.1 and some additional efforts to remove the residual term $Q_{\mathrm{LRT},s}(\Sigma \cdot b^{-1}(\Sigma))$, we have the following asymptotic power formula for $\Psi_{\mathrm{LRT},s}(X)$; see Section 6.1.4 for its proof.

THEOREM 3.9. *Suppose $p/N \le 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Then there exists some constant $C = C(\varepsilon,\alpha) > 0$ such that*

$$\biggl|\mathbb{E}_\Sigma \Psi_{\mathrm{LRT},s}(X) - \mathbb{P}\biggl(\mathcal{N}\biggl(-\frac{N\log\det(\Sigma \cdot b^{-1}(\Sigma))}{2\sigma_{I;s}}, 1\biggr) > z_\alpha\biggr)\biggr| \le C \cdot p^{-1/3}.$$

Consequently, in the asymptotic regime $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < 1$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{LRT},s}(X) \sim 1 - \Phi\biggl(z_\alpha - \frac{-\log\det(\Sigma \cdot b^{-1}(\Sigma))}{\sqrt{2(-\frac{p}{N} - \log(1-\frac{p}{N}))}}\biggr).$$

To the best of our knowledge, in the high-dimensional regime $N \wedge p \to \infty$, the LRT for (3.8) was only studied in [16, 17], where formal theory was missing on the power behavior of $\Psi_{\mathrm{LRT},s}$. Theorem 3.9 fills this gap.

3.2.2. *John's test.* Consider testing (3.8) using the (rescaled) John's trace statistic [18]:

$$T_{\mathrm{J}}(X) \equiv \frac{N}{4}\operatorname{tr}\left[\left(\frac{S}{p^{-1}\operatorname{tr}(S)} - I\right)^2\right]. \tag{3.12}$$

Clearly the law of $T_J(X)$ is invariant under H_0, and the above statistic is nondegenerate for all configurations of (n, p). The general principle in Theorem 2.1 thereby applies in view of the regularity of T_{J} (see Appendix B). We will use $(m_{\Sigma;\mathrm{J}}, \sigma_{\Sigma;\mathrm{J}}, V_{\Sigma;\mathrm{J}})$ to represent their generic versions defined in (1.4) and (2.1).

The next proposition establishes a quantitative CLT for $T_{\mathrm{J}}(X)$ under H_0; its proof is given in Section 6.2.2.

PROPOSITION 3.10. *There exists some absolute constant $C > 0$, such that, under H_0,*

$$d_{\mathrm{TV}}\left(\frac{T_{\mathrm{J}}(X) - m_{I;\mathrm{J}}}{\sigma_{I;\mathrm{J}}}, \mathcal{N}(0,1)\right) \le \frac{C}{N \wedge p}.$$

CLTs for $T_{\mathrm{J}}(X)$ under H_0 in high dimensions are first obtained in [22]. We improve these results both in terms of nonasymptotic normal approximation bound and the removal of the condition $0 < \underline{\lim}(p/N) \le \overline{\lim}(p/N) < \infty$.

The following result establishes the contiguity condition (2.6) for T_{J}; its proof is presented in Section 6.2.3. Recall the definition of $b(\Sigma)$ in (3.10).

PROPOSITION 3.11. *Suppose $p/N \le M$ for some $M > 1$. Then the following hold for N larger than a big enough absolute constant*:

1. *There exists some constant $C_1 = C_1(M) > 0$ such that*

$$V_{\Sigma;\mathrm{J}}^2 \le C_1 \cdot N\big(\|\Sigma \cdot b^{-1}(\Sigma)\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2.$$

2. *The mean difference is given by*

$$m_{\Sigma;\mathrm{J}} - m_{I;\mathrm{J}} = \frac{N}{4}\big[\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2 + Q_{\mathrm{J}}(\Sigma \cdot b^{-1}(\Sigma))\big].$$

   Here

$$\big|Q_{\mathrm{J}}(\Sigma \cdot b^{-1}(\Sigma))\big| \le C_2 \cdot N^{-1/2}\big(p^{-1}\|\Sigma \cdot b^{-1}(\Sigma)\|_F^2 + 1\big)\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F$$

   for some $C_2 = C_2(M) > 0$.

3. *In the asymptotic regime $N \wedge p \to \infty$,*

$$\sigma_{I;\mathrm{J}}^2 \sim \frac{p^2}{4}.$$

4. *There exists some $C_3 = C_3(M) > 0$ such that*

$$\frac{V_{\Sigma;\mathrm{J}}}{|m_{\Sigma;\mathrm{J}} - m_{I;\mathrm{J}}| \vee \sigma_{I;\mathrm{J}}} \le \frac{C_3}{p^{1/2}}.$$

The proof of the above contiguity is the most complicated among the examples studied in this paper. The main complication is due to the existence of the $\operatorname{tr}(S)$ term in the denominator in (3.12), which leads to the complications both in the control of $V_{\Sigma;J}^2$ and the "residual term" $Q_{\mathrm{J}}(\Sigma \cdot b^{-1}(\Sigma))$.

Let $\Psi_{\mathrm{J}}(X)$ be the test built from (1.10) and the statistic in (3.12). Now we have the following theorem.

THEOREM 3.12. *Suppose $p/N \le M$ for some $M > 1$. Then there exists some constant $C = C(\alpha, M) > 0$ such that*

$$\left|\mathbb{E}_\Sigma \Psi_{\mathrm{J}} - \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot \|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2}{4\sigma_{I;\mathrm{J}}}, 1\right) > z_\alpha\right)\right| \le C \cdot p^{-1/3}.$$

Consequently, in the asymptotic regime $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < \infty$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{J}} \sim 1 - \Phi\left(z_\alpha - \frac{\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2}{2(p/N)}\right).$$

See Section 6.2.4 for the proof. The power behavior for John's test is previous studied in [25, 26, 31] for a special class of alternatives under the spiked covariance model with a fixed number of spikes; see Section 3.3 ahead for a detailed discussion. To the best of our knowledge, the theorem above gives the first complete characterization of the power behavior for John's test for arbitrary alternatives in the high-dimensional regime $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < \infty$.

3.3. *Case study*: *Spiked covariance models*. In this subsection, we consider a special class of alternatives known as the spiked covariance model [19]:

$$\Sigma(a) \equiv \mathrm{diag}(1 + a_1, \ldots, 1 + a_p), \tag{3.13}$$

where $a = (a_1, \ldots, a_p) \in (-1, \infty)^p$. Write $\bar{a} = \sum_{j=1}^p a_j/p$. Specializing the results obtained in Sections 3.1 and 3.2, we have the following.

COROLLARY 3.13. *The following hold.*

1. *The power for the likelihood ratio test of $\Sigma = I$ satisfies*

$$\mathbb{E}_{\Sigma(a)} \Psi_{\mathrm{LRT}} \sim 1 - \Phi\left(z_\alpha - \frac{\sum_{j=1}^p (a_j - \log(1 + a_j))}{\sqrt{2(-\frac{p}{N} - \log(1 - \frac{p}{N}))}}\right) \equiv \beta_{\mathrm{LRT}}(a),$$

   under $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < 1$.
2. *The power for Ledoit–Nagao–Wolf test of $\Sigma = I$ satisfies*

$$\mathbb{E}_{\Sigma(a)} \Psi_{\mathrm{LNW}} \sim 1 - \Phi\left(z_\alpha - \frac{\sum_{j=1}^p a_j^2}{2(p/N)}\right) \equiv \beta_{\mathrm{LNW}}(a),$$

   under $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < \infty$.
3. *The power for the likelihood ratio test of $\Sigma = \lambda I$ satisfies*

$$\mathbb{E}_{\Sigma(a)} \Psi_{\mathrm{LRT};s} \sim 1 - \Phi\left(z_\alpha - \frac{\sum_{j=1}^p \log \frac{1+\bar{a}}{1+a_j}}{\sqrt{2(-\frac{p}{N} - \log(1 - \frac{p}{N}))}}\right) \equiv \beta_{\mathrm{LRT};s}(a),$$

   under $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < 1$.
4. *The power for John's test of $\Sigma = \lambda I$ satisfies*

$$\mathbb{E}_{\Sigma(a)} \Psi_{\mathrm{J}} \sim 1 - \Phi\left(z_\alpha - \frac{\sum_{j=1}^p (a_j - \bar{a})^2/(1 + \bar{a})^2}{2(p/N)}\right) \equiv \beta_{\mathrm{J}}(a),$$

   under $N \wedge p \to \infty$ with $\overline{\lim}(p/N) < \infty$.

(1), (2) and (4) above recover [26], Proposition 8 (i)–(ii), while (3)–(4) above recover [31], equations (4.5) and (4.8). Both [25, 31] considered the case where $r \equiv \|a\|_0$ and the nonzero elements of a are fixed. The techniques in [26] work with a further restriction $\|a\|_\infty < \sqrt{y}$ where y is the limiting value of the ratio p/N. As mentioned in the Introduction, this restriction coincides with the Baik–Ben Arous–Péché (BBP) phase transition [3], and is essential for the techniques of [26], due to the singular nature of the likelihood ratio process when $\|a\|_\infty > \sqrt{y}$ already in the case $r = 1$, see [25], Theorem 8. The restriction $\|a\|_\infty < \sqrt{y}$ is removed in [31] for the likelihood ratio test $\Psi_{\mathrm{LRT};s}$ and John's test Ψ_{J} for sphericity, by variations of Bai–Silverstein techniques developed in [1, 2].

It is easy to see that in the setting of [25, 31] with a fixed number of spikes as described above, the asymptotic powers are the same for the following two group of tests:

1. Likelihood ratio tests Ψ_{LRT}, $\Psi_{\mathrm{LRT};s}$: $\beta_{\mathrm{LRT}} = \beta_{\mathrm{LRT};s}$.
2. Ledoit–Nagao–Wolf and John's tests: $\beta_{\mathrm{LNW}} = \beta_{\mathrm{J}}$.

Clearly, neither group of tests universally dominates the other in terms of the power behavior. For instance, the power of tests in (1) dominates that of (2) when some of a_j's are close to -1 (i.e., Σ is near singular), while the reversed phenomenon occurs when some of a_j's are close to ∞.

In general, the asymptotic power equivalence of the above two groups may not hold when the number of spikes are no longer fixed. Instead, we have the following power ordering within each group.

COROLLARY 3.14.

1. *Likelihood ratio tests* Ψ_{LRT}, $\Psi_{\mathrm{LRT};s}$ *have the power ordering*

$$\beta_{\mathrm{LRT}}(a) \geq \beta_{\mathrm{LRT};s}(a).$$

2. *Ledoit–Nagao–Wolf and John's tests* Ψ_{LNW}, Ψ_{J} *have the power ordering*

$$\beta_{\mathrm{LNW}}(a) \begin{cases} \geq \beta_{\mathrm{J}}(a), & \overline{a^2}\big(1-(1+\bar{a})^2\big) \leq \bar{a}^2; \\ < \beta_{\mathrm{J}}(a), & \overline{a^2}\big(1-(1+\bar{a})^2\big) > \bar{a}^2. \end{cases}$$

Here $\overline{a^2} \equiv \sum_{j=1}^p a_j^2/p$.

PROOF. (1) follows from the inequality $\sum_{j=1}^p \log(1+\bar{a}) \leq \sum_{j=1}^p \bar{a} = \sum_{j=1}^p a_j$. (2) follows by the following calculation:

$$\frac{\sum_{j=1}^p (a_j - \bar{a})^2}{(1+\bar{a})^2} = \frac{\sum_{j=1}^p (a_j^2 - \bar{a}^2)}{(1+\bar{a})^2} = \sum_{j=1}^p a_j^2 + \frac{\sum_{j=1}^p a_j^2 \cdot (1-(1+\bar{a})^2) - p\bar{a}^2}{(1+\bar{a})^2}.$$

The proof is complete. □

Note that $\{\bar{a} \geq 0\} \subsetneq \{\overline{a^2}(1-(1+\bar{a})^2) \leq \bar{a}^2\}$ (the inclusion is in fact proper), so if $\bar{a} \geq 0$, John's test Ψ_{J} will be less powerful than Ledoit–Nagao–Wolf's Ψ_{LNW}. Furthermore, both inequalities in the above corollary can be strict asymptotically, and similar to the discussion above, there are no universal power dominance relationships between the tests in the two groups.

4. Some spectral estimates. In this section, we will prove some spectral estimate for a class of high-dimensional random matrices that will be useful for the proofs of the results in Section 3.

First we introduce some convention on notation: Let $\mathcal{I}_1, \mathcal{I}_2$ be finite index sets. For $A = (A_{\iota_1,\iota_2})_{\iota_1\in\mathcal{I}_1,\iota_2\in\mathcal{I}_2} \in \mathbb{R}^{\mathcal{I}_1\times\mathcal{I}_2}$, its operator norm is defined as

$$\|A\|_{\mathrm{op}} \equiv \sup_{v\in B_{\mathcal{I}_2}(1)} \|Av\|_{\ell_2(\mathbb{R}^{\mathcal{I}_2})}. \tag{4.1}$$

It can be readily verified that $\|A\|_{\mathrm{op}} = \sup_{u\in B_{\mathcal{I}_1}, v\in B_{\mathcal{I}_2}} \langle u, Av\rangle_{\mathcal{I}_1}$, and for a symmetric matrix $A \in \mathbb{R}^{\mathcal{I}_1\times\mathcal{I}_1}$, $\|A\|_{\mathrm{op}} = \sup_{u\in B_{\mathcal{I}_1}} |\langle u, Au\rangle_{\mathcal{I}_1}|$. Here $\langle\cdot,\cdot\rangle_{\mathcal{I}_1}$ is the standard inner product on $\mathbb{R}^{\mathcal{I}_1}$. Clearly, the definition of the operator norm does not depend on the choice of the ordering of the index sets.

Under this notational convention, with the index set $\Lambda \equiv \{(ij) : i \in [N], j \in [p]\}$, we present below two results on the spectral norm of some special $\Lambda\times\Lambda$ matrices that are crucial to the proof of the quantitative CLTs. We do not specify a particular ordering on Λ as we will be only interested in the operator norm as defined above. In the following we use $\mathbb{N}$ to denote the set of natural numbers. Recall the data matrix $X = [X_1, \ldots, X_N]^\top \in \mathbb{R}^{N\times p}$ and the definition of S in (3.2).

PROPOSITION 4.1.

1. *Suppose $p/N \le 1-\varepsilon$ for some $\varepsilon > 0$. For $\ell, m \in \mathbb{N}$ such that $\ell + m \ge 1$, let $U_{\ell,m} \in \mathbb{R}^{\Lambda\times\Lambda}$ be defined by*

$$(U_{\ell,m})_{(ij),(i'j')} \equiv N^{-1} X_i^\top S^{-\ell} X_{i'} \big(S^{-m}\big)_{jj'}. \tag{4.2}$$

   Then for any $q \in \mathbb{N}$, there exists some $C = C(\varepsilon, \ell, m, q) > 0$ such that $\mathbb{E}\|U_{\ell,m}\|_{\mathrm{op}}^q \le C$ for $p \ge C$.
2. *When X_i, S and N is replaced by $X_i - \bar{X}$, S_* and n, the conclusion of* (1) *still holds.*

When the inverse S^{-1} in (4.2) is replaced by S, the condition on p/N can be substantially relaxed.

PROPOSITION 4.2. *Let $y \equiv p/N$. For $\ell, m \in \mathbb{N}$, let $U_{\ell,m;+} \in \mathbb{R}^{\Lambda\times\Lambda}$ be defined by*

$$(U_{\ell,m;+})_{(ij),(i'j')} \equiv N^{-1} X_i^\top S^{\ell} X_{i'} \big(S^{m}\big)_{jj'}. \tag{4.3}$$

Then for any $q \in \mathbb{N}$, there exists some $C = C(\ell, m, q) > 0$ such that $\mathbb{E}\|U_{\ell,m;+}\|_{\mathrm{op}}^q \le C(\sqrt{y} \vee y)^{q(\ell+m+1)}$.

The proof of Proposition 4.1 relies crucially on the following stable moment estimate for $\|S^{-1}\|_{\mathrm{op}}$. Its proof utilizes two main technical tools: (i) rigidity estimates on the eigenvalues of the sample covariance matrix (cf. [27]); (ii) closed form distributional formula of sample eigenvalues via zonal polynomials [23], Chapter 9.7.

LEMMA 4.3. *Let $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$ where Z_i's are i.i.d. $\mathcal{N}(0, I)$ in $\mathbb{R}^p$. Suppose $p/N \le 1-\varepsilon$ for some fixed $\varepsilon > 0$ and every N, $p \ge 2$. Then for any positive integer $q \le (N-p-1)/8$, we have $\mathbb{E}\|S_Z^{-1}\|_{\mathrm{op}}^q \le C$ for some positive $C = C(\varepsilon, q)$.*

PROOF. Write S_Z for S in the proof for simplicity. Let λ be the smallest eigenvalue of S, and $y \equiv (p-1)/N < 1-\varepsilon$. By [28], Theorem 1.1, on an event E with probability at least $1 - e^{-cN(1-y)}$, $\lambda \ge c(1-\sqrt{y})^2$ for some absolute constant $c > 0$. A similar estimate can

be obtained using rigidity estimate for the eigenvalues of the sample covariance matrix, for example, [27], Theorem 3.1(iii). Hence

$$\begin{aligned}\mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q} &= \mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q}\mathbf{1}_{E} + \mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q}\mathbf{1}_{E^c} \\ &\le c^{-q}(1-\sqrt{y})^{-2q} + \mathbb{E}^{1/2}\|S^{-1}\|_{\mathrm{op}}^{2q}\cdot e^{-cN(1-y)/2}.\end{aligned} \tag{4.4}$$

Now we give an upper bound for $\mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{2q}$. Let $r \equiv (N-p-1)/2$ assumed to be a positive integer. For any nonnegative integer k, we write $\kappa \vdash k$ if $\kappa = (k_1, k_2, \ldots)$, with convention $k_1 \ge k_2 \ge \ldots$, is a partition of k, that is, $\sum_i k_i = k$. Let C_κ denote the zonal polynomial (cf. [23], Chapter 7) with respect to the partition κ. Then it follows from [23], Corollary 9.7.4, that, for any $x > 0$,

$$\begin{aligned}\mathbb{P}\big(\|S^{-1}\|_{\mathrm{op}} > x\big) &= 1 - \mathbb{P}(\lambda > 1/x) \\ &= 1 - e^{-\frac{Np}{2x}}\sum_{k=0}^{pr}\sum_{\kappa\vdash k:k_1\le r}\frac{C_\kappa(NI/(2x))}{k!} \\ &= e^{-\frac{Np}{2x}}\left[\sum_{k=0}^{\infty}\frac{(Np/(2x))^k}{k!} - \sum_{k=0}^{pr}\sum_{\kappa\vdash k:k_1\le r}\frac{C_\kappa(NI/(2x))}{k!}\right] \\ &= e^{-\frac{Np}{2x}}\left[\sum_{k=pr+1}^{\infty}\frac{(Np/(2x))^k}{k!} + \sum_{k=0}^{pr}\frac{1}{k!}\left\{\left(\frac{Np}{2x}\right)^k - \sum_{\kappa\vdash k:k_1\le r}C_\kappa\left(\frac{NI}{2x}\right)\right\}\right] \\ &\overset{(*)}{=} e^{-\frac{Np}{2x}}\left[\sum_{k=pr+1}^{\infty}\frac{(Np/(2x))^k}{k!} + \sum_{k=0}^{pr}\frac{1}{k!}\sum_{\kappa\vdash k:k_1>r}C_\kappa\left(\frac{NI}{2x}\right)\right] \\ &\overset{(**)}{=} e^{-\frac{Np}{2x}}\left[\sum_{k=pr+1}^{\infty}\frac{(Np/(2x))^k}{k!} + \sum_{k=r+1}^{pr}\frac{(N/(2x))^k}{k!}\sum_{\kappa\vdash k:k_1>r}C_\kappa(I))\right] \\ &\overset{(***)}{\le} e^{-\frac{Np}{2x}}\cdot\sum_{k=r+1}^{\infty}\frac{(Np/(2x))^k}{k!}.\end{aligned}$$

Here $(*)$ follows from [23], Definition 7.2.1, (iii): for any $k \ge 0$ and $t > 0$,

$$\sum_{\kappa\vdash k}C_\kappa(t\cdot I) = \big[\mathrm{tr}(t\cdot I)\big]^k = (tp)^k; \tag{4.5}$$

$(**)$ follows from the fact that for each k and partition κ of k, C_κ is a homogeneous polynomial of order k; $(***)$ follows from the nonnegativity of zonal polynomial for I (cf. [23], Corollary 7.2.4) and an application of (4.5) with $t = 1$:

$$\sum_{\kappa\vdash k:k_1>r}C_\kappa(I) \le \sum_{\kappa\vdash k}C_\kappa(I) = p^k.$$

Hence by using the fact that for any $k \ge 2q+1$,

$$\begin{aligned}\int_0^\infty e^{-\frac{Np}{2x}}\left(\frac{Np}{2x}\right)^k\cdot x^{2q-1}\,\mathrm{d}x &= \left(\frac{Np}{2}\right)^{2q}\int_0^\infty e^{-y}y^{k-2q-1}\,\mathrm{d}y \\ &= \left(\frac{Np}{2}\right)^{2q}(k-2q-1)!,\end{aligned}$$

we have for every $r \geq 4q$

$$\begin{aligned}\mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{2q} &= 2q\int_0^\infty x^{2q-1}\mathbb{P}(\|S^{-1}\|_{\mathrm{op}} > x)\,\mathrm{d}x \\ &\leq \left(\frac{Np}{2}\right)^{2q}\sum_{k=r+1}^{\infty}\frac{1}{k(k-1)\cdots(k-2q)} \\ &= \left(\frac{Np}{2}\right)^{2q}\sum_{k=r+1}^{\infty}\frac{1}{2q}\cdot\left\{\frac{1}{(k-1)\cdots(k-2q)} - \frac{1}{k\cdots(k-2q-1)}\right\} \\ &= \left(\frac{Np}{2}\right)^{2q}\frac{1}{2q}\frac{1}{r(r-1)\cdots(r-2q+1)} \lesssim_q \frac{(Np)^{2q}}{r^{2q}}.\end{aligned} \tag{4.6}$$

Combining (4.4) and (4.6), as $p/N \leq 1-\varepsilon$,

$$\mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q} \leq C_\varepsilon^q + C_\varepsilon N^q e^{-c_\varepsilon N} \lesssim_{q,\varepsilon} 1 \tag{4.7}$$

with $r = (N-p-1)/2$ being a positive integer. If r is not an integer, write $S = \frac{N-1}{N}S' + \frac{1}{N}X_N X_N^\top$, where $S' \equiv \frac{1}{N-1}\sum_{i=1}^{N-1} X_i X_i^\top$. Then using Sherman–Morrison formula,

$$\begin{aligned}S^{-1} &= \frac{N}{N-1}(S')^{-1} - \frac{N}{(N-1)^2}\cdot\frac{(S')^{-1}X_N X_N^\top (S')^{-1}}{1+\frac{1}{N-1}X_N^\top (S')^{-1}X_N} \\ &\equiv \frac{N}{N-1}(S')^{-1} - R.\end{aligned} \tag{4.8}$$

As $X_N^\top (S')^{-1}X_N \geq \|X_N\|^2/\lambda_{\max}(S')$, we have

$$\begin{aligned}\mathbb{E}\|R\|_{\mathrm{op}}^q &\leq \left(\frac{N}{N-1}\right)^q \cdot \mathbb{E}\left[\|(S')^{-1}X_N X_N^\top (S')^{-1}\|_{\mathrm{op}}^q \frac{\lambda_{\max}(S')^{2q}}{\|X_N\|^{2q}}\right] \\ &\lesssim_q \mathbb{E}\left(\frac{\lambda_{\max}(S')^{2q}}{\lambda_{\min}(S')^{2q}}\right) \leq \mathbb{E}^{1/2}\|S'\|_{\mathrm{op}}^{4q}\cdot\mathbb{E}^{1/2}\|(S')^{-1}\|_{\mathrm{op}}^{4q} \lesssim_{q,\varepsilon} 1.\end{aligned}$$

The claim for r not being an integer follows from the decomposition (4.7) and the estimate above. □

The following corollary of the Koltchinskii–Lounici theorem [20] will also be repeatedly used.

LEMMA 4.4. *Let $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$ where Z_i's are i.i.d. $\mathcal{N}(0, I)$ in $\mathbb{R}^p$. Then for any positive integer q, there exists some positive $C = C(q)$ such that*

$$\mathbb{E}\|S_Z - I\|_{\mathrm{op}}^q \leq C\cdot\left(\sqrt{\frac{p}{N}} \vee \frac{p}{N}\right)^q.$$

PROOF. This is a direct consequence of [20], Corollary 2. □

Now we prove Propositions 4.1 and 4.2.

PROOF OF PROPOSITION 4.1. We only prove (1); claim (2) follows from completely same arguments by noting that (4.9) and (4.11) below still hold with the prescribed substitution. Note that $U_{\ell,m}$ is symmetric in that $(U_{\ell,m})_{(ij)(i'j')} = (U_{\ell,m})_{(i'j')(ij)}$, and satisfies that

for any nonnegative integers (ℓ_1, m_1) and (ℓ_2, m_2) such that $(\ell_1 + m_1) \wedge (\ell_2 + m_2) \geq 1$,

$$
\begin{aligned}
&(U_{\ell_1,m_1} \cdot U_{\ell_2,m_2})_{(ij),(i'j')} \\
&= N^{-2} \sum_{(\bar{i}\,\bar{j})} X_i^\top S^{-\ell_1} X_{\bar{i}} \big(S^{-m_1}\big)_{j\bar{j}} X_{\bar{i}}^\top S^{-\ell_2} X_{i'} \big(S^{-m_2}\big)_{\bar{j}j'} \\
&= N^{-2} X_i^\top S^{-\ell_1} \Big(\sum_{\bar{i}} X_{\bar{i}} X_{\bar{i}}^\top\Big) S^{-\ell_2} X_{i'} \cdot \Big(\sum_{\bar{j}} \big(S^{-m_1}\big)_{j\bar{j}} \big(S^{-m_2}\big)_{\bar{j}j'}\Big) \qquad (4.9) \\
&= N^{-1} X_i^\top S^{-(\ell_1+\ell_2-1)} X_{i'} \big(S^{-(m_1+m_2)}\big)_{jj'} \\
&= (U_{\ell_1+\ell_2-1,m_1+m_2})_{(ij),(i'j')}.
\end{aligned}
$$

Consequently the above argument entails that for any $q \in \mathbb{N}$, $(U_{\ell,m})^q = U_{\ell',m'}$ with

$$
\ell' \equiv \ell'(q) \equiv q(\ell - 1) + 1, \qquad m' \equiv m'(q) \equiv qm. \tag{4.10}
$$

Using that $\|U_{\ell,m}\|_{\mathrm{op}} = \sup_{u \in B_{N\times p}(1)} |\sum_{(ij)(i'j')} u_{ij} (U_{\ell,m})_{(ij),(i'j')} u_{i'j'}|$, we have

$$
\begin{aligned}
\|U_{\ell,m}\|_{\mathrm{op}} &= \sup_{u \in B_{N\times p}(1)} N^{-1} \Big| \sum_{i,i',j,j'} X_i^\top S^{-\ell} X_{i'} \big(S^{-m}\big)_{jj'} u_{ij} u_{i'j'} \Big| \\
&= N^{-1} \sup_{u \in B_{N\times p}(1)} \Big| \sum_{i,i'} X_i^\top S^{-\ell} X_{i'} \cdot \sum_{jj'} u_{ij} \cdot \big(S^{-m}\big)_{jj'} u_{i'j'} \Big| \\
&= N^{-1} \sup_{u \in B_{N\times p}(1)} \Big| \sum_{i,i'} X_i^\top S^{-\ell} X_{i'} \cdot \big[u \cdot S^{-m} \cdot u^\top\big]_{ii'} \Big|.
\end{aligned}
$$

As the (i, i')th entry of $X^\top S^{-\ell} X$ is $X_i^\top S^{-\ell} X_{i'}$ and that $\mathrm{tr}(AB) = \sum_{i,i'=1}^N A_{ii'} B_{ii'}$ for two symmetric matrices in $\mathbb{R}^{N\times N}$, we have

$$
\begin{aligned}
\|U_{\ell,m}\|_{\mathrm{op}} &= N^{-1} \sup_{u \in B_{N\times p}(1)} \big|\mathrm{tr}\big[X S^{-\ell} X^\top \cdot u S^{-m} u^\top\big]\big| \\
&= N^{-1} \sup_{u \in B_{N\times p}(1)} \big|\mathrm{tr}\big[\big(u^\top X\big) S^{-\ell} \big(X^\top u\big) \cdot S^{-m}\big]\big|.
\end{aligned}
$$

Further using twice the fact that $\mathrm{tr}(AB) \leq \mathrm{tr}(A)\|B\|_{\mathrm{op}}$ for any two p.s.d. and symmetric matrices A, B, we arrive at

$$
\begin{aligned}
\|U_{\ell,m}\|_{\mathrm{op}} &\leq N^{-1} \big\|S^{-m}\big\|_{\mathrm{op}} \big\|S^{-\ell}\big\|_{\mathrm{op}} \cdot \sup_{u \in B_{N\times p}(1)} \big|\mathrm{tr}\big(XX^\top uu^\top\big)\big| \\
&\leq \big\|S^{-(\ell+m)}\big\|_{\mathrm{op}} \cdot N^{-1} \big\|XX^\top\big\|_{\mathrm{op}} \cdot \sup_{u \in B_{N\times p}(1)} \mathrm{tr}\big(uu^\top\big) \qquad (4.11) \\
&= \big\|S^{-(\ell+m)}\big\|_{\mathrm{op}} \cdot \|S\|_{\mathrm{op}} = \big\|S^{-1}\big\|_{\mathrm{op}}^{\ell+m-1}.
\end{aligned}
$$

Hence for any $q \in \mathbb{N}$, equations (4.9)–(4.11) and Lemma 4.3 entail that

$$
\begin{aligned}
\mathbb{E}\|U_{\ell,m}\|_{\mathrm{op}}^q &= \mathbb{E}\big\|U_{\ell,m}^q\big\|_{\mathrm{op}} = \mathbb{E}\|U_{\ell',m'}\|_{\mathrm{op}} \leq \mathbb{E}\big\|S^{-1}\big\|_{\mathrm{op}}^{\ell'+m'-1} \\
&= \mathbb{E}\big\|S^{-1}\big\|_{\mathrm{op}}^{q(\ell+m-1)} \leq C_{\ell,m,q},
\end{aligned}
$$

completing the proof. □

PROOF OF PROPOSITION 4.2. The proof largely follows that of Proposition 4.1 with modifications. We sketch the difference below. Using the same calculations as in (4.9), we have

$$\begin{aligned}
&(U_{\ell_1,m_1;+} \cdot U_{\ell_2,m_2;+})_{(ij),(i'j')} \\
&\quad = N^{-2} \sum_{(\bar{i}\,\bar{j})} X_i^\top S^{\ell_1} X_{\bar{i}} \big(S^{m_1}\big)_{j\bar{j}} X_{\bar{i}}^\top S^{\ell_2} X_{i'} \big(S^{m_2}\big)_{\bar{j}j'} \\
&\quad = N^{-2} X_i^\top S^{\ell_1} \Big(\sum_{\bar{i}} X_{\bar{i}} X_{\bar{i}}^\top \Big) S^{\ell_2} X_{i'} \cdot \Big(\sum_{\bar{j}} \big(S^{m_1}\big)_{j\bar{j}} \big(S^{m_2}\big)_{\bar{j}j'} \Big) \\
&\quad = N^{-1} X_i^\top S^{\ell_1+\ell_2+1} X_{i'} \big(S^{m_1+m_2}\big)_{jj'} = (U_{\ell_1+\ell_2+1,m_1+m_2;+})_{(ij),(i'j')}.
\end{aligned}$$

Hence for any $q \in \mathbb{N}$, $(U_{\ell,m;+})^q = U_{\ell',m';+}$ with ℓ' now defined by $\ell' \equiv \ell'(q) \equiv \ell + (q-1)(\ell+1)$ and $m' = qm$ remains the same as in (4.10).

Then using the same arguments as in (4.11), we have $\|U_{\ell,m;+}\|_{\mathrm{op}} \le \|S\|_{\mathrm{op}}^{\ell+m+1}$, hence for any $q \in \mathbb{N}$,

$$\mathbb{E}\|U_{\ell,m;+}\|_{\mathrm{op}}^q \le \mathbb{E}\|S\|_{\mathrm{op}}^{\ell'+m'+1} \lesssim_{\ell,m,q} (\sqrt{y} \vee y)^{q(\ell+m+1)},$$

where the last inequality follows by Lemma 4.4 and the fact that $\ell' + m' + 1 = \ell + (q-1)(\ell+1) + qm + 1 = q(\ell+m+1)$. □

5. Proofs for Section 3.1 (testing identity).

5.1. *Proofs for Section 3.1.1 (LRT).* Recall $\Lambda = \{(ij) : i \in [N], j \in [p]\}$. In the following sections, for a sufficiently smooth function $T : \mathbb{R}^\Lambda \to \mathbb{R}$, its gradient $\nabla T : \mathbb{R}^\Lambda \to \mathbb{R}^\Lambda$ and Hessian $\nabla^2 T : \mathbb{R}^\Lambda \to \mathbb{R}^{\Lambda\times\Lambda}$ are defined respectively by

$$\big(\nabla T(x)\big)_{(ij)} \equiv \frac{\partial T}{\partial x_{(ij)}}(x) \quad \text{and} \quad \big(\nabla^2 T(x)\big)_{(ij),(i'j')} \equiv \frac{\partial^2 T}{\partial x_{(ij)} \partial x_{(i'j')}}(x),$$

with $x = (x_{(ij)}) \in \mathbb{R}^\Lambda$. Slightly abusing notation, we use $\|\nabla T(x)\|_F \equiv \|\nabla T(x)\|_{\ell_2(\mathbb{R}^\Lambda)}$. The operator norm $\|\nabla^2 T(x)\|_{\mathrm{op}}$ is defined in (4.1).

5.1.1. *Evaluation of derivatives.* In the following, we use $\{e_j\}_{j=1}^p$ to represent the canonical basis in $\mathbb{R}^p$. Let δ_{ij} be the Kronecker delta.

LEMMA 5.1. *Recall the form of $T_{\mathrm{LRT}}(X)$ in (3.4). We assume without loss of generality that $\mu = 0$. Then, for any $(i, j), (i', j') \in [N] \times [p]$:*

1. $(\nabla T_{\mathrm{LRT}}(X))_{(ij)} = (X(I - S^{-1}))_{(ij)} = e_j^\top (I - S^{-1}) X_i$.
2. $(\nabla^2 T_{\mathrm{LRT}}(X))_{(ij),(i'j')} = N^{-1} X_i^\top S^{-1} (e_{j'} X_{i'}^\top + X_{i'} e_{j'}^\top) S^{-1} e_j + \delta_{ii'} e_j^\top (I - S^{-1}) e_{j'}$.

PROOF. We use T as a shorthand for T_{LRT}.

(1). By definition, we have

$$\partial_{(ij)} T(X) = \frac{N}{2} \big(\partial_{(ij)} \operatorname{tr}(S) - \partial_{(ij)} \log\det S\big).$$

For the first partial derivative, using $\partial_{(ij)} X_k = \delta_{ik} e_j$, we have

$$\begin{aligned}
\partial_{(ij)} \operatorname{tr}\big(S(X)\big) &= N^{-1} \sum_k \frac{\partial}{\partial X_{ij}} \operatorname{tr}\big(X_k X_k^\top\big) \\
&= N^{-1} \sum_k \delta_{ik} \operatorname{tr}\big[e_j X_k^\top + X_k e_j^\top\big] = N^{-1} \sum_k \delta_{ik} \cdot 2X_{kj} = 2N^{-1} X_{ij}.
\end{aligned} \tag{5.1}$$

For the second partial derivative, using the well-known fact that $\nabla \log\det A = A^{-1}$ for any invertible and symmetric matrix A (see, e.g., [6], Section A.4.1), we have

$$\begin{aligned}\partial_{(ij)}\log\det S &= \sum_{k,\ell}\frac{\partial \log\det S}{\partial S_{k\ell}}\frac{\partial S_{k\ell}}{\partial X_{ij}} \overset{(*)}{=} \sum_{k,\ell}\big(S^{-1}\big)_{k\ell}\cdot\frac{1}{N}(\delta_{jk}X_{i\ell}+\delta_{j\ell}X_{ik})\\ &= \frac{1}{N}\sum_{\ell}\big(S^{-1}\big)_{j\ell}X_{i\ell}+\sum_{k}\big(S^{-1}\big)_{kj}X_{ik} = \frac{2}{N}\big(XS^{-1}\big)_{ij},\end{aligned} \tag{5.2}$$

where in $(*)$, we use

$$\frac{\partial S_{k\ell}}{\partial X_{ij}} = \frac{1}{N}\frac{\partial}{\partial X_{ij}}\langle Xe_k, Xe_\ell\rangle = \frac{1}{N}(\delta_{jk}X_{i\ell}+\delta_{j\ell}X_{ik}). \tag{5.3}$$

Combining (5.1) and (5.2) yields the first claim.

(2). Again by definition, we have

$$\partial_{(ij)(i'j')}T(X) = \frac{N}{2}\big(\partial_{(ij)(i'j')}\operatorname{tr}(S) - \partial_{(ij)(i'j')}\log\det S\big).$$

For the first derivative, it follows from (5.1) that

$$\partial_{(ij)(i'j')}\operatorname{tr}(S) = 2N^{-1}\partial_{(i'j')}X_{ij} = 2N^{-1}\delta_{ii'}\delta_{jj'}. \tag{5.4}$$

For the second derivative, it follows from (5.2) that

$$\begin{aligned}&\partial_{(ij)(i'j')}\log\det S\\ &= \frac{2}{N}\frac{\partial}{\partial X_{i'j'}}X_i^\top S^{-1}e_j = \frac{2}{N}\bigg(\frac{\partial X_i}{\partial X_{i'j'}}^\top S^{-1}e_j + X_i^\top\frac{\partial S^{-1}}{\partial X_{i'j'}}e_j\bigg)\\ &\overset{(**)}{=} \frac{2}{N}\big(\delta_{ii'}e_{j'}^\top S^{-1}e_j - N^{-1}X_i^\top S^{-1}\big(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top\big)S^{-1}e_j\big),\end{aligned} \tag{5.5}$$

where in $(**)$ we use the following calculation with the help of (5.3):

$$\begin{aligned}\frac{\partial S^{-1}}{\partial X_{i'j'}} &= -S^{-1}\frac{\partial S}{\partial X_{i'j'}}S^{-1} = -S^{-1}\bigg(\sum_{k\ell}e_ke_\ell^\top\frac{\partial S_{k\ell}}{\partial X_{i'j'}}\bigg)S^{-1}\\ &= -\frac{1}{N}S^{-1}\cdot\bigg(\sum_{k\ell}e_ke_\ell^\top(\delta_{j'k}X_{i'\ell}+\delta_{j'\ell}X_{i'k})\bigg)S^{-1}\\ &= -\frac{1}{N}S^{-1}\big(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top\big)S^{-1}.\end{aligned} \tag{5.6}$$

We obtain the second claim by combining (5.4) and (5.5). □

5.1.2. *Normal approximation.*

PROOF OF PROPOSITION 3.1. We again shorthand $T_{\mathrm{LRT};\Sigma}$ by T. By Lemma 5.1,

$$\|\nabla T(X)\|_F^2 = \sum_{i,j}\big(\partial_{(ij)}T(X)\big)^2 = \sum_i\big\|(I-S^{-1})X_i\big\|^2 \le \big\|S^{-1}\big\|_{\mathrm{op}}^2\|I-S\|_{\mathrm{op}}^2\sum_i\|X_i\|^2.$$

Using Lemma 4.3 and Lemma 4.4,

$$\begin{aligned}\mathbb{E}\|\nabla T(X)\|_F^4 &\lesssim \mathbb{E}\Big(\|S^{-1}\|_{\mathrm{op}}^2\|I-S\|_{\mathrm{op}}^2\sum_i\|X_i\|^2\Big)^2\\ &= \sum_{i,i'}\mathbb{E}\big[\|S^{-1}\|_{\mathrm{op}}^4\|I-S\|_{\mathrm{op}}^4\|X_i\|^2\|X_{i'}\|^2\big]\\ &\le \sum_{i,i'}\mathbb{E}^{1/4}\|S^{-1}\|_{\mathrm{op}}^{16}\cdot\mathbb{E}^{1/4}\|I-S\|_{\mathrm{op}}^{16}\cdot\mathbb{E}^{1/4}\|X_i\|^8\cdot\mathbb{E}^{1/4}\|X_{i'}\|^8\\ &\lesssim N^2\cdot\Big(\sqrt{\frac{p}{N}}\Big)^4\cdot p\cdot p = p^4.\end{aligned} \tag{5.7}$$

Again by Lemma 5.1, the second derivatives are

$$\begin{aligned}\partial_{(ij),(i'j')}T(X) &= N^{-1}X_i^\top S^{-1}e_{j'}X_{i'}^\top S^{-1}e_j + N^{-1}X_i^\top S^{-1}X_{i'}\big(S^{-1}\big)_{jj'} + \delta_{ii'}\big(I-S^{-1}\big)_{jj'}\\ &\equiv (T_1+T_2+T_3)_{(ij),(i'j')}.\end{aligned}$$

Recall the definition of $U_{\ell,m}$ in Proposition 4.1-(1). Then

$$\begin{aligned}\big(T_1^2\big)_{(ij),(i'j')} &= N^{-2}\sum_{(\bar i\bar j)}X_i^\top S^{-1}e_{\bar j}\cdot X_{\bar i}^\top S^{-1}e_j\cdot X_{\bar i}^\top S^{-1}e_{j'}\cdot X_{i'}^\top S^{-1}e_{\bar j}\\ &= N^{-2}\Big(\sum_{\bar i}e_j^\top S^{-1}X_{\bar i}X_{\bar i}^\top S^{-1}e_{j'}\Big)\cdot\Big(\sum_{\bar j}X_i^\top S^{-1}e_{\bar j}e_{\bar j}^\top S^{-1}X_{i'}\Big)\\ &= N^{-1}\big(S^{-1}\big)_{jj'}\cdot X_i^\top S^{-2}X_{i'} = (U_{2,1})_{(ij),(i'j')},\end{aligned} \tag{5.8}$$

and $T_2 = U_{1,1}$. Proposition 4.1-(1) entails that

$$\mathbb{E}\|T_1\|_{\mathrm{op}}^4 \vee \mathbb{E}\|T_2\|_{\mathrm{op}}^4 = \mathcal{O}(1). \tag{5.9}$$

On the other hand, T_3 has a block diagonal structure with respect to the index (i,i'), so its spectral norm equals that of $(I-S^{-1}) \in \mathbb{R}^{p\times p}$, and hence

$$\mathbb{E}\|T_3\|_{\mathrm{op}}^4 = \mathbb{E}\|I-S^{-1}\|_{\mathrm{op}}^4 \le \mathbb{E}\big[\|S^{-1}\|_{\mathrm{op}}^4\|I-S\|_{\mathrm{op}}^4\big] = \mathcal{O}(1). \tag{5.10}$$

Combining all the estimates above, we find that

$$\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \lesssim \mathbb{E}\|T_1\|_{\mathrm{op}}^4 + \mathbb{E}\|T_2\|_{\mathrm{op}}^4 + \mathbb{E}\|T_3\|_{\mathrm{op}}^4 = \mathcal{O}(1). \tag{5.11}$$

Let X' be an independent copy of X and let $X_t' \equiv \sqrt{t}X + \sqrt{1-t}X' \in \mathbb{R}^{N\times p}$. Let $\mathbb{E}'$ denote expectation only with respect to X' and

$$\bar T(X) \equiv \int_0^1 \frac{1}{2\sqrt{t}}\big\langle\nabla T(X), \mathbb{E}'\nabla T\big(X_t'\big)\big\rangle\,\mathrm{d}t.$$

Then by the Gaussian–Poincaré inequality

$$\mathrm{Var}\big(\bar T(X)\big) \le \mathbb{E}\|\nabla\bar T(X)\|_F^2 \lesssim \sqrt{\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4}\sqrt{\mathbb{E}\|\nabla T(X)\|_F^4} \lesssim p^2.$$

The claim now follows from the second-order Poincaré inequality in Lemma A.1 and Proposition 3.2-(4). □

5.1.3. *Contiguity.*

PROOF OF PROPOSITION 3.2. We shorthand $(T_{\mathrm{LRT}}, m_{\Sigma;\mathrm{LRT}}, \sigma_{\Sigma;\mathrm{LRT}}, V_{\Sigma;\mathrm{LRT}})$ by $(T, m_\Sigma, \sigma_\Sigma, V_\Sigma)$.

(1). Recall that $Z_1, \dots, Z_n$ are i.i.d. samples from $\mathcal{N}(0, I_p)$. By Lemma 5.1, with $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$, we have

$$\mathscr{T}_\Sigma(Z) = Z\Sigma^{1/2}\big(I - \Sigma^{-1/2}S_Z^{-1}\Sigma^{-1/2}\big)\Sigma^{1/2} = Z\big(\Sigma - S_Z^{-1}\big).$$

Hence with $\{\lambda_j\}_{j=1}^p$ denoting the eigenvalues of Σ, we have

$$\begin{aligned} V_\Sigma^2 &= \mathbb{E}\big\|\mathscr{T}_\Sigma(Z) - \mathscr{T}_I(Z)\big\|_F^2 = \mathbb{E}\big\|Z(\Sigma - I)\big\|_F^2 \\ &= \mathbb{E}\,\mathrm{tr}\big((\Sigma - I)Z^\top Z(\Sigma - I)\big) = \mathrm{tr}\big(\mathbb{E}Z^\top Z(\Sigma - I)^2\big) \\ &= N\|\Sigma - I\|_F^2 = N\sum_j(\lambda_j - 1)^2. \end{aligned}$$

(2). Note that

$$\begin{aligned} m_\Sigma &= (N/2)\mathbb{E}\big[\mathrm{tr}\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big) - \log\det\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big) - p\big] \\ &= (N/2)\big[\mathrm{tr}(\Sigma) - \log\det\Sigma - p\big] - (N/2)\mathbb{E}\log\det S_Z, \end{aligned}$$

so

$$\begin{aligned} m_\Sigma - m_I &= (N/2)\mathcal{L}_S(\Sigma, I) \\ &= (N/2)\sum_{j=1}^p(\lambda_j - \log\lambda_j - 1) \gtrsim N\sum_{j=1}^p\big[|\lambda_j - 1| \wedge (\lambda_j - 1)^2\big]. \end{aligned}$$

(3). It is shown by the proof of [10], Theorem 1, that with $\mu_{n,0}$, $\sigma_{n,0}^2$ defined in [10], Corollary 1, and $Y_n \equiv (T(X) - \mu_{n,0})/(n\sigma_{n,0})$, for $s \in (-s_0, s_0)$ for some $s_0 > 0$,

$$\lim_{n\wedge p\to\infty, n\geq p+2} M_{Y_n}(s) = M_{\mathcal{N}(0,1)}(s) = e^{s^2/2},$$

where $M_Y(s) \equiv \mathbb{E}e^{sY}$ denotes the moment generating function of a generic random variable Y. Now using that for any $s \in (0, s_0)$,

$$\mathbb{E}Y_n^4 = 4\int_0^\infty t^3\mathbb{P}(Y_n > t)\,\mathrm{d}t \leq 4\int_0^\infty t^3 e^{-st}M_{Y_n}(s)\,\mathrm{d}t = \big(6/s^3\big)M_{Y_n}(s),$$

it follows that $\sup_n \mathbb{E}Y_n^4 < \infty$, and hence convergence of moments yields that $\mathbb{E}Y_n \to 0$, $\mathbb{E}Y_n^2 \to 1$. This implies $\sigma_I^2/(n^2\sigma_{n,0}^2) = \mathrm{Var}(Y_n) = \mathbb{E}Y_n^2 - (\mathbb{E}Y_n)^2 \to 1$. Hence the asymptotic formula for σ_I^2 holds. In particular, this means that there exists some sufficiently large M such that for $n \wedge p \geq M$, $n \geq p + 2$,

$$\sigma_I^2 \geq M^{-1}\cdot n^2\sigma_{n,0}^2 = \frac{1}{2M}\cdot n^2\bigg[-\frac{p}{N} - \log\bigg(1 - \frac{p}{N}\bigg)\bigg] \overset{(*)}{\geq} \frac{1}{4M}n^2\cdot\frac{p^2}{N^2} \geq \frac{p^2}{4M},$$

where in $(*)$ we used the inequality $-x - \log(1-x) \geq x^2/2$ that holds for $x \in (0, 1)$.

(4). Recall that $\{\lambda_j\}_{j=1}^p$ are eigenvalues of Σ. By (1)–(3), we only need to show that for some universal $C > 0$,

$$\frac{\sqrt{N\sum_j(\lambda_j - 1)^2}}{(N\sum_j(|\lambda_j - 1| \wedge (\lambda_j - 1)^2)) \vee \sigma_I} \leq \frac{C}{(\sigma_I \wedge N)^{1/2}}. \tag{5.12}$$

To see this, let $v_j \equiv |\lambda_j - 1|$, and $J \equiv \{j \in [p] : v_j \le 1\}$, it suffices to prove

$$\frac{\sqrt{N\sum_{j\in J} v_j^2} \vee \sqrt{N\sum_{j\in J^c} v_j^2}}{(N\sum_{j\in J} v_j^2) \vee (N\sum_{j\in J^c} v_j) \vee \sigma_I} \le \frac{C}{(\sigma_I \wedge N)^{1/2}}. \tag{5.13}$$

This follows as

$$\begin{aligned}\text{LHS of (5.13)} &\le \frac{\sqrt{N\sum_{j\in J} v_j^2}}{(N\sum_{j\in J} v_j^2) \vee \sigma_I} + \frac{\sqrt{N}\sum_{j\in J^c} v_j}{(N\sum_{j\in J^c} v_j) \vee \sigma_I} \\ &\le \frac{1}{\inf_{x\ge 0}(x \vee \frac{\sigma_I}{x})} + N^{-1/2} \lesssim (\sigma_I \wedge N)^{-1/2}.\end{aligned}$$

The proof is complete. □

5.2. *Proofs for Section 3.1.2 (Ledoit–Nagao–Wolf's test).*

5.2.1. *Evaluation of derivatives.*

LEMMA 5.2. *Recall the form of $T_{\mathrm{LNW}}(X)$ in (3.7). We assume without loss of generality that $\mu = 0$. Then for any $(i, j), (i', j') \in [N] \times [p]$:*

1. $(\nabla T_{\mathrm{LNW}}(X))_{(ij)} = (X(S-I) - (\mathrm{tr}(S)/N)X)_{ij} = e_j^\top (S-I)X_i - (\mathrm{tr}(S)/N)X_{ij}$.
2. $(\nabla^2 T_{\mathrm{LNW}}(X))_{(ij)(i'j')} = N^{-1}\delta_{jj'}X_i^\top X_{i'} + N^{-1}X_{i'j}X_{ij'} + \delta_{ii'}(S-I)_{jj'} - (2/N^2)X_{ij}X_{i'j'} - (\mathrm{tr}(S)/N)\delta_{ii'}\delta_{jj'}$.

Furthermore, for any $(i_\ell, j_\ell) \in [N] \times [p]$, $\ell = 1, 2, 3, 4$,

$$\begin{aligned}&\partial_{(i_1j_1)(i_2j_2)(i_3j_3)(i_4j_4)} T_{\mathrm{LNW}}(X) \\ &\quad= N^{-1}(\delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_2}\delta_{j_3j_4} + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_2}\delta_{j_3j_4} \\ &\qquad+ \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_4}\delta_{j_2j_3} \\ &\qquad+ \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_4}\delta_{j_2j_3}) \\ &\qquad- 2N^{-2}(\delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_4}\delta_{j_2j_3} + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_2}\delta_{j_3j_4}).\end{aligned}$$

PROOF. We shorthand T_{LNW} as T. As $\partial_{ij}S(X) = N^{-1}(e_jX_i^\top + X_ie_j^\top)$, for the first-order derivatives we have

$$\begin{aligned}\partial_{(ij)}T(X) &= \frac{N}{4}\Big(\mathrm{tr}\big[\partial_{(ij)}(S-I)^2\big] - \frac{1}{N}\cdot 2\,\mathrm{tr}(S)\,\mathrm{tr}[\partial_{(ij)}S]\Big) \\ &= \frac{1}{2}\,\mathrm{tr}\big[(S-I)\big(e_jX_i^\top + X_ie_j^\top\big)\big] - \frac{\mathrm{tr}(S)X_{ij}}{N} \\ &= \big(X(S-I)\big)_{ij} - \frac{\mathrm{tr}(S)}{N}X_{ij} = e_j^\top(S-I)X_i - \frac{\mathrm{tr}(S)}{N}X_{ij}.\end{aligned}$$

For the second-order derivatives we have

$$\begin{aligned}&\partial_{(ij),(i'j')}T(X) \\ &\quad= \partial_{(i'j')}\big(e_j^\top(S-I)X_i\big) - N^{-1}\partial_{(i'j')}\big(\mathrm{tr}(S)X_{ij}\big) \\ &\quad= N^{-1}e_j^\top\big(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top\big)\cdot X_i + \delta_{ii'}e_j^\top(S-I)e_{j'} - N^{-1}\big((2/N)X_{ij}X_{i'j'} + \delta_{ii'}\delta_{jj'}\,\mathrm{tr}(S)\big) \\ &\quad= N^{-1}\delta_{jj'}X_i^\top X_{i'} + N^{-1}X_{i'j}X_{ij'} + \delta_{ii'}(S-I)_{jj'} - 2N^{-2}X_{ij}X_{i'j'} - N^{-1}\,\mathrm{tr}(S)\delta_{ii'}\delta_{jj'}.\end{aligned}$$

For the third-order derivatives we have

$$\begin{aligned}
&\partial_{(i_1j_1)(i_2j_2)(i_3j_3)}T(X)\\
&\quad= N^{-1}\delta_{j_1j_2}\partial_{(i_3j_3)}\big(X_{i_1}^\top X_{i_2}\big) + N^{-1}\partial_{(i_3j_3)}(X_{i_2j_1}X_{i_1j_2})\\
&\qquad+ N^{-1}\delta_{i_1i_2}e_{j_1}^\top\big(e_{j_3}X_{i_3}^\top + X_{i_3}e_{j_3}^\top\big)e_{j_2} - 2N^{-2}\partial_{(i_3j_3)}(X_{i_1j_1}X_{i_2j_2}) - 2N^{-2}\delta_{i_1i_2}\delta_{j_1j_2}X_{i_3j_3}\\
&\quad= N^{-1}(\delta_{i_1i_3}\delta_{j_1j_2}X_{i_2j_3} + \delta_{i_2i_3}\delta_{j_1j_2}X_{i_1j_3})\\
&\qquad+ N^{-1}(\delta_{i_2i_3}\delta_{j_1j_3}X_{i_1j_2} + \delta_{i_1i_3}\delta_{j_2j_3}X_{i_2j_1}) + N^{-1}(\delta_{i_1i_2}\delta_{j_1j_3}X_{i_3j_2} + \delta_{i_1i_2}\delta_{j_2j_3}X_{i_3j_1})\\
&\qquad- 2N^{-2}(\delta_{i_1i_3}\delta_{j_1j_3}X_{i_2j_2} + \delta_{i_2i_3}\delta_{j_2j_3}X_{i_1j_1} + \delta_{i_1i_2}\delta_{j_1j_2}X_{i_3j_3}).
\end{aligned}$$

For the fourth-order derivatives we have

$$\begin{aligned}
&\partial_{(i_1j_1)(i_2j_2)(i_3j_3)(i_4j_4)}T(X)\\
&\quad= N^{-1}(\delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_2}\delta_{j_3j_4} + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_2}\delta_{j_3j_4}\\
&\qquad+ \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_4}\delta_{j_2j_3}\\
&\qquad+ \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_4}\delta_{j_2j_3})\\
&\qquad- 2N^{-2}(\delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_4}\delta_{j_2j_3} + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_2}\delta_{j_3j_4}).
\end{aligned}$$

The proof is complete. □

5.2.2. *Normal approximation.*

PROOF OF PROPOSITION 3.4. Let $y \equiv p/N$. We start by showing that

$$\mathbb{E}\big\|\nabla^2 T(X)\big\|_{\mathrm{op}}^4 \le C(1\vee y)^4 \tag{5.14}$$

for some absolute constant $C > 0$. Reorganizing the terms in Lemma 5.2, we have

$$\begin{aligned}
\big(\nabla^2 T(X)\big)_{(ij),(i'j')} &= N^{-1}X_i^\top X_{i'}\delta_{jj'} + N^{-1}X_{ij'}X_{i'j} - 2N^{-2}X_{i'j'}X_{ij}\\
&\quad+ \delta_{ii'}e_{j'}^\top(S - I - N^{-1}\operatorname{tr}(S)I)e_j\\
&\equiv (T_{2,1} + T_{2,2} - T_{2,3} + T_{2,4})_{(ij),(i'j')}.
\end{aligned}$$

Recall the definition of $U_{\ell,m;+}$ from Proposition 4.2. As $T_{2,1} = U_{0,0;+}$ and

$$\begin{aligned}
\big(T_{2,2}^2\big)_{(ij),(i'j')} &= N^{-2}\sum_{(\bar i\,\bar j)} X_{\bar i j}X_{i\bar j}X_{i'\bar j}X_{\bar i j'} = N^{-2}\Big(\sum_{\bar i} X_{\bar i j}X_{\bar i j'}\Big)\Big(\sum_{\bar j} X_{i\bar j}X_{i'\bar j}\Big)\\
&= N^{-1}S_{jj'}X_i^\top X_{i'} = (U_{0,1;+})_{(ij),(i'j')},
\end{aligned}$$

Proposition 4.2 entails that $\mathbb{E}\|T_{2,1}\|_{\mathrm{op}}^4 \vee \mathbb{E}\|T_{2,2}\|_{\mathrm{op}}^4 = \mathcal{O}((1\vee y)^4)$. For $T_{2,3}$, as

$$\|T_{2,3}\|_{\mathrm{op}} = \big(2/N^2\big)\sup_{u,v\in B_{N\times p}}\Big|\sum_{(ij),(i'j')} u_{ij}X_{ij}X_{i'j'}v_{i'j'}\Big| = \big(2/N^2\big)\|X\|_F^2.$$

Hence $\mathbb{E}\|T_{2,3}\|_{\mathrm{op}}^4 = \mathcal{O}(y^4) = \mathcal{O}((1\vee y)^4)$. For $T_{2,4}$, it holds by the block diagonal structure that

$$\|T_{2,4}\|_{\mathrm{op}} = \big\|S - I - N^{-1}\operatorname{tr}(S)I\big\|_{\mathrm{op}} \le \|S - I\|_{\mathrm{op}} + N^{-1}\operatorname{tr}(S).$$

Hence it holds by Lemma 4.4 that

$$\mathbb{E}\|T_{2,4}\|_{\mathrm{op}}^4 \lesssim (y \vee \sqrt{y})^4 + N^{-4} \cdot N^{-4}\mathbb{E}\|X\|_F^8 \lesssim (1 \vee y)^4.$$

By collecting the estimates of $T_{2,1}$-$T_{2,4}$, we complete the proof of (5.14).

Next we show that $\mathbb{E}\|\nabla T(X)\|_F^4 \lesssim p^4$. This will be done by two estimates below.

(*Estimate 1*) By Lemma 5.2-(1),

$$\begin{aligned}\|\nabla T(X)\|_F^2 &\lesssim \sum_i \|(S-I)X_i\|^2 + N^{-2}\operatorname{tr}^2(S)\|X\|_F^2 \\ &\le \big(\|S-I\|_{\mathrm{op}}^2 + N^{-2}\operatorname{tr}^2(S)\big)\sum_i \|X_i\|^2,\end{aligned}$$

so by Lemma 4.4 and Proposition 4.2,

$$\begin{aligned}\mathbb{E}\|\nabla T(X)\|_F^4 &\lesssim \mathbb{E}\bigg[\big(\|S-I\|_{\mathrm{op}}^2 + N^{-2}\operatorname{tr}^2(S)\big)\sum_i \|X_i\|^2\bigg]^2 \\ &\lesssim \sum_{i,i'} \mathbb{E}\big[\big(\|S-I\|_{\mathrm{op}}^4 + N^{-4}\operatorname{tr}^4(S)\big)\|X_i\|^2\|X_{i'}\|^2\big] \\ &\le \sum_{i,i'} \big(\mathbb{E}^{1/2}\|S-I\|_{\mathrm{op}}^8 + N^{-4}\mathbb{E}^{1/2}\operatorname{tr}^8(S)\big)\cdot \mathbb{E}^{1/4}\|X_i\|^8 \cdot \mathbb{E}^{1/4}\|X_{i'}\|^8 \\ &\lesssim N^2 \cdot \bigg[\Big(\frac{p}{N}\Big)^2 + \Big(\frac{p}{N}\Big)^4 \cdot \mathbb{E}^{1/2}\|S\|_{\mathrm{op}}^8\bigg] \cdot p \cdot p \lesssim p^4(1+y^6).\end{aligned}$$

(*Estimate 2*) Note that

$$\nabla T(X) = X\big(S - N^{-1}\operatorname{tr}(S)I\big) - X \equiv T_{1,1} + T_{1,2}.$$

It is clear that $\mathbb{E}\|T_{1,2}\|_F^2 \lesssim Np$. To handle $T_{1,1}$, note that

$$\begin{aligned}\|T_{1,1}\|_F^2 &= N\operatorname{tr}\big(\big(S - N^{-1}\operatorname{tr}(S)I\big)^2 S\big) \\ &= N\operatorname{tr}\big(S^3 + N^{-2}\operatorname{tr}^2(S)S - 2N^{-1}\operatorname{tr}(S)S^2\big) \\ &= N\big[\operatorname{tr}(S^3) + N^{-2}\operatorname{tr}^3(S) - 2N^{-1}\operatorname{tr}(S)\operatorname{tr}(S^2)\big].\end{aligned}$$

Then using Lemma C.5-(1)(2)(3), we have under the prescribed asymptotics that

$$\begin{aligned}\mathbb{E}\|T_{1,1}\|_F^2 &= N\big[py^2 + 3py + p + 3y^2 + 3y + 4N^{-1}y + N^{-2}\big(p^3 + 6py + 8N^{-1}y\big) \\ &\quad - 2N^{-1}\big(p^2y + p^2 + py + 4\big(y^2+y\big) + 4N^{-1}y\big)\big] \\ &= p^2\bigg(1 + \frac{N}{p} + \frac{1}{N} + \frac{3}{p} - \frac{4}{Np} - \frac{2}{N^2}\bigg) \\ &= p^2\big[1 + \mathcal{O}\big((N\wedge p)^{-1}\big)\big] + pN.\end{aligned}$$

Hence we have

$$\begin{aligned}\mathbb{E}\|\nabla T(X)\|_F^4 &= \big(\mathbb{E}\|\nabla T(X)\|_F^2\big)^2 + \operatorname{Var}\big(\|\nabla T(X)\|_F^2\big) \\ &= \mathcal{O}\big(p^4\big(1+y^{-2}\big)\big) + \operatorname{Var}\big(\|\nabla T(X)\|_F^2\big).\end{aligned} \tag{5.15}$$

By the Gaussian–Poincaré inequality, we have

$$\begin{aligned}\operatorname{Var}\big(\|\nabla T(X)\|_F^2\big) &\le \mathbb{E}\big\|\nabla\|\nabla T(X)\|_F^2\big\|_F^2 = 4\mathbb{E}\big\|\big(\nabla^2 T(X)\big)^\top \nabla T(X)\big\|_F^2 \\ &\le 4\mathbb{E}^{1/2}\big\|\nabla^2 T(X)\big\|_{\mathrm{op}}^4 \cdot \mathbb{E}^{1/2}\|\nabla T(X)\|_F^4.\end{aligned}$$

Combining the above display with (5.15) yields that

$$\mathbb{E}\|\nabla T(X)\|_F^4 \le \mathcal{O}\big(p^4(1+y^{-2})\big) + 4\mathbb{E}^{1/2}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \cdot \mathbb{E}^{1/2}\|\nabla T(X)\|_F^4.$$

Solving the quadratic inequality above and using (5.14), we arrive at

$$\mathbb{E}\|\nabla T(X)\|_F^4 = \mathcal{O}\big(p^4(1+y^{-2}) \vee \mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4\big) = \mathcal{O}\big(p^4(1+y^{-2})\big).$$

Combining the above two estimates, we have

$$\mathbb{E}\|\nabla T(X)\|_F^4 \lesssim p^4 \max_{y\ge 0} \min\big\{(1+y^6), (1+y^{-2})\big\} \asymp p^4.$$

The rest of the proof proceeds along the lines in the proof of Proposition 3.1, with the help of the variance formula in Proposition 3.5-(3). □

5.2.3. *Contiguity.*

PROOF OF PROPOSITION 3.5. (1). Recall that $Z_1, \dots, Z_n$ are i.i.d. samples from $\mathcal{N}(0, I_p)$. By Lemma 5.2, with $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$,

$$\begin{aligned}\mathscr{T}_{\Sigma;\mathrm{LNW}}(Z) &= \big[Z\Sigma^{1/2}\big(\Sigma^{1/2} S_Z \Sigma^{1/2} - I\big) - N^{-1}\operatorname{tr}(\Sigma S_Z) Z \Sigma^{1/2}\big]\Sigma^{1/2} \\ &= Z\Sigma S_Z \Sigma - Z\Sigma - N^{-1}\operatorname{tr}(\Sigma S_Z) Z\Sigma,\end{aligned}$$

so

$$\begin{aligned}&\mathscr{T}_{\Sigma;\mathrm{LNW}}(Z) - \mathscr{T}_{I;\mathrm{LNW}}(Z) \\ &= \big[Z\Sigma(S_Z\Sigma - I) - Z(S_Z - I)\big] - \frac{1}{N}\big[\operatorname{tr}(\Sigma S_Z) Z\Sigma - \operatorname{tr}(S_Z) Z\big] \\ &= \big[Z\Sigma(S_Z\Sigma - I) - Z(S_Z\Sigma - I) + Z(S_Z\Sigma - I) - Z(S_Z - I)\big] \\ &\quad - \frac{1}{N}\big[\operatorname{tr}(\Sigma S_Z) Z\Sigma - \operatorname{tr}(\Sigma S_Z) Z + \operatorname{tr}(\Sigma S_Z) Z - \operatorname{tr}(S_Z) Z\big] \\ &= Z(\Sigma - I)(S_Z\Sigma - I) + Z S_Z(\Sigma - I) - \frac{1}{N}\operatorname{tr}(\Sigma S_Z) Z(\Sigma - I) - \frac{1}{N}\operatorname{tr}\big((\Sigma - I) S_Z\big) Z \\ &\equiv V_1(Z) + V_2(Z) + V_3(Z) + V_4(Z).\end{aligned}$$

Note that

$$\begin{aligned}\|V_1(Z)\|_F^2 &\le \|S_Z\Sigma - I\|_{\mathrm{op}}^2 \|Z(\Sigma - I)\|_F^2 \le \|S_Z\Sigma - I\|_{\mathrm{op}}^2 \|Z\|_{\mathrm{op}}^2 \|\Sigma - I\|_F^2, \\ \|V_2(Z)\|_F^2 &\le \|Z S_Z\|_{\mathrm{op}}^2 \|\Sigma - I\|_F^2 \le \|Z\|_{\mathrm{op}}^2 \|S_Z\|_{\mathrm{op}}^2 \|\Sigma - I\|_F^2, \\ \|V_3(Z)\|_F^2 &\le N^{-2}\operatorname{tr}^2(\Sigma S_Z)\|Z(\Sigma - I)\|_F^2 \le p^2 N^{-2}\|\Sigma\|_{\mathrm{op}}^2 \|S_Z\|_{\mathrm{op}}^2 \|Z(\Sigma - I)\|_F^2, \\ \|V_4(Z)\|_F^2 &\le N^{-2}\operatorname{tr}^2\big((\Sigma - I) S_Z\big)\|Z\|_F^2 \\ &\le N^{-2}\|S_Z\|_F^2 \|Z\|_F^2 \|\Sigma - I\|_F^2 \le p N^{-2}\|S_Z\|_{\mathrm{op}}^2 \|Z\|_F^2 \|\Sigma - I\|_F^2.\end{aligned}$$

Under $p/N \le M$, we have

$$V_{\Sigma;\mathrm{LNW}}^2 \lesssim_M N\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.$$

(2). By Lemma C.4, with $\delta_N \equiv N^{-1} - 2N^{-2}$,

$$\begin{aligned} m_\Sigma &= \frac{N}{4}\Big[\mathbb{E}\operatorname{tr}(S-I)^2 - \frac{1}{N}\mathbb{E}\operatorname{tr}^2(S)\Big] = \frac{N}{4}\Big[\mathbb{E}\operatorname{tr}(S^2) - 2\mathbb{E}\operatorname{tr}(S) + p - \frac{1}{N}\mathbb{E}\operatorname{tr}^2(S)\Big] \\ &= \frac{N}{4}\big[\big(1+N^{-1}\big)\operatorname{tr}\big(\Sigma^2\big) + N^{-1}\operatorname{tr}^2(\Sigma) - 2\operatorname{tr}(\Sigma) + p - N^{-1}\operatorname{tr}^2(\Sigma) - 2N^{-2}\operatorname{tr}\big(\Sigma^2\big)\big] \\ &= \frac{N}{4}\big[(1+\delta_N)\operatorname{tr}\big(\Sigma^2\big) - 2\operatorname{tr}(\Sigma) + p\big]. \end{aligned}$$

Hence

$$\begin{aligned} m_\Sigma - m_I &= \frac{N}{4}\big[(1+\delta_N)\operatorname{tr}\big(\Sigma^2 - I\big) - 2\operatorname{tr}(\Sigma - I)\big] \\ &= \frac{N}{4}\big[\|\Sigma - I\|_F^2 + \delta_N \operatorname{tr}\big(\Sigma^2 - I\big)\big]. \end{aligned}$$

(3). By Plancherel's theorem (i.e., [9], formula (6.2)), we have

$$\begin{aligned} \sigma_{I;\mathrm{LNW}}^2 &= \sum_{(ij)}\big[\mathbb{E}\partial_{(ij)}T(X)\big]^2 + \frac{1}{2!}\sum_{(i_1j_1)(i_2j_2)}\big[\mathbb{E}\partial_{(i_1j_1)(i_2j_2)}T(X)\big]^2 \\ &\quad + \frac{1}{3!}\sum_{(i_1j_1)(i_2j_2)(i_3j_3)}\big[\mathbb{E}\partial_{(i_1j_1)(i_2j_2)(i_3j_3)}T(X)\big]^2 \\ &\quad + \frac{1}{4!}\sum_{(i_1j_1)(i_2j_2)(i_3j_3)(i_4j_4)}\big[\mathbb{E}\partial_{(i_1j_1)(i_2j_2)(i_3j_3)(i_4j_4)}T(X)\big]^2 \\ &\equiv \mathrm{(I)} + \mathrm{(II)} + \mathrm{(III)} + \mathrm{(IV)}. \end{aligned}$$

Terms (I)–(IV) are handled as follows:

- To handle (I), note that

$$\mathbb{E}\partial_{(ij)}T(X) = \mathbb{E}e_j^\top(S-I)X_i - \mathbb{E}\big[\big(\operatorname{tr}(S)/N\big)X_{ij}\big].$$

The first term satisfies

$$\begin{aligned} \mathbb{E}e_j^\top(S-I)X_i &= \mathbb{E}e_j^\top\left(\frac{1}{N}\sum_{k=1}^N X_kX_k^\top\right)X_i = N^{-1}e_j^\top\mathbb{E}\big(X_i\cdot\|X_i\|^2\big) \\ &= N^{-1}e_j^\top\mathbb{E}\left(\frac{X_i}{\|X_i\|}\cdot\|X_i\|^3\right) = N^{-1}e_j^\top\mathbb{E}\left(\frac{X_i}{\|X_i\|}\right)\cdot\mathbb{E}\|X_i\|^3 = 0. \end{aligned}$$

A similar identity holds for the second term, so (I) $= 0$.

- (II) $\lesssim p/N = \mathfrak{o}(p^2)$ by noting that $\mathbb{E}\partial_{(i_1j_1)(i_2j_2)}T(X) = (N^{-1} - 2N^{-2})\cdot\delta_{i_1i_2}\delta_{j_1j_2}$.
- (III) $= 0$ by direct calculation.
- (IV) $= 6p^2(1+\mathfrak{o}(1))$ by direct calculation.

The proof is now complete by collecting all of the estimates.

(4). By (1)–(3), $\|\Sigma\|_{\mathrm{op}} \le \|\Sigma - I\|_F + 1$ and the condition $p/N \le M$, we only need to show that

$$\frac{\sqrt{N}\|\Sigma - I\|_F^2 \vee \sqrt{N\|\Sigma - I\|_F^2}}{(N\|\Sigma - I\|_F^2 - N\delta_N|\operatorname{tr}(\Sigma^2 - I)|)_+ \vee \sigma_{I;\mathrm{LNW}}} \le \frac{C_M}{(\sigma_{I;\mathrm{LNW}} \wedge N)^{1/2}}. \tag{5.16}$$

Note that with $\{\lambda_j\}_{j=1}^p$ denoting the eigenvalues of Σ,

$$|\mathrm{tr}(\Sigma^2 - I)| = \left|\sum_{j=1}^p (\lambda_j^2 - 1)\right| \le \max_j(\lambda_j + 1) \cdot \sum_{j=1}^p |\lambda_j - 1| \tag{5.17}$$
$$\le \sqrt{p}(\|\Sigma\|_{\mathrm{op}} + 1)\|\Sigma - I\|_F \lesssim_M \sqrt{N}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F,$$

so for N large enough, (5.16) is satisfied provided that

$$\frac{\sqrt{N}\|\Sigma - I\|_F^2 \vee \sqrt{N\|\Sigma - I\|_F^2}}{(N\|\Sigma - I\|_F^2 - C'_M\sqrt{N}\|\Sigma - I\|_F)_+ \vee \sigma_{I;\mathrm{LNW}}} \le \frac{C_M}{(\sigma_{I;\mathrm{LNW}} \wedge N)^{1/2}}. \tag{5.18}$$

To see this, note that the left-hand side of the above display is bounded, up to a constant that may depend on M, by

$$\mathbf{1}_{\sqrt{N}\|\Sigma - I\|_F \le 2C'_M} \frac{1}{\sigma_{I;\mathrm{LNW}}} + \mathbf{1}_{\sqrt{N}\|\Sigma - I\|_F > 2C'_M} \frac{\sqrt{N}\|\Sigma - I\|_F^2 \vee \sqrt{N\|\Sigma - I\|_F^2}}{N\|\Sigma - I\|_F^2 \vee \sigma_{I;\mathrm{LNW}}}$$
$$\lesssim \frac{1}{\sigma_{I;\mathrm{LNW}}} + \frac{\sqrt{N}\|\Sigma - I\|_F^2}{N\|\Sigma - I\|_F^2 \vee \sigma_{I;\mathrm{LNW}}} + \frac{\sqrt{N\|\Sigma - I\|_F^2}}{N\|\Sigma - I\|_F^2 \vee \sigma_{I;\mathrm{LNW}}}$$
$$\le \frac{1}{\sigma_{I;\mathrm{LNW}}} + \frac{1}{N^{1/2}} + \frac{1}{\inf_{x\ge 0}(x \vee \frac{\sigma_{I;\mathrm{LNW}}}{x})} \le \text{RHS of (5.18)}.$$

This completes the proof. □

5.2.4. *Completing the proof for power expansion.*

PROOF OF THEOREM 3.6. Abbreviate Ψ_{LNW} by Ψ. By Proposition 3.4 and Proposition 3.5, we have

$$\left|\mathbb{E}_\Sigma \Psi(X) - \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot (\|\Sigma - I\|_F^2 + Q_{\mathrm{LNW}}(\Sigma))}{4\sigma_{I;\mathrm{LNW}}}, 1\right) > z_\alpha\right)\right| \le C \cdot p^{-1/3}.$$

We only need to remove the residual term $Q_{\mathrm{LNW}}(\Sigma)$. To see this, note that by (5.17),

$$|Q_{\mathrm{LNW}}(\Sigma)| \le C_M N^{-1/2}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F.$$

So using Lemma 2.5 we have

$$\Delta P \le \frac{C_{\alpha,M}(\|\Sigma - I\|_F \vee 1)}{N^{1/2}\|\Sigma - I\|_F},$$

where

$$\Delta P \equiv \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot (\|\Sigma - I\|_F^2 + Q_{\mathrm{LNW}}(\Sigma))}{4\sigma_{I;\mathrm{LNW}}}, 1\right) > z_\alpha\right) - \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot \|\Sigma - I\|_F^2}{4\sigma_{I;\mathrm{LNW}}}, 1\right) > z_\alpha\right).$$

On the other hand, by anticoncentration of normal random variable,

$$\Delta P \le C_M \frac{N^{1/2}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F}{\sigma_{I;\mathrm{LNW}}}.$$

Hence

$$\begin{aligned}\Delta P &\lesssim_{\alpha,M} \frac{(\|\Sigma - I\|_F \vee 1)}{N^{1/2}\|\Sigma - I\|_F} \wedge \frac{N^{1/2}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F}{\sigma_{I;\mathrm{LNW}}} \\ &\le \mathbf{1}_{\|\Sigma - I\|_F > 1} \frac{1}{N^{1/2}} + \mathbf{1}_{\|\Sigma - I\|_F \le 1}\left[\frac{1}{N^{1/2}\|\Sigma - I\|_F} \wedge \frac{N^{1/2}\|\Sigma - I\|_F}{\sigma_{I;\mathrm{LNW}}}\right] \\ &\le \frac{1}{N^{1/2}} + \frac{1}{\inf_{\alpha \ge 0}(x \vee \frac{\sigma_{I;\mathrm{LNW}}}{x})} \asymp \frac{1}{(\sigma_{I;\mathrm{LNW}} \wedge N)^{1/2}}.\end{aligned}$$

Similarly we may get a lower bound for ΔP. The proof is complete. □

6. Proofs for Section 3.2 (testing sphericity).

6.1. *Proofs for Section 3.2.1 (LRT).*

6.1.1. *Evaluation of derivatives.*

LEMMA 6.1. *Recall the form of $T_{\mathrm{LRT},s}(X)$ in (3.9) and the definition of $b(S)$ in (3.10). Then for any $(i, j), (i', j') \in [N] \times [p]$:*

1. $\partial_{(ij)} T_{\mathrm{LRT},s}(X) = (X(I - S^{-1}))_{(ij)} + (1/b(S) - 1)X_{ij} = e_j^\top[(I - S^{-1})X_i + (1/b(S) - 1)X_i]$.
2. $\partial_{(ij),(i'j')} T_{\mathrm{LRT},s}(X) = N^{-1}X_i^\top S^{-1}(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top)S^{-1}e_j + \delta_{ii'}e_j^\top(I - S^{-1})e_{j'} + (1/b(S) - 1)\delta_{ii'}\delta_{jj'} - (2/Np)X_{ij}X_{i'j'}/b^2(S)$.

PROOF. (1). We shorthand $T_{\mathrm{LRT},s}(X)$ as T. By definition, (5.1) and (5.3), we have

$$\begin{aligned}\partial_{(ij)} T(X) &= \frac{N}{2}\big(p \cdot \partial_{(ij)} \log \mathrm{tr}(S) - \partial_{ij} \log\det S\big) \\ &= \frac{N}{2}\left(p\frac{\partial_{(ij)}\,\mathrm{tr}(S)}{\mathrm{tr}(S)} - \sum_{k,\ell=1}^{p} \frac{\partial \log\det S}{\partial S_{k\ell}}\frac{\partial S_{k\ell}}{\partial X_{ij}}\right) \\ &= \frac{N}{2}\left[\frac{2p}{N}\frac{X_{ij}}{\mathrm{tr}(S)} - \sum_{k,\ell=1}^{p}(S^{-1})_{k\ell} \cdot \frac{1}{N}(\delta_{kj}X_{i\ell} + \delta_{\ell j}X_{ik})\right] \\ &= \frac{p}{\mathrm{tr}(S)}X_{ij} - \sum_{k=1}^{p}(S^{-1})_{kj}X_{ik} \\ &= \big(X(I - S^{-1})\big)_{ij} + \left(\frac{p}{\mathrm{tr}(S)} - 1\right)X_{ij}.\end{aligned}$$

(2). By the previous part, we have

$$\partial_{(ij),(i'j')} T(X) = \partial_{(i'j')}\big(X(I - S^{-1})\big)_{ij} + \partial_{(i'j')}\left(\frac{p}{\mathrm{tr}(S)} - 1\right)X_{ij} \equiv \mathrm{(I)} + \mathrm{(II)}.$$

The first term above is already calculated in Lemma 5.1-(2):

$$\mathrm{(I)} = N^{-1}X_i^\top S^{-1}\big(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top\big)S^{-1}e_j + \delta_{ii'}e_j^\top(I - S^{-1})e_{j'}.$$

So we only need to evaluate the second term:

$$\begin{aligned}\text{(II)} &= p \cdot \partial_{(i'j')} \operatorname{tr}^{-1}(S) \cdot X_{ij} + \left(\frac{p}{\operatorname{tr}(S)} - 1\right)\partial_{(i'j')} X_{ij} \\ &= -p \cdot \partial_{(i'j')} \operatorname{tr}(S) \cdot X_{ij} \cdot \operatorname{tr}^{-2}(S) + \left(\frac{p}{\operatorname{tr}(S)} - 1\right)\delta_{ii'}\delta_{jj'} \\ &= -\frac{2p}{N} X_{ij} X_{i'j'} \cdot \operatorname{tr}^{-2}(S) + \left(\frac{p}{\operatorname{tr}(S)} - 1\right)\delta_{ii'}\delta_{jj'}.\end{aligned}$$

The proof is complete. □

6.1.2. *Normal approximation.*

PROOF OF THEOREM 3.7. We abbreviate $T_{\mathrm{LRT},s}(X)$ as T. First we bound the norm for the gradient. Comparing Lemmas 5.1-(1) and 6.1-(1), we only need to control

$$\begin{aligned}\mathbb{E}\big\|\big(b^{-1}(S) - 1\big)X\big\|_F^4 &= \mathbb{E}\big(N\big(b^{-1}(S) - 1\big)^2 \operatorname{tr}(S)\big)^2 \\ &\le N^2 p^2 \cdot \mathbb{E}^{1/2} b^4(S) \cdot \mathbb{E}^{1/2}\big(b^{-1}(S) - 1\big)^8 \lesssim N^2 p^2 \cdot \left(\frac{p}{N}\right)^2 = p^4.\end{aligned}$$

The inequality in the final line of the above display follows as

$$\mathbb{E} b^4(S) \le \mathbb{E}\|S\|_{\mathrm{op}}^4 \lesssim 1, \tag{6.1}$$

$$\mathbb{E}\big(b^{-1}(S) - 1\big)^8 = \mathbb{E}^{1/2} b^{-16}(S) \cdot \mathbb{E}^{1/2}\big(b(S) - 1\big)^{16} \overset{(*)}{\lesssim} \big(pN^{-1}\big)^4. \tag{6.2}$$

Here $(*)$ follows from Lemma C.4-(3). Now by combining with (5.7) derived in the proof of Proposition 3.1, we see that $\mathbb{E}\|\nabla T(X)\|_F^4 \lesssim p^4$.

Next we bound the spectral norm of the Hessian. Comparing Lemmas 5.1-(1) and 6.1-(1), we only need to control the spectral norms of T_4 and T_5, where

$$(T_4)_{(ij),(i'j')} \equiv \big(b^{-1}(S) - 1\big)\delta_{ii'}\delta_{jj'}, (T_5)_{(ij),(i'j')} \equiv -\frac{2}{Np} X_{ij} X_{i'j'} \cdot b^{-2}(S).$$

For T_4, clearly $\|T_4\|_{\mathrm{op}} = |1/b(S) - 1|$, so $\mathbb{E}\|T_4\|_{\mathrm{op}}^4 = \mathbb{E}(1/b(S) - 1)^4 \lesssim (p/N)^2$ by (6.1). For T_5, note that

$$\|T_5\|_{\mathrm{op}} = \frac{2}{Np \cdot b^2(S)} \sup_{u,v \in B_{N\times p}(1)} \left| \sum_{(ij),(i'j')} u_{ij} X_{ij} X_{i'j'} v_{i'j'} \right| = \frac{2}{Np \cdot b^2(S)} \|X\|_F^2 = \frac{2}{b(S)}.$$

So $\mathbb{E}\|T_5\|_{\mathrm{op}}^4 \lesssim \mathbb{E} b^{-4}(S) = \mathcal{O}(1)$ by Lemma C.4-(3). By combining with (5.11) derived in the proof of Proposition 3.1, we see that $\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 = \mathcal{O}(1)$. The rest of the proof proceeds along the lines in the proof of Proposition 3.1, with the help of the variance formula in Proposition 3.8-(3). □

6.1.3. *Contiguity.*

PROOF OF PROPOSITION 3.8. We will abbreviate $(T_{\mathrm{LRT},s}, m_{\Sigma;\mathrm{LRT},s}, \sigma_{\Sigma;\mathrm{LRT},s}, V_{\Sigma;\mathrm{LRT},s})$ as $(T, m_{\Sigma;s}, \sigma_{\Sigma;s}, V_{\Sigma;s})$, and assume without loss of generality that $b(\Sigma) = \operatorname{tr}(\Sigma)/p = 1$ (otherwise we may replace Σ by $\Sigma \cdot b^{-1}(\Sigma)$).

(1). By Lemma 6.1, with $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$, we have

$$\begin{aligned}\mathscr{T}_{\Sigma;s} &= \Big[Z\Sigma^{1/2}(I-\Sigma^{-1/2}S_Z^{-1}\Sigma^{-1/2}) + \Big(\frac{1}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})}-1\Big)Z\Sigma^{1/2}\Big]\Sigma^{1/2}\\ &= Z(\Sigma - S_Z^{-1}) + \Big(\frac{1}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})}-1\Big)Z\Sigma = \frac{Z\Sigma}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})} - ZS_Z^{-1}.\end{aligned}$$

Hence

$$\begin{aligned}V_{\Sigma;s}^2 &= \mathbb{E}\|\mathscr{T}_{\Sigma;s}-\mathscr{T}_{I;s}\|_F^2 = \mathbb{E}\Big\|\frac{Z\Sigma}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{Z}{b(S_Z)}\Big\|_F^2\\ &\le 2\Big\{\mathbb{E}\Big[\Big(\frac{1}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{b(S_Z)}\Big)^2\|Z\Sigma\|_F^2\Big] + \mathbb{E}\big[b^{-2}(S_Z)\|Z(\Sigma-I)\|_F^2\big]\Big\}\\ &\equiv 2\mathbb{E}\big((\mathrm{I})+(\mathrm{II})\big).\end{aligned}$$

We bound (I) and (II) separately:

$$\begin{aligned}(\mathrm{I}) &= b^{-2}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-2}(S_Z)b^2((\Sigma-I)S_Z)\|Z\Sigma\|_F^2\\ &\le b^{-2}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-2}(S_Z)\|S_Z\|_{\mathrm{op}}^2\cdot(\|\Sigma\|_F^2/p)\|Z\|_{\mathrm{op}}^2\|\Sigma-I\|_F^2;\\ (\mathrm{II}) &\le b^{-2}(S_Z)\cdot\|Z\|_{\mathrm{op}}^2\|\Sigma-I\|_F^2.\end{aligned}$$

Using Lemmas C.2 and 4.4, we have

$$V_{\Sigma;s}^2 \lesssim (p^{-1}\|\Sigma-I\|_F^2+1)N\|(\Sigma-I)\|_F^2.$$

On the other hand, a trivial bound for $V_{\Sigma;s}^2$ is

$$\begin{aligned}V_{\Sigma;s}^2 &= \mathbb{E}\Big\|\frac{Z\Sigma}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{Z}{b(S_Z)}\Big\|_F^2\\ &\lesssim \mathbb{E}b^{-2}(\Sigma^{1/2}S_Z\Sigma^{1/2})\|Z\Sigma\|_F^2 + \mathbb{E}b^{-2}(S_Z)\|Z\|_F^2 \lesssim N(\|\Sigma-I\|_F^2\vee p).\end{aligned}$$

Collecting the two bounds, we have

$$V_{\Sigma;s}^2 \lesssim [(p^{-1}\|\Sigma-I\|_F^2+1)N\|(\Sigma-I)\|_F^2]\wedge N(\|\Sigma-I\|_F^2\vee p) \asymp N\|(\Sigma-I)\|_F^2.$$

(2). As

$$m_{\Sigma;s} = \frac{N}{2}[p\cdot\mathbb{E}\log\mathrm{tr}(\Sigma S_Z) - \log\det(\Sigma) - p\log p - \mathbb{E}\log\det(S_Z)],$$

by Lemma C.3 we have

$$m_{\Sigma;s} - m_{I;s} = \frac{N}{2}[-\log\det(\Sigma) + Q_s(\Sigma)],$$

where

$$\begin{aligned}|Q_s(\Sigma)| &\equiv |p(\mathbb{E}\log\mathrm{tr}(\Sigma S_Z) - \mathbb{E}\log\mathrm{tr}(S_Z))|\\ &\lesssim N^{-1}\{1+b(\Sigma^2)+e^{-cN}[1+b^{1/2}(\Sigma^2)]\} \lesssim N^{-1}[1+b(\Sigma^2)] \lesssim N^{-1}b(\Sigma^2),\end{aligned}$$

where the last inequality follows as $b(\Sigma^2) = p^{-1}\sum_{j=1}^p \lambda_j^2 \ge p^{-2}(\sum_{j=1}^p\lambda_j)^2 = 1$.

(3). Recall T_{LRT} defined in (3.4). Define

$$\Delta(X) \equiv T_{\mathrm{LRT}}(X) - T_{\mathrm{LRT},s}(X).$$

Then for any $\varepsilon > 0$, there exists some $C_\varepsilon > 0$ such that under the null (i.e., $X_1, \dots, X_n$ are i.i.d. $\mathcal{N}(0, I_p)$),

$$\big[(1-\varepsilon)\sigma^2_{I;\mathrm{LRT}} - C_\varepsilon \operatorname{Var}_I(\Delta)\big]_+ \le \sigma^2_{I;\mathrm{LRT},s} \le (1+\varepsilon)\sigma^2_{I;\mathrm{LRT}} + C_\varepsilon \operatorname{Var}_I(\Delta). \tag{6.3}$$

We will now bound $\operatorname{Var}_I(\Delta)$. By Lemmas 5.1-(1) and 6.1-(1), we have for any $i, j \in [N] \times [p]$

$$\partial_{(ij)}\Delta(X) = \partial_{(ij)} T_{\mathrm{LRT}}(X) - \partial_{(ij)} T_{\mathrm{LRT},s}(X) = X_{ij}\big[b^{-1}(S) - 1\big].$$

By the Gaussian–Poincaré inequality [5], Theorem 3.20,

$$\begin{aligned}\operatorname{Var}_I \Delta(X) &\le \mathbb{E}\big[b^{-1}(S) - 1\big]^2 \|X\|_F^2 = Np\,\mathbb{E}\big[b(S) - 1\big]^2 b^{-1}(S) \\ &\le Np \cdot \mathbb{E}^{1/2}\big(b(S) - 1\big)^4 \cdot \mathbb{E}^{1/2} b^{-2}(S) \overset{(*)}{\lesssim} Np \cdot (Np)^{-1} = 1.\end{aligned}$$

Here $(*)$ follows from Lemma C.4-(3). Hence by choosing ε in (6.3) to be decaying to 0 slowly enough, $\sigma^2_{I;\mathrm{LRT}}$ and $\sigma^2_{I;\mathrm{LRT},s}$ share the same asymptotic formula in Proposition 3.2-(3).

(4). By (1)–(2), and using that $b(\Sigma^2) = \|\Sigma\|_F^2/p$, we only need to prove that for a given constant $C_0 > 0$, there exists some constant $C = C(C_0) > 0$ such that

$$\frac{\sqrt{N\|\Sigma - I\|_F^2}}{(-N\log\det(\Sigma) - C_0(1 + \frac{\|\Sigma\|_F^2}{p}) - C_0 e^{-cN}(\frac{\|\Sigma\|_F}{p^{1/2}} + 1))_+ \vee \sigma_{I;s}} \le \frac{C}{(\sigma_{I;s} \wedge N)^{1/2}}.$$

Equivalently, with $\lambda = (\lambda_1, \dots, \lambda_p) \in (0,\infty)^p$ and $\bar\lambda \equiv p^{-1}\sum_j \lambda_j = 1$, we only need to show

$$\frac{\sqrt{N\sum_j(\lambda_j - 1)^2}}{(N\sum_j - \log(1 + (\lambda_j - 1)) - C_0 - C_0(\frac{\sum_j \lambda_j^2}{p}) - C_0 e^{-cN}\frac{(\sum_j \lambda_j^2)^{1/2}}{p^{1/2}})_+ \vee \sigma_{I;s}}$$

is at most a multiple of $(\sigma_{I;s} \wedge N)^{-1/2}$. Let $J \equiv \{j : |\lambda_j - 1| \le 1\}$ and $J^c \equiv \{j : |\lambda_j - 1| > 1\}$. As $|\lambda_j - 1| \lesssim p$, so the first term in the denominator becomes

$$\begin{aligned}&N\sum_j\big[-\log\big(1 + (\lambda_j - 1)\big) + (\lambda_j - 1)\big] - C_0 - C_0\frac{\sum_j \lambda_j^2}{p} - C_0 e^{-cN}\frac{(\sum_j \lambda_j^2)^{1/2}}{p^{1/2}} \\ &\gtrsim N\sum_j (\lambda_j - 1)^2 \wedge |\lambda_j - 1| - C_1 p^{-1}\sum_j (\lambda_j - 1)^2 - C_2.\end{aligned}$$

Next, by breaking the summation in $\sum_j(\lambda_j - 1)^2$ into J and J^c, the above display equals

$$\begin{aligned}&N\sum_{j\in J}(\lambda_j - 1)^2 + N\sum_{j\in J}|\lambda_j - 1| - C_1\frac{\sum_{j\in J}(\lambda_j - 1)^2 + \sum_{j\in J^c}(\lambda_j - 1)^2}{p} - C_2 \\ &\ge \big(N - C_1 p^{-1}\big)\sum_{j\in J}(\lambda_j - 1)^2 + \big(N - \mathcal{O}(1)\big)\sum_{j\in J^c}|\lambda_j - 1| - C_2 \\ &\ge \frac{N}{2}\sum_j (\lambda_j - 1)^2 \wedge |\lambda_j - 1| - C_2\end{aligned}$$

for N and p large enough. Hence with $\nu_j \equiv |\lambda_j - 1|$, we only need to show that for given $C_0 > 0$,

$$\frac{\sqrt{N\sum_{j\in J}\nu_j^2} \vee \sqrt{N\sum_{j\in J^c}\nu_j^2}}{(N\sum_{j\in J}\nu_j^2 + N\sum_{j\in J^c}\nu_j - C_0)_+ \vee \sigma_{I;s}} \le \frac{C}{(\sigma_{I;s}\wedge N)^{1/2}}.$$

Equivalently, we only need to show

$$\frac{\sqrt{N\sum_{j\in J}\nu_j^2}}{(N\sum_{j\in J}\nu_j^2 - C_0)_+ \vee \sigma_{I;s}} \le \frac{C}{\sigma_{I;s}^{1/2}}, \tag{6.4}$$

$$\frac{\sqrt{N\sum_{j\in J^c}\nu_j^2}}{(N\sum_{j\in J^c}\nu_j - C_0)_+ \vee \sigma_{I;s}} \le \frac{C}{N^{1/2}}. \tag{6.5}$$

To see these inequalities, note that the left side of (6.4) is bounded by

$$\begin{aligned}&\mathbf{1}_{N\sum_{j\in J}\nu_j^2\le 2C_0}\frac{(2C_0)^{1/2}}{\sigma_{I;s}} + \mathbf{1}_{N\sum_{j\in J}\nu_j^2>2C_0}\frac{\sqrt{N\sum_{j\in J}\nu_j^2}}{(N/2)\sum_{j\in J}\nu_j^2 \vee \sigma_{I;s}}\\ &\lesssim \frac{1}{\sigma_{I;s}} + \frac{1}{\inf_{x\ge 0}(x \vee \frac{\sigma_{I;s}}{x})} \lesssim \sigma_{I;s}^{-1/2}.\end{aligned}$$

Also, the left side of (6.5) is bounded by

$$\begin{aligned}&\frac{\sqrt{N}\sum_{j\in J^c}\nu_j}{(N\sum_{j\in J^c}\nu_j - C_0)_+ \vee \sigma_{I;s}}\\ &\le \mathbf{1}_{N\sum_{j\in J^c}\nu_j\le 2C_0}\frac{(2C_0)^{1/2}}{\sqrt{N}\sigma_{I;s}} + \mathbf{1}_{N\sum_{j\in J^c}\nu_j>2C_0}\frac{\sqrt{N}\sum_{j\in J^c}\nu_j}{N\sum_{j\in J^c}\nu_j \vee \sigma_{I;s}}\\ &\lesssim \frac{1}{\sqrt{N}\sigma_{I;s}} + \frac{1}{\sqrt{N}} \lesssim \frac{1}{N^{1/2}},\end{aligned}$$

proving the claim. □

6.1.4. *Completing the proof for power expansion.*

PROOF OF THEOREM 3.9. The proof is similar to that of Theorem 3.6, we provide some details for the convenience of the reader. Without loss of generality we assume $b(\Sigma) = 1$. Abbreviate $\Psi_{\mathrm{LRT},s}$ by Ψ and $Q_{\mathrm{LRT},s}(\Sigma)$ by $Q(\Sigma)$. By Theorem 3.7 and Proposition 3.8, we have

$$\left|\mathbb{E}_\Sigma\Psi(X) - \mathbb{P}\left(\mathcal{N}\left(\frac{N\cdot(-\log\det(\Sigma) + Q(\Sigma))}{2\sigma_{I;s}}, 1\right) > z_\alpha\right)\right| \le C\cdot p^{-1/3}.$$

We only need to remove the residual term $Q(\Sigma)$. To this end, we claim that

$$|\Delta P| \le C_\alpha\left[\frac{NQ(\Sigma)}{\sigma_{I;s}} \wedge \frac{Q(\Sigma)}{|\log\det(\Sigma)|}\right], \tag{6.6}$$

where

$$\Delta P \equiv \mathbb{P}\left(\mathcal{N}\left(\frac{N\cdot(-\log\det(\Sigma) + Q(\Sigma))}{2\sigma_{I;s}}, 1\right) > z_\alpha\right) - \mathbb{P}\left(\mathcal{N}\left(\frac{-N\log\det(\Sigma)}{2\sigma_{I;s}}, 1\right) > z_\alpha\right).$$

Here the first bound in (6.6) is by anticoncentration of the normal distribution, and the second bound in (6.6) follows from Lemma 2.5.

Let $\{\lambda_j\}_{j=1}^p$ be the eigenvalues of Σ so that $\sum_{j=1}^p \lambda_j = p$. Then by (3.11), $Q(\Sigma) \lesssim 2(Np)^{-1}\sum_{j=1}^p \lambda_j^2$. Hence using the bound $\sigma_{I;s} \geq cp$, (6.6) entails that

$$|\Delta P| \leq C'_\alpha \cdot \Big[\frac{p^{-1}\sum_{j=1}^p \lambda_j^2}{p} \wedge \frac{(Np)^{-1}\sum_{j=1}^p \lambda_j^2}{\sum_{j=1}^p \lambda_j - \log \lambda_j - 1}\Big]. \tag{6.7}$$

If $\max_j \lambda_j \leq 10$, we use the first bound in (6.7) to conclude that $\Delta P \lesssim_\alpha p^{-1}$. Otherwise, by writing $J \equiv \{j \in [p] : |\lambda_j - 1| \geq 1\}$ and $J^c \equiv [p]\backslash J$, the second bound in (6.7) yields that

$$\begin{aligned}|\Delta P| &\lesssim_\alpha \frac{(Np)^{-1}\sum_{j=1}^p \lambda_j^2}{\sum_{j=1}^p |\lambda_j - 1| \wedge (\lambda_j - 1)^2} \lesssim \frac{(Np)^{-1}\sum_{j\in J}(\lambda_j - 1)^2 + (Np)^{-1}(|J| + |J^c|)}{\sum_{j\in J}|\lambda_j - 1|} \\ &\leq \frac{(Np)^{-1}\sum_{j\in J}(\lambda_j - 1)^2}{\sum_{j\in J}|\lambda_j - 1|} + \frac{N^{-1}}{\sum_{j\in J}|\lambda_j - 1|} \equiv \text{(I)} + \text{(II)}.\end{aligned}$$

Now (II) $\lesssim N^{-1}$ as $\max_j \lambda_j > 10$, and (I) satisfies

$$\text{(I)} \leq (Np)^{-1}\max_{j\in J}|\lambda_j - 1| \lesssim N^{-1}$$

by using the trivial bound that $\max_j \lambda_j \leq p$ due to the normalization $b(\Sigma) = 1$. The proof is complete. □

6.2. *Proofs for Section 3.2.2 (John's test).*

6.2.1. *Evaluation of derivatives.*

LEMMA 6.2. *Recall the form of $T_J(X)$ in (3.12) and the definition of $b_\ell(S)$ in (3.10). Then the following hold*:

1. *For the first-order partial derivatives*: *for any $(i, j) \in [N] \times [p]$,*

$$\partial_{(ij)} T_J(X) = \Big(\frac{XS}{b^2(S)} - X \cdot \frac{b_2(S)}{b^3(S)}\Big)_{ij} = \frac{X_i^\top S e_j}{b^2(S)} - X_{ij}\frac{b_2(S)}{b^3(S)}.$$

2. *For the second-order partial derivatives*: *for any $(i, j), (i', j') \in [N] \times [p]$,*

$$\begin{aligned}&\partial_{(ij),(i'j')} T_J(X) \\ &= b(S)^{-2}\big(N^{-1}\delta_{jj'}X_i^\top X_{i'} + N^{-1}X_{i'j}X_{ij'} + \delta_{ii'}S_{jj'}\big) - \delta_{ii'}\delta_{jj'}\frac{b_2(S)}{b^3(S)} \\ &\quad + X_{ij}X_{i'j'}\frac{6b_2(S)}{b^4(S)Np} - \frac{4}{b^3(S)Np}\big[X_i^\top S e_j \cdot X_{i'j'} + X_{i'}^\top S e_{j'} \cdot X_{ij}\big].\end{aligned}$$

PROOF. We abbreviate $T_J(X)$ by $T(X)$ and write $b = b(S)$ in the proof if no confusion could arise.

(1). Note that $\partial_{ij} S(X) = N^{-1}(e_j X_i^\top + X_i e_j^\top)$, $\partial_{(ij)}\operatorname{tr}(S) = 2N^{-1}X_{ij}$ and

$$\partial_{(ij)} b(S) = \frac{2}{Np}X_{ij},\ \partial_{(ij)} b_2(S) = \frac{2}{Np}\operatorname{tr}\big(S\big(e_j X_i^\top + X_i e_j^\top\big)\big) = \frac{4}{Np}X_i^\top S e_j. \tag{6.8}$$

For the first-order derivatives we have

$$\begin{aligned}
\partial_{(ij)}T(X) &= \frac{N}{4}\operatorname{tr}\biggl[2\Bigl(\frac{S}{b}-I\Bigr)\partial_{(ij)}\Bigl(\frac{S}{b}\Bigr)\biggr] \\
&= \frac{N}{2}\operatorname{tr}\biggl[\Bigl(\frac{S}{b}-I\Bigr)\cdot\frac{N^{-1}(e_jX_i^\top+X_ie_j^\top)b-2(Np)^{-1}SX_{ij}}{b^2}\biggr] \\
&= \frac{1}{2b^2}\operatorname{tr}\bigl[(S-bI)\bigl(e_jX_i^\top+X_ie_j^\top\bigr)\bigr]-\frac{X_{ij}}{b^3p}\operatorname{tr}\bigl[(S-bI)S\bigr] \\
&= \frac{(XS)_{ij}}{b^2}-\frac{X_{ij}}{b}-\biggl[\frac{X_{ij}b_2}{b^3}-\frac{X_{ij}}{b}\biggr] \\
&= \frac{(XS)_{ij}}{b^2}-X_{ij}\cdot\frac{b_2}{b^3}\equiv T_{1,(ij)}(X)-T_{2,(ij)}(X).
\end{aligned}$$

(2). For the second-order derivatives,

$$\begin{aligned}
\partial_{(i'j')}T_{1,(ij)}(X) &= \frac{\partial_{(i'j')}(X_i^\top Se_j)b^2-(X_i^\top Se_j)\partial_{(i'j')}b^2}{b^4} \\
&= \frac{\delta_{ii'}e_{j'}^\top Se_j+N^{-1}X_i^\top(e_{j'}X_{i'}^\top+X_{i'}e_{j'}^\top)e_j}{b^2}-\frac{4}{pN}\frac{X_i^\top Se_j\cdot X_{i'j'}}{b^3} \\
&= \frac{N^{-1}\delta_{jj'}X_i^\top X_{i'}+N^{-1}X_{i'j}X_{ij'}+\delta_{ii'}S_{jj'}}{b^2}-\frac{4}{pN}\frac{X_i^\top Se_j\cdot X_{i'j'}}{b^3}, \\
\partial_{(i'j')}T_{2,(ij)}(X) &= \delta_{ii'}\delta_{jj'}\frac{b_2}{b^3}+X_{ij}\cdot\partial_{(i'j')}\biggl[\frac{b_2}{b^3}\biggr] \\
&= \delta_{ii'}\delta_{jj'}\frac{b_2}{b^3}+X_{ij}\cdot\biggl[\frac{4X_{i'}^\top Se_{j'}}{b^3Np}-\frac{6b_2X_{i'j'}}{b^4Np}\biggr] \\
&= \delta_{ii'}\delta_{jj'}\frac{b_2}{b^3}-X_{ij}X_{i'j'}\frac{6b_2}{b^4Np}+4X_{i'}^\top Se_{j'}\cdot X_{ij}\frac{1}{b^3Np}.
\end{aligned}$$

Combining the above two displays, we have

$$\begin{aligned}
\partial_{(ij),(i'j')}T(X) &= b^{-2}\bigl(N^{-1}\delta_{jj'}X_i^\top X_{i'}+N^{-1}X_{i'j}X_{ij'}+\delta_{ii'}S_{jj'}\bigr)-\delta_{ii'}\delta_{jj'}\frac{b_2}{b^3} \\
&\quad+X_{ij}X_{i'j'}\frac{6b_2}{b^4Np}-\frac{4}{b^3Np}\bigl[X_i^\top Se_j\cdot X_{i'j'}+X_{i'}^\top Se_{j'}\cdot X_{ij}\bigr].
\end{aligned}$$

The proof is complete. □

6.2.2. *Normal approximation*.

PROOF OF PROPOSITION 3.10. We abbreviate T_J by T and write $b=b(S)$ in the proof if no confusion could arise. First we bound the operator norm of the Hessian. By Lemma 6.2-(2),

$$\begin{aligned}
\partial_{(ij),(i'j')}T(X) &= b^{-2}\bigl(N^{-1}\delta_{jj'}X_i^\top X_{i'}+N^{-1}X_{i'j}X_{ij'}+\delta_{ii'}S_{jj'}\bigr) \\
&\quad-\delta_{ii'}\delta_{jj'}\frac{b_2}{b^3}+X_{ij}X_{i'j'}\frac{6b_2}{b^4Np}-\frac{4}{b^3Np}\bigl[X_i^\top Se_j\cdot X_{i'j'}+X_{i'}^\top Se_{j'}\cdot X_{ij}\bigr] \\
&\equiv (T_1-T_2+T_3-T_4)_{(ij),(i'j')}.
\end{aligned}$$

Following the proof of Proposition 3.4 along with Lemma C.2, we have $\mathbb{E}\|T_1\|_{\mathrm{op}}^4 \lesssim (1 \vee y)^4$. Next for T_2, we have by Lemma C.2 and Lemma 4.4 that

$$\mathbb{E}\|T_2\|_{\mathrm{op}}^4 \lesssim \mathbb{E}(b_2^4 \cdot b^{-12}) \le \mathbb{E}^{1/2} b_2^8 \cdot \mathbb{E}^{1/2} b^{-24} \lesssim \mathbb{E}^{1/2}\|S\|_{\mathrm{op}}^8 \lesssim (1 \vee y)^4.$$

The operator norm of T_3 can be similarly bounded by

$$\begin{aligned}\mathbb{E}\|T_3\|_{\mathrm{op}}^4 &= \frac{6^4}{(Np)^4}\mathbb{E}\left[\left(\frac{b_2}{b^4}\right)^4 \|X\|_F^8\right] \lesssim (Np)^{-4}\mathbb{E}^{1/2} b_2^8 \cdot \mathbb{E}^{1/4} b^{-64}\mathbb{E}^{1/4}\|X\|_F^{32} \\ &\lesssim (Np)^{-4} \cdot \mathbb{E}^{1/2} b_2^8 \cdot (Np)^4 \lesssim (1 \vee y)^4.\end{aligned}$$

Last,

$$\begin{aligned}\|T_4\|_{\mathrm{op}} &\lesssim \frac{1}{b^3 Np} \cdot \sup_{u,v \in B_{N\times p}(1)} \left| \sum_{(ij),(i'j')} X_i^\top S e_j \cdot X_{i'j'} u_{ij} v_{i'j'} \right| \\ &= \frac{1}{b^3 Np} \cdot \sup_{u,v \in B_{N\times p}(1)} \left| \left(\sum_{i,j} X_i^\top S e_j u_{ij}\right)\left(\sum_{i'j'} X_{i'j'} v_{i'j'}\right) \right| \\ &\le \frac{1}{b^3 Np} \cdot \|XS\|_F \cdot \|X\|_F \le \frac{1}{b^3 Np} \cdot \|S\|_{\mathrm{op}} \|X\|_F^2.\end{aligned}$$

Hence by Lemma 4.4 and Lemma C.2, $\mathbb{E}\|T_4(X)\|_{\mathrm{op}}^4 \lesssim (1 \vee y)^4$. Putting together the bounds for T_1–T_4 yields that $\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \lesssim (1 \vee y)^4$.

Next we bound the norm of the gradient. We will show that $\mathbb{E}\|\nabla T(X)\|_F^2 \lesssim p^2$ by considering the two cases $p/N \le 1$ and $p/N > 1$ separately.

(*Case* $p/N \le 1$) By Lemma 6.2-(1), we may write

$$\nabla T(X) = b^{-1} X\big(b^{-1} S - I\big) - b^{-1} X \cdot b\big(b^{-1} S - I\big)^2,$$

so

$$\begin{aligned}\big\|\nabla T(X)\big\|_F^4 &\lesssim b^{-8}\|X\|_F^4 \|S - bI\|_{\mathrm{op}}^4 + b^{-12}\|X\|_F^4 \|S - bI\|_{\mathrm{op}}^8 \\ &\lesssim \|X\|_F^4 \big(b^{-8}\|S - I\|_{\mathrm{op}}^4 + b^{-8}|b-1|^4 + b^{-12}\|S - I\|_{\mathrm{op}}^8 + b^{-12}|b-1|^8\big).\end{aligned}$$

By Lemma 4.4 and Lemma C.4, it holds under the condition $p/N \le 1$ that

$$\mathbb{E}\big\|T(X)\big\|_F^4 \lesssim (Np)^2\big(\big(N^{-1}p\big)^2 + \big(N^{-1}p\big)^4\big) \lesssim p^4.$$

(*Case* $p/N > 1$) By Lemma 6.2-(1), we have

$$\begin{aligned}\mathbb{E}\big\|\nabla T(X)\big\|_F^2 &= \mathbb{E}\left[\left\|\frac{XS}{b^2}\right\|_F^2 + \left\|\frac{Xb_2}{b^3}\right\|_F^2 - 2\left\langle \frac{XS}{b^2}, \frac{Xb_2}{b^3}\right\rangle\right] = Np \cdot \mathbb{E}\left[\frac{bb_3 - b_2^2}{b^5}\right] \\ &= Np \cdot \mathbb{E}\big(bb_3 - b_2^2\big) + Np \cdot \mathbb{E}\big(bb_3 - b_2^2\big)\big(b^{-5} - 1\big) \equiv \mathrm{(I)} + \mathrm{(II)}.\end{aligned}$$

To handle (I), it holds by Lemma C.5-(4)(5) that under $p > N$,

$$\mathrm{(I)} = \frac{N}{p}\mathbb{E}\big[\mathrm{tr}(S)\,\mathrm{tr}\big(S^3\big) - \mathrm{tr}^2\big(S^2\big)\big] = \frac{N}{p} N^{-4}\mathcal{O}\big(N^3 p^3\big) = \mathcal{O}\big(p^2\big).$$

To handle (II), it holds by Lemmas C.2, C.4-(3), and C.5-(6) that under $p > N$,

$$\begin{aligned}\mathrm{(II)} &= Np \cdot \mathbb{E}\big(bb_3 - b_2^2\big)b^{-5}\big(1 - b^5\big) \le Np \cdot \mathbb{E}^{1/2}\big(bb_3 - b_2^2\big)^2 \mathbb{E}^{1/4} b^{-20}\mathbb{E}^{1/4}\big(b^5 - 1\big)^4 \\ &\le Np \cdot \big(\big|\mathbb{E}\big(bb_3 - b_2^2\big)\big| + \mathrm{Var}^{1/2}\big(bb_3 - b_2^2\big)\big) \cdot \mathbb{E}^{1/4} b^{-20}\mathbb{E}^{1/4}\big(b^5 - 1\big)^4 \\ &= Np \cdot \mathcal{O}\big(N^{-1}p\big) \cdot \mathcal{O}(1) \cdot \mathcal{O}\big((Np)^{-1/2}\big) = \mathfrak{o}\big(p^2\big).\end{aligned}$$

Putting together the estimates for (I) and (II) yield that $\mathbb{E}\|\nabla T(X)\|_F^2 = \mathcal{O}(p^2)$ under the considered case $p > N$. The rest of the proof proceeds along the lines in the proof of Proposition 3.4, with the help of the variance formula in Proposition 3.11-(3). The normal approximation error bound becomes a constant multiple of

$$\frac{(1 \vee y) \cdot p}{p^2} = \frac{\frac{p}{n} \vee 1}{p} = \frac{1}{n \wedge p},$$

as desired. □

6.2.3. *Contiguity.*

PROOF OF PROPOSITION 3.11. We assume without loss of generality that $b(\Sigma) = \operatorname{tr}(\Sigma)/p = 1$ (otherwise we replace Σ by $\Sigma \cdot b^{-1}(\Sigma)$).

(1). By Lemma 6.2, with $S_Z \equiv N^{-1} \sum_{i=1}^N Z_i Z_i^\top$, we have

$$\begin{aligned}\mathscr{T}_{\Sigma;\mathrm{J}} &= \left\{ \frac{Z\Sigma^{1/2}\Sigma^{1/2} S_Z \Sigma^{1/2}}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - Z\Sigma^{1/2} \frac{b_2(\Sigma^{1/2} S_Z \Sigma^{1/2})}{b^3(\Sigma^{1/2} S_Z \Sigma^{1/2})} \right\} \Sigma^{1/2} \\ &= \frac{Z\Sigma S_Z \Sigma}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - Z\Sigma \frac{b_2(\Sigma^{1/2} S_Z \Sigma^{1/2})}{b^3(\Sigma^{1/2} S_Z \Sigma^{1/2})},\end{aligned}$$

so

$$\begin{aligned}&\mathscr{T}_{\Sigma;\mathrm{J}} - \mathscr{T}_{I;\mathrm{J}} \\ &= \left\{ \frac{Z\Sigma S_Z \Sigma}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{Z S_Z}{b^2(S_Z)} \right\} - \left\{ Z\Sigma \frac{b_2(\Sigma^{1/2} S_Z \Sigma^{1/2})}{b^3(\Sigma^{1/2} S_Z \Sigma^{1/2})} - Z \frac{b_2(S_Z)}{b^3(S_Z)} \right\} \\ &= \left\{ \frac{Z\Sigma S_Z \Sigma}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{Z S_Z}{b^2(S_Z)} \right\} - Z \left\{ \frac{b_2(\Sigma^{1/2} S_Z \Sigma^{1/2})}{b^3(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{b_2(S_Z)}{b^3(S_Z)} \right\} \\ &\quad - (Z\Sigma - Z) \cdot \frac{b_2(\Sigma^{1/2} S_Z \Sigma^{1/2})}{b^3(\Sigma^{1/2} S_Z \Sigma^{1/2})} \\ &\equiv V_1(Z) + V_2(Z) + V_3(Z).\end{aligned}$$

We will handle the Frobenius norms of $V_1(Z)$, $V_2(Z)$, $V_3(Z)$ separately below. For $V_1(Z)$,

$$\begin{aligned}\|V_1(Z)\|_F^2 &\lesssim \left\| \frac{Z\Sigma S_Z \Sigma}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{Z S_Z}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} \right\|_F^2 + \left\| \frac{Z S_Z}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{Z S_Z}{b^2(S_Z)} \right\|_F^2 \\ &= \|Z\Sigma S_Z \Sigma - Z S_Z\|_F^2 \cdot \frac{1}{b^4(\Sigma^{1/2} S_Z \Sigma^{1/2})} \\ &\quad + \|Z S_Z\|_F^2 \cdot \left[\frac{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2}) - b^2(S_Z)}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2}) b^2(S_Z)} \right]^2 \\ &\equiv V_{1,1} + V_{1,2}.\end{aligned}$$

Note that

$$\begin{aligned}V_{1,1} &\lesssim b^{-4}(\Sigma^{1/2} S_Z \Sigma^{1/2}) \big(\|Z\Sigma S_Z(\Sigma - I)\|_F^2 + \|Z(\Sigma - I) S_Z\|_F^2 \big) \\ &\lesssim \big[b^{-4}(\Sigma^{1/2} S_Z \Sigma^{1/2}) \cdot \|S_Z\|_{\mathrm{op}}^2 \big] \cdot \big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1 \big) \cdot \|Z\|_{\mathrm{op}}^2 \|\Sigma - I\|_F^2, \\ V_{1,2} &\le \|S_Z\|_{\mathrm{op}}^2 \|Z\|_F^2 b^{-4}\big(\Sigma^{1/2} S_Z \Sigma^{1/2}\big) b^{-4}(S_Z) \\ &\quad \times \big(\operatorname{tr}\big((\Sigma - I) S_Z\big)/p \big)^2 \big(b^2\big(\Sigma^{1/2} S_Z \Sigma^{1/2}\big) \vee b^2(S_Z) \big)\end{aligned}$$

$$\lesssim \big[\|S_Z\|_{\mathrm{op}}^4 b^{-4}(\Sigma^{1/2}S_Z\Sigma^{1/2}) b^{-4}(S_Z)\big(b^2(\Sigma^{1/2}S_Z\Sigma^{1/2}) \vee b^2(S_Z)\big)\big] \times p^{-1}\|Z\|_F^2\|\Sigma - I\|_F^2.$$

So under $p/N \le M$, by Lemma C.2 and Lemma 4.4, we have

$$\mathbb{E}\|V_1(Z)\|_F^2 \lesssim_M N\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.$$

For $V_2(Z)$,

$$\begin{aligned}\|V_2(Z)\|_F^2 &= \|Z\|_F^2\left(\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{b_2(S_Z)}{b^3(S_Z)}\right)^2 \\ &\lesssim \|Z\|_F^2\left\{\left(\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{b_2(S_Z)}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})}\right)^2\right. \\ &\quad \left. + \left(\frac{b_2(S_Z)}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{b_2(S_Z)}{b^3(S_Z)}\right)^2\right\} \\ &\equiv V_{2,1} + V_{2,2}.\end{aligned}$$

Note that

$$\begin{aligned}\big(b_2(\Sigma^{1/2}S_Z\Sigma^{1/2}) - b_2(S_Z)\big)^2 &= p^{-2}\operatorname{tr}^2\big(S_Z\Sigma S_Z\Sigma - S_Z^2\big) \\ &\lesssim p^{-2}\big\{\operatorname{tr}^2\big(S_Z(\Sigma - I)S_Z\Sigma\big) + \operatorname{tr}^2\big(S_Z^2(\Sigma - I)\big)\big\} \\ &\lesssim p^{-1}\|S_Z\|_{\mathrm{op}}^4\big(\|\Sigma\|_F^2/p + 1\big)\|\Sigma - I\|_F^2 \\ &\lesssim p^{-1}\|S_Z\|_{\mathrm{op}}^4\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2,\end{aligned} \tag{6.9}$$

so

$$\begin{aligned}V_{2,1} &\lesssim \big[b^{-6}(\Sigma^{1/2}S_Z\Sigma^{1/2})\|S_Z\|_{\mathrm{op}}^4\big]\cdot\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\cdot\big(p^{-1}\|Z\|_F^2\big)\cdot\|\Sigma - I\|_F^2, \\ V_{2,2} &\le \|Z\|_F^2 b^{-6}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-6}(S_Z)b_2^2(S_Z)\big(b(\Sigma^{1/2}S_Z\Sigma^{1/2}) - b(S_Z)\big)^2 \\ &\quad \times \big(b^2(\Sigma^{1/2}S_Z\Sigma^{1/2}) + b(\Sigma^{1/2}S_Z\Sigma^{1/2})b(S_Z) + b^2(S_Z)\big)^2 \\ &\lesssim \big[b^{-6}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-6}(S_Z)\big(b^4(\Sigma^{1/2}S_Z\Sigma^{1/2}) \vee b^4(S_Z)\big)\|S_Z\|_{\mathrm{op}}^6\big] \\ &\quad \times p^{-1}\|Z\|_F^2\|\Sigma - I\|_F^2.\end{aligned}$$

Hence under $p/N \le M$, by Lemma C.2 and Lemma 4.4, we have

$$\mathbb{E}\|V_2(Z)\|_F^2 \lesssim_M N\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.$$

Last, recall that $\operatorname{tr}(\Sigma) = p$ so using trace Hölder inequality we have $\operatorname{tr}(S_Z\Sigma S_Z\Sigma) \le \operatorname{tr}(\Sigma)\|S_Z\Sigma S_Z\|_{\mathrm{op}} \le p\|S_Z\|_{\mathrm{op}}^2\|\Sigma\|_{\mathrm{op}}$, so $V_3(Z)$ satisfies

$$\begin{aligned}\|V_3(Z)\|_F^2 &\le p^{-2}\|\Sigma - I\|_F^2\cdot\|Z\|_{\mathrm{op}}^2\cdot b^{-6}(\Sigma S_Z)\cdot\operatorname{tr}^2(S_Z\Sigma S_Z\Sigma) \\ &\le \big[b^{-6}(\Sigma^{1/2}S_Z\Sigma^{1/2})\|S_Z\|_{\mathrm{op}}^4\big]\cdot\|\Sigma\|_{\mathrm{op}}^2\cdot\|Z\|_{\mathrm{op}}^2\|\Sigma - I\|_F^2.\end{aligned}$$

Hence under $p/N \le M$, by Lemma C.2 and Lemma 4.4, we have

$$\mathbb{E}\|V_3(Z)\|_F^2 \lesssim_M N\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.$$

Combining the estimates proves the claim.

(2). Recall the normalization $b(\Sigma)=1$. Note that

$$\begin{aligned}
&\mathbb{E}\Big[\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\Big]\\
&\quad=\frac{\mathbb{E}b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}+\mathbb{E}\Big[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\Big(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\Big)\Big]\\
&\quad\overset{(*)}{=}\frac{(1+N^{-1})\frac{b_2(\Sigma)}{b^2(\Sigma)}+\frac{p}{N}}{1+2\operatorname{tr}(\Sigma^2)/(Np^2)}\\
&\qquad+\mathbb{E}\Big[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\Big(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\Big)\Big]\\
&\quad=\frac{b_2(\Sigma)}{b^2(\Sigma)}+\frac{p}{N}+\Big\{\mathbb{E}\Big[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\Big(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\Big)\Big]\\
&\qquad+\Big[(1+N^{-1})\frac{b_2(\Sigma)}{b^2(\Sigma)}+\frac{p}{N}\Big]\Big(\frac{1}{1+2\operatorname{tr}(\Sigma^2)\cdot(Np^2)^{-1}}-1\Big)+N^{-1}\frac{b_2(\Sigma)}{b^2(\Sigma)}\Big\}\\
&\quad\equiv\frac{b_2(\Sigma)}{b^2(\Sigma)}+\frac{p}{N}+R(\Sigma).
\end{aligned}$$

Here we use Lemma C.4-(1) in $(*)$ and

$$\begin{aligned}
R(\Sigma)&=\mathbb{E}\Big[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\Big(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\Big)\Big]\\
&\quad+\Big[(1+N^{-1})\frac{b_2(\Sigma)}{b^2(\Sigma)}+\frac{p}{N}\Big]\Big(\frac{1}{1+2\operatorname{tr}(\Sigma^2)\cdot(Np^2)^{-1}}-1\Big)+N^{-1}\frac{b_2(\Sigma)}{b^2(\Sigma)}\\
&\equiv R_1(\Sigma)+R_2(\Sigma)+R_3(\Sigma).
\end{aligned}$$

As

$$m_{\Sigma;\mathrm{J}}=\frac{N}{4}\mathbb{E}\operatorname{tr}\Big(\frac{\Sigma^{1/2}S_Z\Sigma^{1/2}}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})}-I\Big)^2=\frac{Np}{4}\Big\{\mathbb{E}\Big[\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^2(\Sigma S_Z)}\Big]-1\Big\},$$

we have

$$m_{\Sigma;\mathrm{J}}-m_{I;\mathrm{J}}=\frac{Np}{4}\big(p^{-1}\|\Sigma-I\|_F^2+R(\Sigma)-R(I)\big).$$

Now we handle $R_\ell(\Sigma)-R_\ell(I)$ for $\ell=1,2,3$.

For $\ell=1$,

$$\begin{aligned}
&\big|R_1(\Sigma)-R_1(I)\big|\\
&\quad=\Big|\mathbb{E}\Big[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\Big(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\Big)\\
&\qquad-b_2(S_Z)\Big(\frac{1}{b^2(S_Z)}-\frac{1}{\mathbb{E}b^2(S_Z)}\Big)\Big]\Big|\\
&\quad\le\Big|\mathbb{E}\Big[(b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})-b_2(S_Z))\Big(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\Big)\Big]\Big|\\
&\qquad+\Big|\mathbb{E}\Big[b_2(S_Z)\Big\{\Big(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{b^2(S_Z)}\Big)\\
&\qquad-\Big(\frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{1}{\mathbb{E}b^2(S_Z)}\Big)\Big\}\Big]\Big|
\end{aligned}$$

$$\equiv R_{1,1} + R_{1,2}.$$

The term $R_{1,1}$ can be handled as follows: by (6.9) Lemmas 4.4, C.2 and C.4, under $p/N \le M$,

$$\begin{aligned}
R_{1,1} &\lesssim \mathbb{E}^{1/4}\big(b_2(\Sigma^{1/2}S_Z\Sigma^{1/2}) - b_2(S_Z)\big)^4 \cdot \mathbb{E}^{1/4} b^{-8}(\Sigma^{1/2}S\Sigma^{1/2}) \\
&\quad \times \mathrm{Var}^{1/2}\big(b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big) \cdot \big(\mathbb{E}b^2(\Sigma^{1/2}S\Sigma^{1/2})\big)^{-1} \\
&\lesssim_M p^{-2} \cdot \big[p^{-1/2}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F\big] \cdot \mathrm{Var}^{1/2}\big(\mathrm{tr}^2(\Sigma S_Z)\big) \\
&\lesssim_M \big(N^{1/2}p\big)^{-1}\big(p^{-1}\|\Sigma\|_F^2 + 1\big)\|\Sigma - I\|_F.
\end{aligned}$$

For $R_{1,2}$, we have by Lemmas 4.4 and C.4 that, under $p/N \le M$,

$$\begin{aligned}
R_{1,2} &= \left|\mathbb{E}\left[b_2(S_Z)\left(\frac{b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})b^2(S_Z)} - \frac{\mathbb{E}b^2(S_Z) - \mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\mathbb{E}b^2(S_Z)}\right)\right]\right| \\
&\le \mathbb{E}b_2(S_Z)b^{-2}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-2}(S_Z) \\
&\quad \times \big|b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2}) - \mathbb{E}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big)\big| \\
&\quad + \big|\mathbb{E}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big)\big| \cdot \mathbb{E}b_2(S_Z) \\
&\quad \times \left|\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})b^2(S_Z)} - \frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\mathbb{E}b^2(S_Z)}\right| \\
&\lesssim_M \mathrm{Var}^{1/2}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big) \\
&\quad + \big|\mathbb{E}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big)\big| \cdot \big(\mathrm{Var}^{1/2}\big(b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big) \vee \mathrm{Var}^{1/2}\big(b^2(S_Z)\big)\big) \\
&\overset{(*)}{\lesssim}_M \big(N^{1/2}p\big)^{-1}\|\Sigma - I\|_F + p^{-1/2}\|\Sigma - I\|_F \cdot \big(N^{1/2}p\big)^{-1}\big(\|\Sigma\|_F \vee p^{1/2}\big) \\
&\lesssim \big(N^{1/2}p\big)^{-1}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F.
\end{aligned}$$

Here in $(*)$ we use the fact that

$$\begin{aligned}
\big|\mathbb{E}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big)\big| &\lesssim p^{-1}\mathbb{E}^{1/2}\,\mathrm{tr}^2\big((\Sigma - I)S_Z\big) \\
&\le p^{-1/2}\|\Sigma - I\|_F \cdot \mathbb{E}^{1/2}\|S_Z\|_{\mathrm{op}}^2 \lesssim_M p^{-1/2}\|\Sigma - I\|_F.
\end{aligned}$$

Hence

$$\big|R_1(\Sigma) - R_1(I)\big| \lesssim_M \big(N^{1/2}p\big)^{-1}\big(p^{-1}\|\Sigma\|_F^2 + 1\big)\|\Sigma - I\|_F.$$

For $\ell = 2$, with $\mathfrak{a}(\Sigma) \equiv 1/(1 + 2\,\mathrm{tr}(\Sigma^2)/(Np^2)) - 1$ (then $|\mathfrak{a}(\Sigma)| \le 2/N$ and $|\mathfrak{a}(I)| \le 2/(Np)$), we have

$$R_2(\Sigma) = \big(1 + N^{-1}\big)b_2(\Sigma)\mathfrak{a}(\Sigma) + N^{-1}p\mathfrak{a}(\Sigma),$$

so

$$\big|R_2(\Sigma) - R_2(I)\big| \lesssim_M \big|b_2(\Sigma)\mathfrak{a}(\Sigma) - b_2(I)\mathfrak{a}(I)\big| + \big|\mathfrak{a}(\Sigma) - \mathfrak{a}(I)\big| \equiv R_{2,1} + R_{2,2}.$$

The two terms $R_{2,1}$, $R_{2,2}$ can be handled as follows: using $\mathrm{tr}(\Sigma^2) \le p^2$ under $b(\Sigma) = 1$, we have

$$\begin{aligned}
R_{2,1} &\lesssim b_2(\Sigma)\big|\mathfrak{a}(\Sigma) - \mathfrak{a}(I)\big| + \big|\mathfrak{a}(I)\big|\big|b_2(\Sigma) - b_2(I)\big| \\
&\lesssim p^{-1}\,\mathrm{tr}\big(\Sigma^2\big)\big(Np^2\big)^{-1}\big|\mathrm{tr}\big(\Sigma^2 - I\big)\big| + (Np)^{-1} \cdot p^{-1} \cdot \big|\mathrm{tr}\big(\Sigma^2 - I\big)\big| \\
&\lesssim \big(Np^{1/2}\big)^{-1}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F, \\
R_{2,2} &\lesssim \big(Np^{1/2}\big)^{-1}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F,
\end{aligned}$$

so

$$\big|R_2(\Sigma) - R_2(I)\big| \lesssim_M \big(Np^{1/2}\big)^{-1}\big(\|\Sigma\|_F/p^{1/2} + 1\big)\|\Sigma - I\|_F.$$

For $\ell = 3$,

$$\big|R_3(\Sigma) - R_3(I)\big| = N^{-1}\big|b_2(\Sigma) - b_2(\mathrm{I})\big| \lesssim \big(Np^{1/2}\big)^{-1}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F.$$

Now with $Q_{\mathrm{J}}(\Sigma) \equiv p(R(\Sigma) - R(I))$, we have

$$\begin{aligned}\big|Q_{\mathrm{J}}(\Sigma)\big| &\lesssim_M p\max\big\{\big(N^{1/2}p\big)^{-1}, \big(Np^{1/2}\big)^{-1}\big\}\big(p^{-1}\|\Sigma\|_F^2 + 1\big)\|\Sigma - I\|_F \\ &\lesssim_M N^{-1/2}\big(p^{-1}\|\Sigma\|_F^2 + 1\big)\|\Sigma - I\|_F,\end{aligned}$$

and

$$m_{\Sigma;\mathrm{J}} - m_{I;\mathrm{J}} = \frac{N}{4}\big(\|\Sigma - I\|_F^2 + Q_{\mathrm{J}}(\Sigma)\big).$$

(3). Recall T_{LNW} defined in (3.7). Let

$$\Delta(X) \equiv T_{\mathrm{LNW}}(X) - T_{\mathrm{J}}(X).$$

Then for any $\varepsilon > 0$, there exists some $C_\varepsilon > 0$ such that under the null (i.e., $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(0, I_p)$),

$$\big[(1-\varepsilon)\sigma^2_{I;\mathrm{LNW}} - C_\varepsilon \operatorname{Var}_I(\Delta)\big]_+ \le \sigma^2_{I;\mathrm{J}} \le (1+\varepsilon)\sigma^2_{I;\mathrm{LNW}} + C_\varepsilon \operatorname{Var}_I(\Delta). \tag{6.10}$$

We will now bound $\operatorname{Var}_I(\Delta)$. By Lemmas 5.2-(1) and 6.2-(1), we have for any $i, j \in [N] \times [p]$

$$\begin{aligned}\partial_{(ij)}\Delta(X) &= \partial_{(ij)}T_{\mathrm{LNW}}(X) - \partial_{(ij)}T_{\mathrm{J}}(X) \\ &= \big(X(S-I) - N^{-1}\operatorname{tr}(S)X\big)_{ij} - \bigg[\frac{X}{b}\bigg(\frac{S}{b} - I\bigg) - \frac{X}{b(S)}\cdot b\bigg(\bigg(\frac{S}{b(S)} - I\bigg)^2\bigg)\bigg]_{ij} \\ &= \bigg[X(S-I) - \frac{X}{b}\bigg(\frac{S}{b} - I\bigg)\bigg]_{ij} + \big[Xb\big((S - b(S)I)^2\big)\big(b^{-3}(S) - 1\big)\big]_{ij} \\ &\quad + \big[X\big(b_2(S) - b^2(S) - N^{-1}\operatorname{tr}(S)\big)\big]_{ij} \\ &\equiv (\Delta_1 + \Delta_2 + \Delta_3)_{ij}.\end{aligned}$$

We now handle Δ_1–Δ_3 separately below. For Δ_1, by Lemmas 4.4, C.2 and C.4, we have

$$\begin{aligned}\mathbb{E}\|\Delta_1\|_F^2 &\lesssim \mathbb{E}b^{-2}(1-b)^2\|X(S-I)\|_F^2 + \mathbb{E}b^{-4}(1-b)^2\|XS\|_F^2 \\ &\le N\mathbb{E}b^{-2}(1-b)^2\|S\|_{\mathrm{op}}\|S-I\|_F^2 + N\mathbb{E}b^{-4}(b-1)^2\operatorname{tr}\big(S^3\big) \\ &\lesssim N\cdot(pN)^{-1}\cdot(1\vee y)\big(N^{-1}p^2\big) + N\mathbb{E}^{1/4}b^{-16}\mathbb{E}^{1/4}(b-1)^8\mathbb{E}^{1/2}\operatorname{tr}^2\big(S^3\big) \\ &\overset{(*)}{\lesssim} \mathfrak{o}\big(p^2\big) + N\cdot\mathcal{O}(1)\cdot(Np)^{-1}\cdot\mathcal{O}\big(N^{-2}p^3\vee p^2\big) = \mathfrak{o}\big(p^2\big).\end{aligned}$$

Here in $(*)$ the first bound follows by direct calculation and the second bound follows as: by Lemma C.5-(7),

$$\begin{aligned}\mathbb{E}\operatorname{tr}^2\big(S^3\big) &\le \mathbb{E}\operatorname{tr}^2\big(S^2\big)\|S\|_{\mathrm{op}}^2 \le \mathbb{E}^{1/2}\operatorname{tr}^4\big(S^2\big)\cdot\mathbb{E}^{1/2}\|S\|_{\mathrm{op}}^4 \\ &\lesssim p^4\cdot\big(y^2\vee 1\big) = \mathcal{O}\big(N^{-4}p^6\vee p^4\big).\end{aligned}$$

For Δ_2, using $b((S-b(S)I)^2) \le \|(S-b(S)I)^2\|_{\mathrm{op}} \lesssim \|S\|_{\mathrm{op}}^2 \vee b^2(S)$, we have

$$\begin{aligned}\mathbb{E}\|\Delta_2\|_F^2 &\lesssim \mathbb{E}b^{-6}\big(b^4 \vee b^2 \vee 1\big)(b-1)^2\big(\|S\|_{\mathrm{op}}^2 \vee b^2\big)\|X\|_F^2 \\ &\lesssim (pN)^{-1}\cdot(pN)\cdot\mathbb{E}^{1/2}\big(\|S\|_{\mathrm{op}}^4 \vee b^4\big) \asymp (1\vee y)^2 = \mathfrak{o}\big(p^2\big).\end{aligned}$$

For Δ_3, let $h(S) \equiv b_2(S) - b^2(S) - N^{-1}\operatorname{tr}(S)$, we have

$$\begin{aligned}\mathbb{E}\|\Delta_3\|_F^2 &\lesssim \mathbb{E}\|X\|_F^2 h^2(S) \le N\mathbb{E}^{1/2}\operatorname{tr}^2(S)\cdot\mathbb{E}^{1/2}h^4(S) \\ &\lesssim Np\cdot\big[\big(\mathbb{E}h(S)\big)^4 + \operatorname{Var}^2\big(h(S)\big) + \mathbb{E}\big\|\nabla h(S)\big\|_F^4\big]^{1/2},\end{aligned}$$

where the last inequality follows since

$$\begin{aligned}\mathbb{E}h^4(S) &= \big[\mathbb{E}h^2(S)\big]^2 + \operatorname{Var}\big(h^2(S)\big) \le 2\big[\mathbb{E}h(S)\big)^4 + 2\operatorname{Var}^2\big(h(S)\big) + \operatorname{Var}(h^2(S)\big] \\ &\le 2\big(\mathbb{E}h(S)\big)^4 + 2\operatorname{Var}^2\big(h(S)\big) + 4\mathbb{E}h^2(S)\big\|\nabla h(S)\big\|_F^2 \\ &\le 2\big(\mathbb{E}h(S)\big)^4 + 2\operatorname{Var}^2\big(h(S)\big) + 4\tau\mathbb{E}h^4(S) + C_\tau\mathbb{E}\big\|\nabla h(S)\big\|_F^4\end{aligned}$$

and choosing, say, $\tau = 1/8$. For $\mathbb{E}h(S)$, Lemma C.4 yields the direct evaluation

$$\begin{aligned}\mathbb{E}h(S) &= \frac{(1+N^{-1})p + N^{-1}p^2}{p} - \frac{p^2+2N^{-1}p}{p^2} - \frac{p}{N} \\ &= \big(1+N^{-1}\big) + \frac{p}{N} - \Big(1+\frac{2}{Np}\Big) - \frac{p}{N} = \frac{1}{N} - \frac{2}{Np} = \mathcal{O}\big(N^{-1}\big).\end{aligned}$$

For $\operatorname{Var}(h(S))$, the Gaussian–Poincaré inequality [5], Theorem 3.20, yields that

$$\begin{aligned}\operatorname{Var}\big(h(S)\big) &\le \mathbb{E}\sum_{ij}\big(\partial_{(ij)}h(S)\big)^2 \overset{(**)}{=} \sum_{ij}\mathbb{E}\Big(\frac{4X_i^\top Se_j}{Np} - \frac{4X_{ij}b(S)}{Np} - \frac{2X_{ij}}{N^2}\Big)^2 \\ &\lesssim (Np)^{-2}\big(\mathbb{E}\|XS\|_F^2 + \mathbb{E}b^2(S)\|X\|_F^2\big) + N^{-4}\mathbb{E}\|X\|_F^2 \\ &\lesssim \big(Np^2\big)^{-1}\mathbb{E}\operatorname{tr}\big(S^2\big)\|S\|_{\mathrm{op}} + (Np)^{-2}(Np) + N^{-4}(Np) \\ &\lesssim \big(Np^2\big)^{-1}\mathbb{E}^{1/2}\operatorname{tr}^2\big(S^2\big)\mathbb{E}^{1/2}\|S\|_{\mathrm{op}}^2 + (Np)^{-1} + pN^{-3} \\ &\lesssim \big(Np^2\big)^{-1}\cdot p^2\cdot(1\vee y) + (Np)^{-1} + pN^{-3} = \mathfrak{o}\big(N^{-1}p\big).\end{aligned}$$

Here $(**)$ follows from (6.8). Last $\mathbb{E}\|\nabla h(S)\|_F^4$ can be bounded similarly:

$$\mathbb{E}\big\|\nabla h(S)\big\|_F^4 \lesssim (Np)^{-4}\big(\mathbb{E}\|XS\|_F^4 + \mathbb{E}b^4(S)\|X\|_F^4\big) + N^{-8}\mathbb{E}\|X\|_F^4 = \mathfrak{o}\big(N^{-2}p^2\big).$$

Now by the Gaussian–Poincaré inequality ([5], Theorem 3.20),

$$\operatorname{Var}_I(\Delta) \le \mathbb{E}\sum_{ij}\big(\partial_{(ij)}\Delta(X)\big)^2 \lesssim \mathbb{E}\|\Delta_1\|_F^2 + \mathbb{E}\|\Delta_2\|_F^2 + \mathbb{E}\|\Delta_3\|_F^2 = \mathfrak{o}\big(p^2\big).$$

As $\sigma_{I;\mathrm{LNW}}^2 \sim p^2/4 \to \infty$ whenever $N\wedge p \to \infty$, by taking ε in (6.10) slowly decaying to 0 we conclude $\sigma_{I;\mathrm{LNW}}^2 \sim \sigma_{I;\mathrm{J}}^2$.

(4). By (1)–(2), as $\|\Sigma\|_F^2/p \lesssim \|\Sigma - I\|_F^2/p + 1 \lesssim \|\Sigma - I\|_F \vee 1$ [where we use $\|\Sigma - I\|_F \le \|\Sigma\|_F + \sqrt{p} \le p + \sqrt{p}$ under $\operatorname{tr}(\Sigma) = p$], we only need to prove that given C_0, we may find some constant $C_1 > 0$,

$$\frac{\sqrt{N}\|\Sigma - I\|_F(1\vee\|\Sigma - I\|_F)}{(N\|\Sigma - I\|_F^2 - C_0N^{1/2}\|\Sigma - I\|_F(1\vee\|\Sigma - I\|_F))_+ \vee \sigma_{I;\mathrm{J}}} \le \frac{C_1}{(\sigma_{I;\mathrm{J}}\wedge N)^{1/2}}. \tag{6.11}$$

Write $\alpha = \|\Sigma - I\|_F$, we only need to prove that

$$\frac{\sqrt{N}\alpha \vee \sqrt{N}\alpha^2}{(N\alpha^2 - C_0 N^{1/2}\alpha)_+ \vee \sigma_{I;\mathrm{J}}} \leq \frac{C_1}{(\sigma_{I;\mathrm{J}} \wedge N)^{1/2}}. \tag{6.12}$$

This follows as

$$\begin{aligned}
\text{LHS of (6.12)} &\lesssim \mathbf{1}_{\alpha \leq 2C_0 N^{-1/2}} \frac{1}{\sigma_{I;J}} + \mathbf{1}_{\alpha > 2C_0 N^{-1/2}} \frac{\sqrt{N}\alpha \vee \sqrt{N}\alpha^2}{N\alpha^2 \vee \sigma_{I;\mathrm{J}}} \\
&\lesssim \frac{1}{\sigma_{I;J}} + \frac{1}{N^{1/2} \inf_{\alpha \geq 0}(\alpha \vee \frac{\sigma_{I;\mathrm{J}}/N}{\alpha})} + \frac{1}{N^{1/2}} \\
&\lesssim \frac{1}{\sigma_{I;J}} + \frac{1}{\sigma_{I;\mathrm{J}}^{1/2}} + \frac{1}{N^{1/2}} \asymp \frac{1}{(\sigma_{I;\mathrm{J}} \wedge N)^{1/2}}.
\end{aligned}$$

The proof is complete. □

6.2.4. *Completing of the proof for power expansion.*

PROOF OF THEOREM 3.12. The proof essentially follows that of Theorem 3.6 by noting that the key property used therein is $|Q_{\mathrm{LNW}}(\Sigma)| \leq C_M N^{-1/2}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F$, while here we have $|Q_J(\Sigma \cdot b^{-1}(\Sigma))| \leq C_M N^{-1/2}\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F \leq N^{-1/2}(\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F \vee 1)\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F$. □

Acknowledgments. The authors would like to thank an anonymous referee, an Associate Editor and the Editor for their very helpful comments and suggestions that significantly improved the quality of the paper.

Funding. Q. Han was supported by NSF Grants DMS-1916221 and DMS-2143468. T. Jiang was supported by NSF Grant DMS-1916014.

SUPPLEMENTARY MATERIAL

Supplement: Additional proofs (DOI: 10.1214/22-AAP1917SUPP; .pdf). In the supplement [12], we collect the proof of auxiliary lemmas in the appendix.

REFERENCES

[1] BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444 https://doi.org/10.1214/09-AOS694

[2] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

[3] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

[4] BIRKE, M. and DETTE, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. MR2189467 https://doi.org/10.1016/j.spl.2005.04.051

[5] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*: *A nonasymptotic theory of independence*. Oxford Univ. Press, Oxford. With a foreword by Michel Ledoux. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[6] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441

[7] CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

[8] CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121 https://doi.org/10.1007/s00440-007-0118-6

[9] CHATTERJEE, S. (2014). *Superconcentration and Related Topics*. *Springer Monographs in Mathematics*. Springer, Cham. MR3157205 https://doi.org/10.1007/978-3-319-03886-5

[10] CHEN, H. and JIANG, T. (2018). A study of two high-dimensional likelihood ratio tests under alternative hypotheses. *Random Matrices Theory Appl.* **7** 1750016, 23. MR3756424 https://doi.org/10.1142/S2010326317500162

[11] CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863 https://doi.org/10.1198/jasa.2010.tm09560

[12] HAN, Q., JIANG, T. and SHEN, Y. (2023). Supplement to "Contiguity under high-dimensional Gaussianity with applications to covariance testing." https://doi.org/10.1214/22-AAP1917SUPP

[13] HAN, Q., SEN, B. and SHEN, Y. (2022). High-dimensional asymptotics of likelihood ratio tests in the Gaussian sequence model under convex constraints. *Ann. Statist.* **50** 376–406. MR4382021 https://doi.org/10.1214/21-aos2111

[14] JIANG, D., JIANG, T. and YANG, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Statist. Plann. Inference* **142** 2241–2256. MR2911842 https://doi.org/10.1016/j.jspi.2012.02.057

[15] JIANG, T. (2019). Determinant of sample correlation matrix with application. *Ann. Appl. Probab.* **29** 1356–1397. MR3914547 https://doi.org/10.1214/17-AAP1362

[16] JIANG, T. and QI, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.* **42** 988–1009. MR3426306 https://doi.org/10.1111/sjos.12147

[17] JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857 https://doi.org/10.1214/13-AOS1134

[18] JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568 https://doi.org/10.1093/biomet/58.1.123

[19] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[20] KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

[21] LE CAM, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.* **3** 37–98. MR0126903

[22] LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169 https://doi.org/10.1214/aos/1031689018

[23] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0652932

[24] NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

[25] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 https://doi.org/10.1214/13-AOS1100

[26] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485 https://doi.org/10.1214/13-AOS1181

[27] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. MR3199978 https://doi.org/10.1214/13-AAP939

[28] RUDELSON, M. and VERSHYNIN, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707–1739. MR2569075 https://doi.org/10.1002/cpa.20294

[29] SRIVASTAVA, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.* **35** 251–272. MR2328427 https://doi.org/10.14490/jjss.35.251

[30] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[31] WANG, Q. and YAO, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.* **7** 2164–2192. MR3104916 https://doi.org/10.1214/13-EJS842

[32] ZHENG, S., BAI, Z. and YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292