Electronic Journal of Statistics
Vol. 10 (2016) 1751–1806
ISSN: 1935-7524
DOI: 10.1214/15-EJS1048

Bernstein-von Mises theorems for functionals of the covariance matrix

Chao Gao and Harrison H. Zhou*

Department of Statistics
Yale University
New Haven, CT 06511, USA
e-mail: chao.gao@yale.edu; huibin.zhou@yale.edu

Abstract: We provide a general theoretical framework to derive Bernstein-von Mises theorems for functionals of the covariance matrix and its inverse. The conditions on functionals and priors are explicit and easy to check. Results are obtained for various functionals including entries of covariance matrix, entries of precision matrix, quadratic forms, log-determinant, eigenvalues in the Bayesian Gaussian covariance/precision matrix estimation setting, as well as for Bayesian linear and quadratic discriminant analysis.

MSC 2010 subject classifications: 62G05, 62G20.
Keywords and phrases: Bernstein-von Mises theorem, Bayes nonparametrics, covariance matrix.

Received December 2014.

Contents

* The research of Chao Gao and Harrison H. Zhou is supported in part by NSF Grant DMS-1209191.

1. Introduction

The celebrated Bernstein-von Mises (BvM) theorem [24, 3, 30, 25, 28] justifies Bayesian methods from a frequentist point of view. It bridges the gap between Bayesians and frequentists. Consider a parametric model $(P_\theta : \theta \in \Theta)$, and a prior distribution $\theta \sim \Pi$. Suppose we have i.i.d. observations $X^n = (X_1, \ldots, X_n)$ from the product measure $P_{\theta^*}^n$. Under some weak assumptions, Bernstein-von Mises theorem shows that the conditional distribution of

$$\sqrt{n}(\theta - \hat{\theta})|X^n$$

is asymptotically $N(0, V^2)$ under the distribution $P_{\theta^*}^n$ with some centering $\hat{\theta}$ and covariance V^2 when $n \to \infty$. In a local asymptotic normal (LAN) family, the centering $\hat{\theta}$ can be taken as the maximum likelihood estimator (MLE) and V^2 as the inverse of the Fisher information matrix. An immediate consequence of the Bernstein-von Mises theorem is that the distributions

$$\sqrt{n}(\theta - \hat{\theta})|X^n \quad \text{and} \quad \sqrt{n}(\hat{\theta} - \theta)|\theta = \theta^*$$

are asymptotically the same under the sampling distribution $P_{\theta^*}^n$. Note that the first one, known as the posterior, is of interest to Bayesians, and the second one is of interest to frequentists in the large sample theory. Applications of Bernstein-von Mises theorem include constructing confidence sets from Bayesian methods with frequentist coverage guarantees.

Besides the classical parametric setting where the parameter has a fixed dimension, BvM results can also be established when the unknown parameter is of increasing dimensions. For example, [19, 4, 11] established BvM results in high-dimensional linear models, and [18, 20, 5, 15] established BvM results in

high-dimensional exponential models, given certain conditions on the growth rate of the dimension. In this paper, we consider the question whether it is possible to have BvM results for matrix functionals, such as matrix entries and eigenvalues, when the dimension of the matrix p grows with the sample size n.

This paper provides some positive answers to this question. To be specific, we consider a multivariate Gaussian likelihood and put a prior on the covariance matrix. We prove that the posterior distribution has a BvM behavior for various matrix functionals including entries of the covariance matrix, entries of the precision matrix, quadratic forms, log-determinant, and eigenvalues. All of these conclusions are obtained from a general theoretical framework we provide in Section 2, where we propose explicit easy-to-check conditions on both functionals and priors. We illustrate the theory by both conjugate and non-conjugate priors. A slight extension of the general framework leads to BvM results for discriminant analysis. Both linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) are considered.

This work is inspired by a growing interest in studying the BvM phenomena on a low-dimensional functional of the whole parameter. That is, the asymptotic distribution of

$$\sqrt{n}(f(\theta) - \hat{f})|X^n,$$

with f being a map from Θ to $\mathbb{R}^d$, where d does not grow with n. A special case is the semiparametric setting, where $\theta = (\mu, \eta)$ contains both a parametric part μ and a nonparametric part η. The functional f takes the form of $f(\mu, \eta) = \mu$. To the best of our knowledge, the first general framework for semiparametric BvM with conditions cleanly stated and easy to check is the beautiful work by [9], in which the recent advancement in Bayes nonparametrics such as [21] is nicely absorbed. [27] proves BvM for linear functionals for which the distribution of $\sqrt{n}(f(\theta) - \hat{f})|X^n$ converges to a mixture of normal instead of a normal. At the point when this paper is drafted, the most updated theory is due to [10], which provides conditions for BvM to hold for general functionals. The general framework we provide for matrix functional BvM is greatly inspired by the framework developed in [10] for functionals in nonparametrics. However, the theory in this paper is different from theirs since we can take advantage of the structure in the Gaussian likelihood and avoid unnecessary expansion and approximation. Hence, in the covariance matrix functional case, our assumptions can be significantly weaker. A related line of development is the semiparametric BvM results using profile likelihood [12, 13, 14].

The paper is organized as follows. In Section 2, we state the general theoretical framework of our results. It is illustrated with two priors, one conjugate prior and one non-conjugate prior. Section 3 considers specific examples of matrix functionals and the associated BvM results. The extension to discriminant analysis is developed in Section 4. Finally, we devote Section 5 to some discussions on the assumptions and possible generalizations. Most of the proofs are gathered in Section 6.

1.1. Notation

Given a matrix A, we use $||A||$ to denote its spectral norm, and $||A||_F$ to denote its Frobenius norm. The norm $||\cdot||$, when applied to a vector, is understood to be the usual vector norm. Let S^{p-1} be the unit sphere in $\mathbb{R}^p$. For any $a, b \in \mathbb{R}$, we use notation $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. The probability P_Σ stands for $N(0, \Sigma)$ and $P_{(\mu,\Omega)}$ is for $N(\mu, \Omega^{-1})$. In most cases, we use Σ to denote the covariance matrix, and Ω to denote the precision matrix (including those with superscripts or subscripts). The notation $\mathbb{P}$ is for a generic probability, whenever the distribution is clear in the context. We use $O_P(\cdot)$ and $o_P(\cdot)$ to denote stochastic orders under the sampling distribution of the data. We use C to indicate constants throughout the paper. They may be different from line to line.

2. A general framework

Consider i.i.d. samples $X^n = (X_1, \ldots, X_n)$ drawn from $N(0, \Sigma^*)$, where Σ^* is a $p \times p$ covariance matrix with inverse Ω^*. A Bayes method puts a prior Π on the precision matrix Ω, and the posterior distribution is defined as

$$\Pi(B|X^n) = \frac{\int_B \exp\Big(l_n(\Omega)\Big) d\Pi(\Omega)}{\int \exp\Big(l_n(\Omega)\Big) d\Pi(\Omega)},$$

where $l_n(\Omega)$ is the log-likelihood of $N(0, \Omega^{-1})$ defined as

$$l_n(\Omega) = \frac{n}{2} \log\det(\Omega) - \frac{n}{2} \mathrm{tr}(\Omega\hat{\Sigma}), \quad \text{where } \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^T.$$

We deliberately omit the logarithmic normalizing constant in $l_n(\Omega)$ for simplicity and it will not affect the definition of the posterior distribution. Note that specifying a prior on the precision matrix Ω is equivalent to specifying a prior on the covariance matrix Ω^{-1}. The goal of this work is to show that the asymptotic distribution of the functional $f(\Omega)$ under the posterior distribution is approximately normal, i.e.,

$$\Pi\Big(\sqrt{n}V^{-1}\big(f(\Omega) - \hat{f}\big) \le t | X^n\Big) \to \mathbb{P}(Z \le t),$$

where $Z \sim N(0, 1)$, as $(n, p) \to \infty$ jointly with some appropriate centering $\hat{f}$ and variance V^2. In this paper, we choose the centering $\hat{f}$ to be the sample version of $f(\Omega) = f(\Sigma^{-1})$, where Σ is replaced by the sample covariance $\hat{\Sigma}$, and compare the BvM results with the classical asymptotical normality for $\hat{f}$ in the frequentist sense. Other centering $\hat{f}$, including bias correction on the sample version, will be considered in the future work. The paper focuses on Gaussian distribution, though the idea of the general framework is possible to be extended to other distributions.

We first provide a framework for approximately linear functionals, and then use the general theory to derive results for specific examples of priors and functionals. For clarity of presentation, we consider the cases of functionals of Σ and functionals of Ω separately. Though a functional of Σ is also a functional of Ω, we treat them separately, since some functional may be "more linear" in Σ than in Ω, or the other way around.

2.1. Functional of the covariance matrix

Let us first consider a functional of Σ, $f = \phi(\Sigma)$. The functional is approximately linear in a neighborhood of the truth. We assume there is a set A_n satisfying

$$A_n \subset \{||\Sigma - \Sigma^*|| \leq \delta_n\}, \tag{2.1}$$

for any sequence $\delta_n = o(1)$, on which $\phi(\Sigma)$ is approximately linear in the sense that there exists a symmetric matrix Φ such that

$$\sup_{A_n} \sqrt{n} \left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^{-1} \left|\phi(\Sigma) - \phi(\hat{\Sigma}) - \operatorname{tr}\Big((\Sigma - \hat{\Sigma})\Phi\Big)\right| = o_P(1). \tag{2.2}$$

When $\phi(\Sigma)$ is exactly linear in the sense that $\phi(\Sigma) - \phi(\hat{\Sigma}) = \operatorname{tr}((\Sigma - \hat{\Sigma})\Phi)$, (2.2) automatically holds. The main result is stated in the following theorem.

Theorem 2.1. *Under the assumptions of (2.2) and* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$*, if for a given prior* Π*, the following two conditions are satisfied:*

1. $\Pi(A_n|X^n) = 1 - o_P(1)$,
2. *For any fixed* $t \in \mathbb{R}$, $\frac{\int_{A_n} \exp(l_n(\Omega_t))d\Pi(\Omega)}{\int_{A_n} \exp(l_n(\Omega))d\Pi(\Omega)} = 1 + o_P(1)$ *for the perturbed precision matrix*

$$\Omega_t = \Omega + \frac{\sqrt{2}t}{\sqrt{n}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F}\Phi,$$

then

$$\sup_{t\in\mathbb{R}} \left|\Pi\left(\frac{\sqrt{n}\big(\phi(\Sigma) - \phi(\hat{\Sigma})\big)}{\sqrt{2}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F} \leq t \middle| X^n\right) - \mathbb{P}\big(Z \leq t\big)\right| = o_P(1),$$

where $Z \sim N(0,1)$.

The theorem gives explicit conditions on both prior and functional. The first condition says that the posterior distribution concentrates on a neighborhood of the truth under the spectral norm, on which the functional is approximately linear. The second condition says that the bias caused by the shifted parameter can be absorbed by the posterior distribution. Under both conditions, Theorem 2.1 shows that the asymptotic posterior distribution of $\phi(\Sigma)$ is

$$N\left(\phi(\hat{\Sigma}), 2n^{-1}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^2\right).$$

2.2. Functional of the precision matrix

We state a corresponding theorem for functionals of precision matrix in this section. The condition for linear approximation is slightly different. Consider the functional $f = \psi(\Omega)$. Let A_n be a set satisfying

$$A_n \subset \{\sqrt{rp}||\Sigma - \Sigma^*|| \leq \delta_n\}, \tag{2.3}$$

for some integer $r > 0$ and any sequence $\delta_n = o(1)$. We assume the functional $\psi(\Omega)$ is approximately linear on A_n in the sense that there exists a symmetric matrix Ψ satisfying $\text{rank}(\Psi) \leq r$, such that

$$\sup_{A_n} \sqrt{n} \left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|_F^{-1} \left|\psi(\Omega) - \psi(\hat{\Sigma}^{-1}) - \text{tr}\Big((\Omega - \hat{\Sigma}^{-1})\Psi\Big)\right| = o_P(1). \tag{2.4}$$

The main result is stated in the following theorem.

Theorem 2.2. *Under the assumptions of (2.4), $rp^2/n = o(1)$ and $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$, if for a given prior Π, the following conditions are satisfied:*

1. $\Pi(A_n|X^n) = 1 - o_P(1)$,
2. *For any fixed $t \in \mathbb{R}$, $\frac{\int_{A_n} \exp(l_n(\Omega_t))d\Pi(\Omega)}{\int_{A_n} \exp(l_n(\Omega))d\Pi(\Omega)} = 1 + o_P(1)$ for the perturbed precision matrix*

$$\Omega_t = \Omega - \frac{\sqrt{2}t}{\sqrt{n}\left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|_F}\Omega^*\Psi\Omega^*,$$

then

$$\sup_{t\in\mathbb{R}} \left|\Pi\left(\frac{\sqrt{n}\big(\psi(\Omega) - \psi(\hat{\Sigma}^{-1})\big)}{\sqrt{2}\left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|_F} \leq t \middle| X^n\right) - \mathbb{P}\big(Z \leq t\big)\right| = o_P(1),$$

where $Z \sim N(0,1)$.

Remark 2.1. The extra condition $rp^2/n = o(1)$ does not appear in Theorem 2.1. We show that this condition is indeed sharp for Theorem 2.2 in Section 5.3 in comparison with the asymptotics of MLE.

2.3. Priors

In this section, we provide examples of priors. In particular, we consider both a conjugate prior and a non-conjugate prior. A covariance matrix prior can be viewed as a vector prior with an additional constraint of positive semidefiniteness. Note that the result of a conjugate prior can be derived by directly exploring the posterior form without applying our general theory. However, the general framework provided in this paper can handle both conjugate and non-conjugate priors in a unified way.

2.3.1. Wishart prior

Consider the Wishart prior $\mathcal{W}_p(I, p+b-1)$ on Ω with density function

$$\frac{d\Pi(\Omega)}{d\Omega} \propto \exp\left(\frac{b-2}{2}\log\det(\Omega) - \frac{1}{2}\mathrm{tr}(\Omega)\right), \tag{2.5}$$

supported on the set of symmetric positive semi-definite matrices.

Lemma 2.1. *Assume* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$ *and* $p/n = o(1)$. *Then, for any integer* $b = O(1)$, *the prior* $\Pi = \mathcal{W}_p(I, p+b-1)$ *satisfies the two conditions in Theorem 2.1 for some* A_n. *If the extra assumption* $rp^2/n = o(1)$ *is made, the two conditions in Theorem 2.2 are also satisfied for some* A_n.

Remark 2.2. In the proof of Lemma 2.1 (Section 6.2), we set

$$A_n = \left\{||\Sigma - \Sigma^*|| \leq M\sqrt{\frac{p}{n}}\right\},$$

for some $M > 0$.

2.3.2. Gaussian prior

Consider Gaussian prior on Ω with density function

$$\frac{d\Pi(\Omega)}{d\Omega} \propto \exp\Big(-\frac{1}{2}||\Omega||_F^2\Big), \tag{2.6}$$

supported on the following set

$$\left\{\Omega = \Omega^T \succeq 0, ||\Omega|| < 2\Lambda, ||\Sigma|| \leq 2\Lambda\right\},$$

for some constant $\Lambda > 0$.

Lemma 2.2. *Assume* $||\Sigma^*|| \vee ||\Omega^*|| \leq \Lambda = O(1)$ *and* $\frac{p^2\log n}{n} = o(1)$. *The Gaussian prior* Π *defined above satisfies the two conditions in Theorem 2.1 for some appropriate* A_n. *If the extra assumption* $\frac{rp^3\log n}{n} = o(1)$ *is made, the two conditions in Theorem 2.2 are also satisfied for some appropriate* A_n.

Remark 2.3. In the proof of Lemma 2.2 (Section 6.3), we set

$$A_n = \left\{||\Sigma - \Sigma^*||_F \leq M\sqrt{\frac{p^2\log n}{n}}\right\},$$

for some constant $M > 0$.

3. Examples of matrix functionals

We consider various examples of functionals in this section. The two conditions of Theorem 2.1 and Theorem 2.2 are satisfied by Wishart prior and Gaussian prior, as is shown in Lemma 2.1 and Lemma 2.2 respectively. Hence, it is sufficient to check the approximate linearity of the functional with respect to Σ or Ω for the BvM result to hold. Among the four examples we consider, the first two are exactly linear and the last two are approximately linear. In the below examples, Z is always a random variable distributed as $N(0,1)$.

3.1. Entry-wise functional

We consider the elementwise functional $\sigma_{ij} = \phi_{ij}(\Sigma)$ and $\omega_{ij} = \psi_{ij}(\Omega)$. Note that these two functionals are linear with respect to Σ and Ω respectively. For σ_{ij}, we write

$$\sigma_{ij} = \mathrm{tr}\Big(\Sigma\big(\frac{1}{2}E_{ij} + \frac{1}{2}E_{ji}\big)\Big),$$

where the matrix E_{ij} is the (i,j)-th basis in $\mathbb{R}^{p\times p}$ with 1 on its (i,j)-the element and 0 elsewhere. For ω_{ij}, we write

$$\omega_{ij} = \mathrm{tr}\Big(\Omega\big(\frac{1}{2}E_{ij} + \frac{1}{2}E_{ji}\big)\Big).$$

Note that $\mathrm{rank}(\frac{1}{2}E_{ij} + \frac{1}{2}E_{ji}) \le 2$. Hence, the corresponding matrices Φ and Ψ in the linear expansion of ϕ and ψ are $\frac{1}{2}E_{ij} + \frac{1}{2}E_{ji}$. In view of Theorem 2.1 and Theorem 2.2, the asymptotic variance for $\sqrt{n}\big(\phi(\Sigma) - \phi(\hat{\Sigma})\big)$ is

$$2\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^2 = \sigma_{ii}^*\sigma_{jj}^* + \sigma_{ij}^{*2}.$$

The asymptotic variance for $\sqrt{n}\big(\psi(\Omega) - \psi(\hat{\Sigma}^{-1})\big)$ is

$$2\left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|_F^2 = \omega_{ii}^*\omega_{jj}^* + \omega_{ij}^{*2}.$$

Plugging these quantities in Theorem 2.1, Theorem 2.2, Lemma 2.1, and Lemma 2.2, we have the following Bernstein-von Mises results.

Corollary 3.1. *Consider the Wishart prior $\Pi = \mathcal{W}_p(I, p+b-1)$ in (2.5) with integer $b = O(1)$. Assume $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$ and $p/n = o(1)$, then we have*

$$P_{\Sigma^*}^n \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(\sigma_{ij} - \hat{\sigma}_{ij})}{\sqrt{\sigma_{ii}^*\sigma_{jj}^* + \sigma_{ij}^{*2}}} \le t \middle| X^n \right) - \mathbb{P}\big(Z \le t\big) \right| \to 0,$$

where $\hat{\sigma}_{ij}$ is the (i,j)-th element of the sample covariance $\hat{\Sigma}$. If we additionally assume $p^2/n = o(1)$, then

$$P_{\Sigma^*}^n \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(\omega_{ij} - \hat{\omega}_{ij})}{\sqrt{\omega_{ii}^*\omega_{jj}^* + \omega_{ij}^{*2}}} \le t \middle| X^n \right) - \mathbb{P}\big(Z \le t\big) \right| \to 0,$$

where $\hat{\omega}_{ij}$ is the (i,j)-th element of $\hat{\Sigma}^{-1}$.

Corollary 3.2. *Consider the Gaussian prior* Π *in (2.6). Assume* $||\Sigma^*||\vee||\Omega^*|| \leq \Lambda = O(1)$ *and* $\frac{p^2 \log n}{n} = o(1)$*, then we have*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(\sigma_{ij} - \hat{\sigma}_{ij})}{\sqrt{\sigma^*_{ii}\sigma^*_{jj} + \sigma^{*2}_{ij}}} \leq t \middle| X^n \right) - \mathbb{P}\big(Z \leq t\big) \right| \to 0.$$

If we additionally assume $\frac{p^3 \log n}{n} = o(1)$*, then*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(\omega_{ij} - \hat{\omega}_{ij})}{\sqrt{\omega^*_{ii}\omega^*_{jj} + \omega^{*2}_{ij}}} \leq t \middle| X^n \right) - \mathbb{P}\big(Z \leq t\big) \right| \to 0,$$

where $\hat{\sigma}_{ij}$ *and* $\hat{\omega}_{ij}$ *are defined in Corollary 3.1.*

3.2. Quadratic form

Consider the functional $\phi_v(\Sigma) = v^T\Sigma v = \mathrm{tr}(\Sigma vv^T)$ and $\psi_v(\Omega) = v^T\Omega v = \mathrm{tr}(\Omega vv^T)$ for some $v \in \mathbb{R}^p$. Therefore, the corresponding matrices Φ and Ψ are vv^T. It is easy to see that $\mathrm{rank}(vv^T) = 1$. The asymptotic variances are

$$2\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|^2_F = 2|v^T\Sigma^* v|^2, \quad 2\left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|^2_F = 2|v^T\Omega^* v|^2.$$

Plugging these representations in Theorem 2.1, Theorem 2.2, Lemma 2.1 and Lemma 2.2, we have the following Bernstein-von Mises results.

Corollary 3.3. *Consider the Wishart prior* $\Pi = \mathcal{W}_p(I, p+b-1)$ *in (2.5) with integer* $b = O(1)$*. Assume* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$ *and* $p/n = o(1)$*, then we have*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(v^T\Sigma v - v^T\hat{\Sigma}v)}{\sqrt{2}|v^T\Sigma^* v|} \leq t \middle| X^n \right) - \mathbb{P}\big(Z \leq t\big) \right| \to 0.$$

If we additionally assume $p^2/n = o(1)$*, then*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(v^T\Omega v - v^T\hat{\Sigma}^{-1}v)}{\sqrt{2}|v^T\Omega^* v|} \leq t \middle| X^n \right) - \mathbb{P}\big(Z \leq t\big) \right| \to 0.$$

Corollary 3.4. *Consider the Gaussian prior* Π *in (2.6). Assume* $||\Sigma^*||\vee||\Omega^*|| \leq \Lambda = O(1)$ *and* $\frac{p^2 \log n}{n} = o(1)$*, then we have*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(v^T\Sigma v - v^T\hat{\Sigma}v)}{\sqrt{2}|v^T\Sigma^* v|} \leq t \middle| X^n \right) - \mathbb{P}\big(Z \leq t\big) \right| \to 0.$$

If we additionally assume $\frac{p^3 \log n}{n} = o(1)$*, then*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(v^T\Omega v - v^T\hat{\Sigma}^{-1}v)}{\sqrt{2}|v^T\Omega^* v|} \leq t \middle| X^n \right) - \mathbb{P}\big(Z \leq t\big) \right| \to 0.$$

Remark 3.1. The entry-wise functional and the quadratic form are both special cases of the functional $u^T\Sigma v$ for some $u, v \in \mathbb{R}^p$. It is direct to apply the general framework to this functional and obtain the result

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(u^T\Sigma v - u^T\hat{\Sigma}v)}{\sqrt{|u^T\Sigma^* v|^2 + |u^T\Sigma^* u||v^T\Sigma^* v|}} \le t \middle| X^n \right) - \mathbb{P}\big(Z \le t\big) \right| \to 0.$$

Similarly, for the functional $u^T\Omega v$ for some $u, v \in \mathbb{R}^p$, we have

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(u^T\Omega v - u^T\hat{\Sigma}^{-1}v)}{\sqrt{|u^T\Omega^* v|^2 + |u^T\Omega^* u||v^T\Omega^* v|}} \le t \middle| X^n \right) - \mathbb{P}\big(Z \le t\big) \right| \to 0,$$

Both results can be derived under the same conditions of Corollary 3.3 and Corollary 3.4.

3.3. Log determinant

In this section, we consider the log-determinant functional. That is $\phi(\Sigma) = \log\det(\Sigma)$. Different from entry-wise functional and quadratic form, we do not need to consider $\log\det(\Omega)$ because of the simple observation

$$\log\det(\Omega) = -\log\det(\Sigma).$$

The following lemma establishes the approximate linearity of $\log\det(\Sigma)$.

Lemma 3.1. *Assume* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$ *and* $p^3/n = o(1)$*, then for any* $\delta_n = o(1)$*, we have*

$$\sup_{\left\{ \begin{smallmatrix} \sqrt{n/p}||\Sigma-\Sigma^*||_F^2 \\ \vee\sqrt{p}||\Sigma-\Sigma^*||_F \le \delta_n \end{smallmatrix} \right\}} \sqrt{\frac{n}{p}} \left| \log\det(\Sigma) - \log\det(\hat{\Sigma}) - tr\Big((\Sigma-\hat{\Sigma})\Omega^*\Big) \right| = o_P(1).$$

By Lemma 3.1, the corresponding matrix Φ is Ω^*. The asymptotic variance of $\sqrt{n}(\phi(\Sigma) - \phi(\hat{\Sigma}))$ is

$$2\left\| \Sigma^{*1/2}\Phi\Sigma^{*1/2} \right\|_F^2 = 2p.$$

Corollary 3.5. *Consider the Wishart prior* $\Pi = \mathcal{W}_p(I, p+b-1)$ *in (2.5) with integer* $b = O(1)$*. Assume* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$ *and* $p^3/n = o(1)$*, then we have*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\sqrt{\frac{n}{2p}}\big(\log\det(\Sigma) - \log\det(\hat{\Sigma})\big) \le t \middle| X^n \right) - \mathbb{P}\Big(Z \le t\Big) \right| \to 0,$$

where $\hat{\Sigma}$ *is the sample covariance matrix.*

Proof. By Theorem 2.1 and Lemma 2.1, we only need to check the approximate linearity of the functional. According to the proof of Lemma 2.1, the choice of A_n such that $\Pi(A_n|X^n) = 1 - o_P(1)$ is

$$A_n = \left\{ ||\Sigma - \Sigma^*|| \leq M\sqrt{\frac{p}{n}} \right\},$$

for some $M > 0$. This implies $||\Sigma - \Sigma^*||_F \leq M\sqrt{\frac{p^2}{n}}$. Therefore,

$$A_n \subset \{\sqrt{n/p}||\Sigma - \Sigma^*||_F^2 \vee \sqrt{p}||\Sigma - \Sigma^*||_F \leq \delta_n\},$$

for some $\delta_n = o(1)$. By Lemma 3.1, we have

$$\sup_{A_n} \sqrt{\frac{n}{p}} \left| \log\det(\Sigma) - \log\det(\hat{\Sigma}) - \operatorname{tr}\Big((\Sigma - \hat{\Sigma})\Omega^*\Big) \right| = o_P(1),$$

and the approximate linearity holds. □

Corollary 3.6. *Consider the Gaussian prior* Π *in (2.6). Assume* $||\Sigma^*|| \vee ||\Omega^*|| \leq \Lambda = O(1)$ *and* $\frac{p^3(\log n)^2}{n} = o(1)$*, then we have*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\sqrt{\frac{n}{2p}} \big(\log\det(\Sigma) - \log\det(\hat{\Sigma}) \big) \leq t \middle| X^n \right) - \mathbb{P}\Big(Z \leq t\Big) \right| \to 0,$$

where $\hat{\Sigma}$ *is the sample covariance matrix.*

Proof. The proof of this corollary is the same as the proof of the last one using Wishart prior. The only difference is that the choice of A_n, according to the proof of Lemma 2.2, is

$$A_n = \left\{ ||\Sigma - \Sigma^*||_F \leq M\sqrt{\frac{p^2 \log n}{n}} \right\},$$

for some $M > 0$. Therefore,

$$A_n \subset \{\sqrt{n/p}||\Sigma - \Sigma^*||_F^2 \vee \sqrt{p}||\Sigma - \Sigma^*||_F \leq \delta_n\},$$

for some $\delta_n = o(1)$ under the assumption, and the approximate linearity holds. □

One immediate consequence of the result is the Bernstein-von Mises result for the entropy functional, defined as

$$H(\Sigma) = \frac{p}{2} + \frac{p\log(2\pi)}{2} + \frac{\log\det(\Sigma)}{2}.$$

Then it is direct that

$$\sqrt{\frac{2n}{p}} \Big(H(\Sigma) - H(\hat{\Sigma}) \Big) \Big| X^n \approx N(0,1).$$

3.4. Eigenvalues

In this section, we consider the eigenvalue functional. In particular, let $\{\lambda_m(\Sigma)\}_{m=1}^p$ be eigenvalues of the matrix Σ with decreasing order. We investigate the posterior distribution of $\lambda_m(\Sigma)$ for each $m = 1, \ldots, p$. Define the eigen-gap

$$\delta = \begin{cases} |\lambda_1(\Sigma^*) - \lambda_2(\Sigma^*)| & m = 1, \\ \min\{|\lambda_m(\Sigma^*) - \lambda_{m-1}(\Sigma^*)|, |\lambda_m(\Sigma^*) - \lambda_{m+1}(\Sigma^*)|\} & m = 2, 3, \ldots, p-1, \\ |\lambda_{m-1}(\Sigma^*) - \lambda_m(\Sigma^*)| & m = p. \end{cases}$$

The asymptotic order of δ plays an important role in the theory. The following lemma characterizes the approximate linearity of $\lambda_m(\Sigma)$.

Lemma 3.2. *Assume* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$ *and* $\frac{p}{\delta\sqrt{n}} = o(1)$, *then for any* $\delta_n = o(1)$, *we have*

$$\sup_{\left\{\substack{\delta^{-1}\sqrt{n}||\Sigma-\Sigma^*||^2 \\ \vee(\delta^{-1}+\sqrt{p})||\Sigma-\Sigma^*||\le\delta_n}\right\}} \frac{\sqrt{n}\left|\lambda_m(\Sigma)-\lambda_m(\hat{\Sigma})-tr\left((\Sigma-\hat{\Sigma})u_m^*u_m^{*T}\right)\right|}{|\lambda_m(\Sigma^*)|} = o_P(1),$$

where u_m^* *is the* m*-th eigenvector of* Σ^*.

Lemma 3.2 implies that the corresponding Φ in the linear expansion of $\phi(\Sigma)$ is $u_m^* u_m^{*T}$, and the asymptotic variance is

$$2\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^2 = 2|\lambda_m(\Sigma^*)|^2.$$

We also consider eigenvalues of the precision matrix. With slight abuse of notation, we define the eigengap of $\lambda_m(\Omega^*)$ to be

$$\delta = \begin{cases} |\lambda_1(\Omega^*) - \lambda_2(\Omega^*)| & m = 1, \\ \min\{|\lambda_m(\Omega^*) - \lambda_{m-1}(\Omega^*)|, |\lambda_m(\Omega^*) - \lambda_{m+1}(\Omega^*)|\} & m = 2, 3, \ldots, p-1, \\ |\lambda_{m-1}(\Omega^*) - \lambda_m(\Omega^*)| & m = p. \end{cases}$$

The approximate linearity of $\lambda_m(\Omega)$ is established in the following lemma.

Lemma 3.3. *Assume* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$, *then for any* $\delta_n = o(1)$, *we have*

$$\sup_{\left\{\substack{\delta^{-1}\sqrt{n}||\Sigma-\Sigma^*||^2 \\ \vee(\delta^{-1}+\sqrt{p})||\Sigma-\Sigma^*||\le\delta_n}\right\}} \frac{\sqrt{n}\left|\lambda_m(\Omega)-\lambda_m(\hat{\Sigma}^{-1})-tr\left((\Omega-\hat{\Sigma}^{-1})u_m^*u_m^{*T}\right)\right|}{|\lambda_m(\Omega^*)|} = o(1),$$

where u_m^* *is the* m*-th eigenvector of* Ω^*.

Similarly, Lemma 3.3 implies that the corresponding Ψ in the linear expansion of $\psi(\Omega)$ is $u_m^* u_m^{*T}$, and the asymptotic variance is

$$2\left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|_F^2 = 2|\lambda_m(\Omega^*)|^2.$$

Plugging the above lemmas into our general framework, we get the following corollaries.

Corollary 3.7. *Consider the Wishart prior* $\Pi = \mathcal{W}_p(I, p+b-1)$ *in (2.5) with integer* $b = O(1)$*. Assume* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$ *and* $\frac{p}{\delta\sqrt{n}} = o(1)$*, then we have*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\sqrt{\frac{n}{\sqrt{2}\lambda_m(\Sigma^*)}} \big(\lambda_m(\Sigma) - \lambda_m(\hat{\Sigma})\big) \le t | X^n \right) - \mathbb{P}\Big(Z \le t\Big) \right| \to 0,$$

where $\hat{\Sigma}$ *is the sample covariance matrix. If we instead assume* $\frac{p}{\delta\sqrt{n}} = o(1)$ *with* δ *being the eigengap of* $\lambda_m(\Omega^*)$*, then*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\Big(\sqrt{\frac{n}{\sqrt{2}\lambda_m(\Omega^*)}} \big(\lambda_m(\Omega) - \lambda_m(\hat{\Sigma}^{-1})\big) \le t | X^n \Big) - \mathbb{P}\Big(Z \le t\Big) \right| \to 0.$$

Proof. We only need to check the approximate linearity. According to Lemma 2.1, the choice of A_n is

$$A_n = \left\{ ||\Sigma - \Sigma^*|| \le M\sqrt{\frac{p}{n}} \right\},$$

for some $M > 0$. The assumption $\frac{p}{\delta\sqrt{n}} = o(1)$ implies

$$\delta^{-1}\sqrt{n}||\Sigma - \Sigma^*||^2 \vee (\delta^{-1} + \sqrt{p})||\Sigma - \Sigma^*|| = o(1),$$

on the set A_n. By Lemma 3.2 and Lemma 3.3, we have

$$\sup_{A_n} \sqrt{n}|\lambda_m(\Sigma^*)|^{-1} \left| \lambda_m(\Sigma) - \lambda_m(\hat{\Sigma}) - \mathrm{tr}\Big((\Sigma - \hat{\Sigma})u^*_m u^{*T}_m\Big) \right| = o_P(1),$$

and

$$\sup_{A_n} \sqrt{n}|\lambda_m(\Omega^*)|^{-1} \left| \lambda_m(\Omega) - \lambda_m(\hat{\Sigma}^{-1}) - \mathrm{tr}\Big((\Omega - \hat{\Sigma}^{-1})u^*_m u^{*T}_m\Big) \right| = o_P(1). \quad \square$$

Corollary 3.8. *Consider the Gaussian prior* Π *in (2.6). Assume* $||\Sigma^*|| \vee ||\Omega^*|| \le \Lambda = O(1)$ *and* $\frac{p^2 \log n}{\delta\sqrt{n}} = o(1)$*, then we have*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\left(\sqrt{\frac{n}{\sqrt{2}\lambda_m(\Sigma^*)}} \big(\lambda_m(\Sigma) - \lambda_m(\hat{\Sigma})\big) \le t | X^n \right) - \mathbb{P}\Big(Z \le t\Big) \right| \to 0,$$

where $\hat{\Sigma}$ *is the sample covariance matrix. If we instead assume* $\frac{p^2 \log n}{\delta\sqrt{n}} = o(1)$ *with* δ *being the eigengap of* $\lambda_m(\Omega^*)$*, then*

$$P^n_{\Sigma^*} \sup_{t\in\mathbb{R}} \left| \Pi\Big(\sqrt{\frac{n}{\sqrt{2}\lambda_m(\Omega^*)}} \big(\lambda_m(\Omega) - \lambda_m(\hat{\Sigma}^{-1})\big) \le t | X^n \Big) - \mathbb{P}\Big(Z \le t\Big) \right| \to 0.$$

Proof. We only need to check the approximate linearity. According to Lemma 2.2, the choice of A_n is

$$A_n = \left\{ ||\Sigma - \Sigma^*||_F \leq M\sqrt{\frac{p^2 \log n}{n}} \right\},$$

for some $M > 0$. The assumption $\frac{p^2 \log n}{\delta\sqrt{n}} = o(1)$ implies

$$\delta^{-1}\sqrt{n}||\Sigma - \Sigma^*||^2 \vee (\delta^{-1} + \sqrt{p})||\Sigma - \Sigma^*|| = o(1),$$

on the set A_n. By Lemma 3.2 and Lemma 3.3, we have

$$\sup_{A_n} \sqrt{n}|\lambda_m(\Sigma^*)|^{-1} \left| \lambda_m(\Sigma) - \lambda_m(\hat{\Sigma}) - \operatorname{tr}\Big((\Sigma - \hat{\Sigma}) u_m^* u_m^{*T} \Big) \right| = o_P(1),$$

and

$$\sup_{A_n} \sqrt{n}|\lambda_m(\Omega^*)|^{-1} \left| \lambda_m(\Omega) - \lambda_m(\hat{\Sigma}^{-1}) - \operatorname{tr}\Big((\Omega - \hat{\Sigma}^{-1}) u_m^* u_m^{*T} \Big) \right| = o_P(1). \quad \square$$

4. Discriminant analysis

In this section, we generalize the theory in Section 2 to handle the BvM theorem in discriminant analysis. Let $X^n = (X_1, \ldots, X_n)$ and $Y^n = (Y_1, \ldots, Y_n)$ be n i.i.d. training samples, where

$$X_i \sim N(\mu_X^*, \Omega_X^{*-1}), \quad Y_i \sim N(\mu_Y^*, \Omega_Y^{*-1}).$$

The discriminant analysis problem is to predict whether an independent new sample z is from the X-class or Y-class. For a given $(\mu_X, \mu_Y, \Omega_X, \Omega_Y)$, Fisher's QDA rule can be written as

$$\begin{aligned} \Delta(\mu_X, \mu_Y, \Omega_X, \Omega_Y) = &- (z - \mu_X)^T \Omega_X (z - \mu_X) \\ &+ (z - \mu_Y)^T \Omega_Y (z - \mu_Y) + \log \frac{\det(\Omega_X)}{\det(\Omega_Y)}. \end{aligned}$$

In this section, we are going to find the asymptotic posterior distribution

$$\sqrt{n} V^{-1} \Big(\Delta(\mu_X, \mu_Y, \Omega_X, \Omega_Y) - \hat{\Delta} \Big) \Big| X^n, Y^n, z,$$

with some appropriate variance V^2 and some prior distribution. Since the result is conditional on the new observation z, we treat it as a fixed (non-random) vector in this section without loss of generality.

Note that when $\Omega_X = \Omega_Y$ is assumed, the QDA rule can be reduced to the LDA rule. We give general results for Bernstein-von Mises theorem to hold in both cases respectively.

4.1. Linear discriminant analysis

Assume $\Omega_X^* = \Omega_Y^*$. For a given prior Π, the posterior distribution for LDA is defined as

$$\Pi(B|X^n, Y^n) = \frac{\int_B \exp\Big(l_n(\mu_X, \mu_Y, \Omega)\Big) d\Pi(\mu_X, \mu_Y, \Omega)}{\int \exp\Big(l_n(\mu_X, \mu_Y, \Omega)\Big) d\Pi(\mu_X, \mu_Y, \Omega)},$$

where $l_n(\mu_X, \mu_Y, \Omega)$ is the log-likelihood function decomposed as

$$l_n(\mu_X, \mu_Y, \Omega) = l_X(\mu_X, \Omega) + l_Y(\mu_Y, \Omega),$$

where

$$l_X(\mu_X, \Omega) = \frac{n}{2}\log\det(\Omega) - \frac{n}{2}\mathrm{tr}(\Omega\tilde{\Sigma}_X),$$

with $\tilde{\Sigma}_X = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_X)(X_i - \mu_X)^T$, and $l_Y(\mu_Y, \Omega)$ is defined in the similar way.

Consider the LDA functional

$$\Delta(\mu_X, \mu_Y, \Omega) = -(z - \mu_X)^T\Omega(z - \mu_X) + (z - \mu_Y)^T\Omega(z - \mu_Y).$$

Define the following quantities

$$\Phi = \frac{1}{2}\Omega^*\Big((z - \mu_X^*)(z - \mu_X^*)^T - (z - \mu_Y^*)(z - \mu_Y^*)^T\Big)\Omega^*,$$

$$\xi_X = 2(z - \mu_X^*), \quad \xi_Y = 2(\mu_Y^* - z),$$

$$\Omega_t = \Omega + \frac{2t}{\sqrt{n}}\Phi,$$

$$\mu_{X,t} = \mu_X + \frac{t}{\sqrt{n}}\xi_X, \quad \mu_{Y,t} = \mu_Y + \frac{t}{\sqrt{n}}\xi_Y,$$

$$\hat{\Sigma} = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T + \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})^T\right),$$

$$V^2 = 4\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^2 + \xi_X^T\Omega^*\xi_X + \xi_Y^T\Omega^*\xi_Y. \tag{4.1}$$

Assume A_n is a set satisfying

$$A_n \subset \Big\{\sqrt{n}\Big(||\mu_X - \mu_X^*||^2 + ||\mu_Y - \mu_Y^*||^2 + ||\Sigma - \Sigma^*||^2\Big)$$

$$\vee\sqrt{p}\Big(||\mu_X - \mu_X^*|| + ||\mu_Y - \mu_Y^*|| + ||\Sigma - \Sigma^*||\Big) \le \delta_n\Big\}, \quad \text{for some } \delta_n = o(1),$$

The main result for LDA is the following theorem.

Theorem 4.1. *Assume that* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$, $p^2/n = o(1)$, *and* $V^{-1} = O(1)$. *If for a given prior* Π, *the following two conditions are satisfied:*

1. $\Pi(A_n|X^n, Y^n) = 1 - o_P(1)$,
2. *For any fixed* $t \in \mathbb{R}$, $\frac{\int_{A_n} \exp(l_n(\mu_{X,t}, \mu_{Y,t}, \Omega_t))d\Pi(\mu_X, \mu_Y, \Omega)}{\int_{A_n} \exp(l_n(\mu_X, \mu_Y, \Omega))d\Pi(\mu_X, \mu_Y, \Omega)} = 1 + o_P(1)$,

then

$$\sup_{t\in\mathbb{R}} \left|\Pi\Big(\sqrt{n}V^{-1}\big(\Delta(\mu_X, \mu_Y, \Omega) - \hat{\Delta}\big) \le t|X^n, Y^n\Big) - \mathbb{P}\Big(Z \le t\Big)\right| = o_P(1),$$

where $Z \sim N(0,1)$ *and the centering is* $\hat{\Delta} = \Delta(\bar{X}, \bar{Y}, \hat{\Sigma}^{-1})$.

A curious condition in the above theorem is $V^{-1} = O(1)$. The following proposition shows it is implied by the separation of the two classes.

Proposition 4.1. *Under the setting of Theorem 4.1, if* $||\mu_X^* - \mu_Y^*|| \ge c$ *for some constant* $c > 0$, *then we have* $V^{-1} = O(1)$.

Proof. By the definition of V^2, we have

$$\begin{aligned} V^2 &\ge \xi_X^T\Omega^*\xi_X + \xi_Y^T\Omega^*\xi_Y \ge C\Big(||\xi_X||^2 + ||\xi_Y||^2\Big) \\ &= 4C\Big(||z - \mu_X^*||^2 + ||z - \mu_Y^*||^2\Big) \\ &\ge 2C||\mu_X^* - \mu_Y^*||^2, \end{aligned}$$

which is greater than a constant under the separation assumption. □

Now we give examples of priors for LDA. Let us use independent priors. That is

$$\Omega \sim \Pi_\Omega, \quad \mu_X \sim \Pi_X, \quad \mu_Y \sim \Pi_Y,$$

independently. The prior for the whole parameter (Ω, μ_X, μ_Y) is a product measure defined as

$$\Pi = \Pi_\Omega \times \Pi_X \times \Pi_Y.$$

Let Π_Ω be the Gaussian prior defined in (2.6). Let both Π_X and Π_Y be $N(0, I_{p\times p})$.

Theorem 4.2. *Assume* $||\Sigma^*|| \vee ||\Omega^*|| \le \Lambda = O(1)$, $V^{-1} = O(1)$, *and* $p^2 = o(\frac{\sqrt{n}}{\log n})$. *The prior defined above satisfies the two conditions in Theorem 4.1 for some appropriate* A_n. *Thus, the Bernstein-von Mises result holds.*

4.2. Quadratic discriminant analysis

For the general case that $\Omega_X^* = \Omega_Y^*$ may not be true, the posterior distribution for QDA is defined as

$$\Pi(B|X^n) = \frac{\int_B \exp\Big(l_n(\mu_X, \mu_Y, \Omega_X, \Omega_Y)\Big)d\Pi(\mu_X, \mu_Y, \Omega_X, \Omega_Y)}{\int \exp\Big(l_n(\mu_X, \mu_Y, \Omega_X, \Omega_Y)\Big)d\Pi(\mu_X, \mu_Y, \Omega_X, \Omega_Y)},$$

where $l_n(\mu_X, \mu_X, \Omega_X, \Omega_Y)$ has decomposition

$$l_n(\mu_X, \mu_X, \Omega_X, \Omega_Y) = l_X(\mu_X, \Omega_X) + l_Y(\mu_Y, \Omega_Y).$$

We define the following quantities,

$$\Phi_X = -\Omega_X^*\Big(\Sigma_X^* - (z-\mu_X^*)(z-\mu_X^*)^T\Big)\Omega_X^*,$$

$$\Phi_Y = \Omega_Y^*\Big(\Sigma_Y^* - (z-\mu_Y^*)(z-\mu_Y^*)^T\Big)\Omega_Y^*,$$

$$\xi_X = 2(z-\mu_X^*), \quad \xi_Y = 2(\mu_Y^* - z),$$

$$\Omega_{X,t} = \Omega_X + \frac{2t}{\sqrt{n}}\Phi_X, \quad \Omega_{Y,t} = \Omega_Y + \frac{2t}{\sqrt{n}}\Phi_Y,$$

$$\mu_{X,t} = \mu_X + \frac{t}{\sqrt{n}}\xi_X, \quad \mu_{Y,t} = \mu_Y + \frac{t}{\sqrt{n}}\xi_Y,$$

$$\hat{\Sigma}_X = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T, \quad \hat{\Sigma}_Y = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})^T,$$

$$V^2 = 2\left\|\Sigma_X^{*1/2}\Phi_X\Sigma_X^{*1/2}\right\|_F^2 + 2\left\|\Sigma_Y^{*1/2}\Phi_Y\Sigma_Y^{*1/2}\right\|_F^2 + \xi_X^T\Omega^*\xi_X + \xi_Y^T\Omega^*\xi_Y. \quad (4.2)$$

Assume A_n is a set satisfying

$$\begin{aligned} A_n \quad \subset \quad & \Big\{\sqrt{n}\Big(||\mu_X - \mu_X^*||^2 + ||\mu_Y - \mu_Y^*||^2 + ||\Sigma - \Sigma^*||^2\Big) \\ & \vee\sqrt{p}\Big(||\mu_X - \mu_X^*|| + ||\mu_Y - \mu_Y^*|| + ||\Sigma - \Sigma^*||\Big) \\ & \vee\sqrt{n/p}\Big(||\Sigma_X - \Sigma_X^*||_F^2 + ||\Sigma_Y - \Sigma_Y^*||_F^2\Big) \\ & \vee\sqrt{p}\Big(||\Sigma_X - \Sigma_X^*||_F + ||\Sigma_Y - \Sigma_Y||_F\Big) \le \delta_n\Big\}, \end{aligned}$$

with some $\delta_n = o(1)$. The main result for QDA is the following theorem.

Theorem 4.3. *Assume $||\Sigma^*|| \vee ||\Omega^*| = O(1)$, $V^{-1} = O(1)$, and $p^3/n = o(1)$. If for a given prior Π, the following two conditions are satisfied:*

1. $\Pi(A_n|X^n, Y^n) = 1 - o_P(1)$,
2. *For any fixed $t \in \mathbb{R}$, $\frac{\int_{A_n}\exp(l_n(\mu_{X,t},\mu_{Y,t},\Omega_{X,t},\Omega_{Y,t}))d\Pi(\mu_X,\mu_Y,\Omega_X,\Omega_Y)}{\int_{A_n}\exp(l_n(\mu_X,\mu_Y,\Omega_X,\Omega_Y))d\Pi(\mu_X,\mu_Y,\Omega_X,\Omega_Y)} = 1 + o_P(1)$,*

then

$$\sup_{t\in\mathbb{R}}\left|\Pi\Big(\sqrt{n}V^{-1}\big(\Delta(\mu_X,\mu_Y,\Omega_X,\Omega_Y) - \hat{\Delta}\big) \le t|X^n, Y^n\Big) - \mathbb{P}\Big(Z \le t\Big)\right| = o_P(1),$$

where $Z \sim N(0,1)$ and the centering is $\hat{\Delta} = \Delta(\bar{X}, \bar{Y}, \hat{\Sigma}_X^{-1}, \hat{\Sigma}_Y^{-1})$.

Remark 4.1. With the new definition of V in QDA, the assumption $V^{-1} = O(1)$ is also implied by the separation condition $||\mu_X - \mu_Y|| > c$ by applying the same argument in Proposition 4.1.

Remark 4.2. For independent prior in the sense that

$$d\Pi(\mu_X, \mu_Y, \Omega_X, \Omega_Y) = d\Pi(\mu_X, \Omega_X) \times d\Pi(\mu_Y, \Omega_Y),$$

the posterior is also independent because of the decomposition of the likelihood. In this case, we have

$$\Pi(A_n | X^n, Y^n) = \Pi_X(A_{X,n} | X^n) \times \Pi_Y(A_{Y,n} | Y^n),$$

with $A_{X,n}$ and $A_{Y,n}$ being versions of A_n involving only (μ_X, Ω_X) and (μ_Y, Ω_Y). In the same way, we also have

$$\begin{aligned} & \frac{\int_{A_n} \exp\Big(l_n(\mu_{X,t}, \mu_{Y,t}, \Omega_{X,t}, \Omega_{Y,t})\Big) d\Pi(\mu_X, \mu_Y, \Omega_X, \Omega_Y)}{\int_{A_n} \exp\Big(l_n(\mu_X, \mu_Y, \Omega_X, \Omega_Y)\Big) d\Pi(\mu_X, \mu_Y, \Omega_X, \Omega_Y)} \\ = & \frac{\int_{A_{X,n}} \exp\Big(l_n(\Omega_{X,t}, \mu_{X,t})\Big) d\Pi_X(\mu_X, \Omega_X)}{\int_{A_{X,n}} \exp\Big(l_n(\Omega_X, \mu_X)\Big) d\Pi_X(\mu_X, \Omega_X)} \\ & \times \frac{\int_{A_{Y,n}} \exp\Big(l_n(\Omega_{Y,t}, \mu_{Y,t})\Big) d\Pi_Y(\mu_Y, \Omega_Y)}{\int_{A_{Y,n}} \exp\Big(l_n(\Omega_Y, \mu_Y)\Big) d\Pi_Y(\mu_Y, \Omega_Y)}. \end{aligned}$$

Hence, for the two conditions in Theorem 4.3, it is sufficient to check

1. $\Pi(A_{X,n} | X^n) = 1 - o_P(1)$,
2. For any fixed $t \in \mathbb{R}$, $\frac{\int_{A_{X,n}} \exp(l_n(\Omega_{X,t}, \mu_{X,t})) d\Pi_X(\mu_X, \Omega_X)}{\int_{A_{X,n}} \exp(l_n(\Omega_X, \mu_X)) d\Pi_X(\mu_X, \Omega_X)} = 1 + o_P(1)$,

and the corresponding conditions for Y, when the prior has an independent structure.

The example of prior we specify for QDA is similar to the one for LDA. Let us use independent priors. That is

$$\Omega_X \sim \Pi_{\Omega_X}, \quad \Omega_Y \sim \Pi_{\Omega_Y}, \quad \mu_X \sim \Pi_X, \quad \mu_Y \sim \Pi_Y,$$

independently. The prior for the whole parameter $(\Omega_X, \Omega_Y, \mu_X, \mu_Y)$ is a product measure defined as

$$\Pi = \Pi_{\Omega_X} \times \Pi_{\Omega_Y} \times \Pi_X \times \Pi_Y.$$

Let Π_{Ω_X} and Π_{Ω_Y} be the Gaussian prior defined in Section 2.3.2. Let both Π_X and Π_Y be $N(0, I_{p \times p})$.

Theorem 4.4. *Assume* $||\Sigma_X^*|| \vee ||\Omega_X^*|| \vee ||\Sigma_Y^*|| \vee ||\Omega_Y^*|| \leq \Lambda = O(1)$, $V^{-1} = O(1)$ *and* $p^2 = o(\frac{\sqrt{n}}{\log n})$. *The prior defined above satisfies the two conditions in Theorem 4.1 for some appropriate* A_n. *Thus, the Bernstein-von Mises result holds.*

5. Discussion

5.1. Comparison: Asymptotic normality of $\phi(\hat{\Sigma})$ and $\psi(\hat{\Sigma}^{-1})$

In this section, we present the classical results for asymptotic normality of the estimators $\phi(\hat{\Sigma})$ and $\psi(\hat{\Sigma}^{-1})$. Note that in many cases, they coincide with MLE. The purpose is to compare them with the BvM results obtained in this paper. We first review and define some notation. Remember $\hat{\sigma}_{ij}$ is the (i,j)-th element of $\hat{\Sigma}$ and $\hat{\omega}_{ij}$ is the (i,j)-th element of $\hat{\Sigma}^{-1}$. We let Δ_L and Δ_Q be the LDA and QDA functionals respectively. The corresponding asymptotic variances are denoted by V_L^2 and V_Q^2, defined in (4.1) and (4.2) respectively. As $p, n \to \infty$ jointly, the asymptotic normality of $\phi(\hat{\Sigma})$ or $\psi(\hat{\Sigma}^{-1})$ holds under different asymptotic regimes for different functionals. For comparison, we assume that V_L, V_Q and the eigengap δ are at constant levels.

Theorem 5.1. *Let $p, n \to \infty$ jointly, then for any asymptotic regime of (p, n),*

$$\frac{\sqrt{n}(\hat{\sigma}_{ij} - \sigma^*_{ij})}{\sqrt{\sigma^{*2}_{ij} + \sigma^*_{ii}\sigma^*_{jj}}} \rightsquigarrow N(0,1),$$

$$\frac{\sqrt{n}(v^T\hat{\Sigma}v - v^T\Sigma^* v)}{\sqrt{2}|v^T\Sigma^* v|} \rightsquigarrow N(0,1).$$

Assume $p^2/n = o(1)$, we have

$$\frac{\sqrt{n}(\hat{\omega}_{ij} - \omega^*_{ij})}{\sqrt{\omega^{*2}_{ij} + \omega^*_{ii}\omega^*_{jj}}} \rightsquigarrow N(0,1), \tag{5.1}$$

$$\frac{\sqrt{n}(v^T\hat{\Sigma}^{-1}v - v^T\Omega^* v)}{\sqrt{2}|v^T\Omega^* v|} \rightsquigarrow N(0,1), \tag{5.2}$$

$$\frac{\sqrt{n}\big(\lambda_m(\hat{\Sigma}) - \lambda_m(\Sigma^*)\big)}{\sqrt{2}\lambda_m(\Sigma^*)} \rightsquigarrow N(0,1), \tag{5.3}$$

$$\frac{\sqrt{n}\big(\lambda_m(\hat{\Sigma}^{-1}) - \lambda_m(\Omega^*)\big)}{\sqrt{2}\lambda_m(\Omega^*)} \rightsquigarrow N(0,1), \tag{5.4}$$

$$\sqrt{n}V_L^{-1}\Big(\Delta_L(\bar{X}, \bar{Y}, \hat{\Sigma}^{-1}) - \Delta_L(\mu^*_X, \mu^*_Y, \Omega^*)\Big) \rightsquigarrow N(0,1). \tag{5.5}$$

Assume $p^3/n = o(1)$, we have

$$\sqrt{\frac{n}{2p}}\Big(\log\det(\hat{\Sigma}) - \log\det(\Sigma^*)\Big) \rightsquigarrow N(0,1), \tag{5.6}$$

$$\sqrt{n}V_Q^{-1}\Big(\Delta_Q(\bar{X}, \bar{Y}, \hat{\Sigma}_X^{-1}, \hat{\Sigma}_Y^{-1}) - \Delta_Q(\mu^*_X, \mu^*_Y, \Omega^*_X, \Omega^*_Y)\Big) \rightsquigarrow N(0,1). \tag{5.7}$$

Since the above results are more or less scattered in the literature, we do not present their proofs in this paper. Readers who are interested can derive these results using delta method.

We remark that the condition $p^2/n = o(1)$ is sharp for (5.1)-(5.5). For (5.1) and (5.2), a common example is $\omega_{11} = e_1^T \Omega e_1$, where $e_1^T = (1, 0, \ldots, 0)$. By distributional facts of inverse Wishart, $\sqrt{n}(\hat{\omega}_{11} - \omega_{11}^*)$ is not asymptotically normal if $p^2/n = o(1)$ does not hold. Since the functional Δ_L is harder than $v^T \Omega v$ (the latter is a special case of the former if μ_X^* and μ_Y^* are known), $p^2/n = o(1)$ is also sharp for (5.5). For (5.3) and (5.4), we have the following proposition to show that $p^2/n = o(1)$ is necessary.

Proposition 5.1. *Consider a diagonal Σ^*. Let the eigengap $\sigma_{11}^* - \sigma_{22}^*$ be at constant level when $p, n \to \infty$ jointly. Assume $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$, $n^{1/2} = o(p)$ and $p = o(n^{2/3})$. Then $\lambda_1(\hat{\Sigma})$ is not $\sqrt{n}$-consistent. As a consequence, $\lambda_p(\hat{\Sigma}^{-1}) = \lambda_1^{-1}(\hat{\Sigma})$ is not $\sqrt{n}$-consistent.*

The condition $p^3/n = o(1)$ is sharp for (5.6) and (5.7). If $p^3/n = o(1)$ does not hold, a bias correction is necessary for (5.6) to hold (see [6]). That the condition $p^3/n = o(1)$ is necessary for (5.7) is because the functional Δ_Q contains the part $\log\det(\Sigma)$.

In the next section, we are going to discuss the asymptotic regime of (p, n) for BvM and compare them with the frequentist results listed in this section.

5.2. The asymptotic regime of (p, n)

For all the BvM results we obtain in this paper, they assume different asymptotic regime of the sample size n and the dimension p. Ignoring the $\log n$ factor and assume constant eigengap δ and asymptotic variances for LDA and QDA, the asymptotic regime for (p, n) is summarized in the following table.

functional	$\phi(\hat{\Sigma})$ or $\psi(\hat{\Sigma}^{-1})$	conjugate	non-conjugate
σ_{ij}	***	$p \ll n$	$p^2 \ll n$
ω_{ij}	$p^2 \ll n$	$p^2 \ll n$	$p^3 \ll n$
$v^T \Sigma v$	***	$p \ll n$	$p^2 \ll n$
$v^T \Omega v$	$p^2 \ll n$	$p^2 \ll n$	$p^3 \ll n$
$\log\det(\Sigma)$	$p^3 \ll n$	$p^3 \ll n$	$p^3 \ll n$
$\lambda_m(\Sigma)$	$p^2 \ll n$	$p^2 \ll n$	$p^4 \ll n$
$\lambda_m(\Omega)$	$p^2 \ll n$	$p^2 \ll n$	$p^4 \ll n$
LDA	$p^2 \ll n$	$p^2 \ll$ n	$p^4 \ll n$
QDA	$p^3 \ll n$	$p^3 \ll$ n	$p^4 \ll n$

The table has three columns for the asymptotic normality of $\phi(\hat{\Sigma})$ and $\psi(\hat{\Sigma}^{-1})$ and for BvM with conjugate and non-conjugate priors respectively. The purpose is to compare our BvM result with the classical frequentist asymptotic normality. The priors are the Wishart prior and Gaussian prior we consider in this paper. For discriminant analysis, we did not consider conjugate prior because of limit of space. The conjugate prior in the LDA and QDA settings is the normal-Wishart

prior. Its posterior distribution can be decomposed as a marginal Wishart times a conditional normal. The analysis of the BvM result for this case is direct, and we claim the asymptotic regimes for LDA and QDA are $p^2 \ll n$ and $p^3 \ll n$ respectively without giving a formal proof.

Comparing the first and the second columns, the condition for p and n we need for the BvM results with conjugate prior matches the conditions for the frequentist results. The two exceptions are σ_{ij} and $v^T\Sigma v$, where for the frequentist asymptotic normality to hold, there is no assumption on p, n. Our technique of proof requires $p \ll n$. This is because our theory requires a set $A_n \subset \{||\Sigma - \Sigma^*|| \leq \delta_n\}$ for some $\delta_n = o(1)$ to satisfy $\Pi(A_n|X^n) = 1 - o_P(1)$. The best rate of convergence for $||\Sigma - \Sigma^*||$ is $\sqrt{p/n}$, which leads to $p \ll n$. Such assumption may be weaken if a different theory than ours can be developed (or through direct calculation by taking advantage of the conjugacy).

The comparison of the second and the third columns suggests that using of non-conjugate prior requires stronger assumptions. We believe these stronger assumptions can all be weakened. The current stronger assumptions on p and n are caused the technique we use in this paper to prove posterior contraction, which is Condition 1 in Theorem 2.1 and Theorem 2.2. The current way of proving posterior contraction in nonparametric Bayes theory only allows loss functions which are at the same order of the Kullback-Leibler divergence. In the covariance matrix estimation setting, we can only deal with Frobenius loss. We choose

$$A_n = \left\{ ||\Sigma - \Sigma^*||_F^2 \leq M \frac{p^2 \log n}{n} \right\}.$$

For functionals of covariance such as σ_{ij} and $v^T\Sigma v$, we need $A_n \subset \{||\Sigma - \Sigma^*|| \leq \delta_n\}$ for some δ_n. We have to bound $||\Sigma - \Sigma^*||$ as

$$||\Sigma - \Sigma^*|| \leq ||\Sigma - \Sigma^*||_F \leq \sqrt{M \frac{p^2 \log n}{n}},$$

and require $\sqrt{M \frac{p^2 \log n}{n}} \leq \delta_n = o(1)$. This leads to $p^2 \ll n$. For functionals of precision matrix, we need $A_n \subset \{\sqrt{p}||\Sigma - \Sigma^*|| \leq \delta_n\}$. Again, we have bound

$$\sqrt{p}||\Sigma - \Sigma^*|| \leq \sqrt{p}||\Sigma - \Sigma^*||_F \leq \sqrt{M \frac{p^3 \log n}{n}},$$

and require $\sqrt{M \frac{p^3 \log n}{n}} = o(1)$. This leads to $p^3 \ll n$. It would be great if we can prove a posterior contraction on $\{||\Sigma - \Sigma^*|| \leq M\sqrt{p/n}\}$ directly without referring to the Frobenius loss. However, under the current technique of Bayes nonparemtrics [21], this is impossible. See a lower bound argument in [22].

5.3. Sharpness of the condition $rp^2/n = o(1)$ in Theorem 2.2

It is curious whether the condition $rp^2/n = o(1)$ is sharp in Theorem 2.2. Let us consider the funcitonal $\psi(\Omega) = \log\det(\Omega)$. In this case, the corresponding

matrix Ψ in the linear expansion of $\psi(\Omega)$ is $\Psi = \Sigma^*$ and $r = \text{rank}(\Sigma^*) = p$. Then, the condition $rp^2 = o(1)$ becomes $p^3/n = o(1)$. Since $\log\det(\Omega) = -\log\det(\Sigma)$ and $p^3/n = o(1)$ is sharp for BvM to hold for $\log\det(\Sigma)$, it is also sharp for $\log\det(\Omega)$.

5.4. Covariance priors

The general framework in Section 2 only considers prior defined on precision matrix Ω. However, sometimes it is more natural to use prior defined on covariance matrix Σ, for example, Gaussian prior on Σ. Then, the first conditions in Theorem 2.1 and Theorem 2.2 are hard to check. We propose a slight variation of this condition, so that our theory can also be user-friendly for covariance priors.

We first consider approximate linear functionals of Σ satisfying (2.2). Then, the first condition of Theorem 2.1 can be replaced by

$$\frac{\int_{A_n} \exp\Big(l_n(\Sigma_t^{-1})\Big) d\Pi(\Sigma)}{\int_{A_n} \exp\Big(l_n(\Sigma^{-1})\Big) d\Pi(\Sigma)} = 1 + o_P(1), \quad \text{for each fixed } t \in \mathbb{R},$$

where $\Sigma_t = \Sigma - \frac{2t}{\sqrt{n}||\Sigma^{*1/2}\Phi\Sigma^{*1/2}||_F}\Sigma^*\Phi\Sigma^*$. Then we consider approximate linear functionals of Ω satisfying (2.4). The first condition of Theorem 2.2 can be replaced by

$$\frac{\int_{A_n} \exp\Big(l_n(\Sigma_t^{-1})\Big) d\Pi(\Sigma)}{\int_{A_n} \exp\Big(l_n(\Sigma^{-1})\Big) d\Pi(\Sigma)} = 1 + o_P(1), \quad \text{for each fixed } t \in \mathbb{R},$$

where $\Sigma_t = \Sigma + \frac{2t}{\sqrt{n}||\Omega^{*1/2}\Psi\Omega^{*1/2}||_F}\Psi$.

With the new conditions, it is direct to check them for covariance priors by change of variable, as is done in the proof of Lemma 2.1 and Lemma 2.2. In particular, for the Gaussian prior on covariance matrix, we claim the conclusion of Lemma 2.2 holds. We avoid expanding the technical details for the covariance priors in this paper due to the limit of space.

5.5. Relation to matrix estimation under Non-Frobenius loss

As we have mentioned in the end of Section 5.2, the current Bayes nonparametric technique for proving posterior contraction rate only covers losses which are at the same order of Kullback-Leiber divergence. It cannot handle other non-intrinsic loss [22]. In the Bayes matrix estimation setting, whether we can show the following conclusion

$$\Pi\Big(||\Sigma - \Sigma^*|| \le M\sqrt{\frac{p}{n}}\Big|X^n\Big) = 1 - o_P(1), \tag{5.8}$$

for a general non-conjugate prior still remains open. This explains why there is so little literature in this field compared to the growing research using frequentist methods. See, for example, [7] and [8].

However, we observe that for the spectral norm loss,

$$||\Sigma - \Sigma^*|| \leq 2 \sup_{v \in \mathcal{N}} |v^T(\Sigma - \Sigma^*)v|,$$

where $\mathcal{N}$ is a subset of S^{p-1} with cardinality bound $\log |\mathcal{N}| \leq cp$ for some $c > 0$. The BvM result we establish for the functional $v^T \Sigma v$ indicates that for each v, the posterior distribution of $|v^T(\Sigma - \Sigma^*)v|$ is at the order of $n^{-1/2}$. Therefore, heuristically, $2 \sup_{v \in \mathcal{N}} |v^T(\Sigma - \Sigma^*)v|$ should be at the order of $\frac{\sqrt{\log |\mathcal{N}|}}{\sqrt{n}}$, which is $\sqrt{p/n}$. We will use this intuition as a key idea in our future research project on the topic of Bayes matrix estimation.

Once (5.8) is established for a non-conjugate prior (e.g. Gaussian prior in this paper), then we may use (5.8) to weaken the conditions in the third column of the table in Section 5.2. In fact, most entries of that column can be weakened to match the conditions in the second column for a conjugate prior. As argued in Section 5.2, (5.8) directly implies the concentration $\Pi(A_n|X^n) = 1 - o_P(1)$, which is Condition 1 in both Theorem 2.1 and Theorem 2.2.

5.6. Matrices with structures

The paper mainly deals with covariance and precision matrices without structural assumptions. Extension of the results to more complex settings where the matrices may admit certain banded or sparsity structures is an interesting problem. In this section, we provide such an example by leveraging a recent posterior contraction result of Bayesian banded precision matrix estimation due to [2].

Define a class of positive semi-definite matrices with a k-banded structure by

$$\mathcal{F}_k = \left\{\Omega = \Omega^T = (\omega_{ij}) : \lambda_{\min}(\Omega) \geq 0, \omega_{ij} = \omega_{ij}\mathbb{I}\{|i-j| \leq k\}\right\}.$$

We are going to consider precision matrices in $\mathcal{F}_k$. A natural conjugate prior to model a banded precision matrix is the G-Wishart distribution, whose density function is given by

$$\frac{d\Pi(\Omega)}{d\Omega} \propto \exp\left(\frac{b-2}{2}\log\det(\Omega) - \frac{1}{2}\mathrm{tr}(\Omega)\right)\mathbb{I}\{\Omega \in \mathcal{F}_k\}.$$

Note that the G-Wishart distribution has the same density function as the Wishart distribution except that it is supported on the set $\mathcal{F}_k$. Banerjee and Ghosal [2] established the posterior contraction rate under the Gaussian covariance model.

Theorem 5.2 ([2])**.** *Consider the G-Wishart prior with $b > 2$ and $b = O(1)$. Assume $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$, $\Omega^* \in \mathcal{F}_k$ and $\frac{k^5 \log p}{n} = o(1)$. Then,*

$$\Pi\left(||\Omega - \Omega^*||_{\ell_1} \leq M\sqrt{\frac{k^5 \log p}{n}}\Big|X\right) = 1 - o_P(1),$$

with some large constant $M > 0$. The symbol $||\cdot||_{\ell_1}$ stands for the matrix ℓ_1 norm defined as the maximal row sum.

Now let us use this result to establish the BvM property of some (approximate) linear functionals of the covariance matrix. In view of Theorem 5.2, Condition 1 of Theorem 2.1 holds with

$$A_n = \left\{ ||\Omega - \Omega^*||_{\ell_1} \leq M\sqrt{\frac{k^5 \log p}{n}} \right\}. \tag{5.9}$$

Some direct calculation shows that Condition 2 also holds (details of the argument are given in the proof of Corollary 5.1). Hence, we have the following result.

Corollary 5.1. *Consider the G-Wishart prior with $b > 2$ and $b = O(1)$, and the functional $\phi(\Sigma)$ satisfying (2.2) with some symmetric matrix $\Phi = (\Phi_{ij}\mathbb{I}\{|i-j| \leq k\})$ of rank r and the set A_n defined by (5.9). Assume $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$, $\Omega^* \in \mathcal{F}_k$, $\frac{k^5 \log p}{n} = o(1)$ and $\frac{r}{n} = o(1)$. Then,*

$$\sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}\big(\phi(\Sigma) - \phi(\hat{\Sigma})\big)}{\sqrt{2}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F} \leq t \middle| X^n \right) - \mathbb{P}\big(Z \leq t\big) \right| = o_P(1),$$

where $Z \sim N(0,1)$.

Remark 5.1. The condition $\frac{k^5 \log p}{n} = o(1)$ only depends on the dimension p logarithmically due to the structural assumption on Ω^*. In [2], a wider class than $\mathcal{F}_k$ was considered. The result can be extended in a routine way, but we choose to stick to the class $\mathcal{F}_k$ for clarity of presentation.

Compared with the BvM results for matrices without structural assumption, Corollary 5.1 works for a more restrictive class of functionals. Namely, the symmetric matrix must also have a k-banded structure. For example, consider $\phi(\Sigma) = \sigma_{ij}$ for any (i,j) satisfying $|i-j| \leq k$, and then we have

$$\sup_{t\in\mathbb{R}} \left| \Pi\left(\frac{\sqrt{n}(\sigma_{ij} - \hat{\sigma}_{ij})}{\sqrt{\sigma^*_{ii}\sigma^*_{jj} + \sigma^{*2}_{ij}}} \leq t \middle| X^n \right) - \mathbb{P}\big(Z \leq t\big) \right| = o_P(1).$$

For $|i-j| > k$, the result may not hold. This is because Condition 2 of Theorem 2.1 requires a change of variable that,

$$\Omega_t = \Omega + \frac{\sqrt{2}t}{\sqrt{n}||\Sigma^{*1/2}\Phi\Sigma^{*1/2}||_F}\Phi$$

with respect to the prior distribution. Since the G-Wishart prior is only supported on the class of banded matrices, we need Φ to be banded so that Ω_t is in the support. For the same reason, we do not have a cleanly stated BvM result

for functionals of precision matrices, which, according to Theorem 2.2, requires a change of variable that

$$\Omega_t = \Omega - \frac{\sqrt{2}t}{\sqrt{n}||\Omega^{*1/2}\Psi\Omega^{*1/2}||_F}\Omega^*\Psi\Omega^*.$$

Therefore, $\Omega^*\Psi\Omega^*$ is required to be k-banded, which does not imply a clear restriction on Ψ.

To summarize, the extension of the BvM results to matrices with more structures still remain open. We hope Corollary 5.1 provides a stimulating attempt towards that direction.

6. Proofs

6.1. Proof of Theorem 2.1 & Theorem 2.2

Before stating the proofs, we first display some lemmas. The following lemma is Lemma 2 in [10]. It allows us to prove BvM results through convergence of moment generating functions.

Lemma 6.1. *Consider the random probability measure P_n and a fixed probability measure P. Suppose for any real t, the Laplace transformation $\int e^{tx}dP(x)$ is finite, and $\int e^{tx}dP_n(x) \to \int e^{tx}dP(x)$ in probability. Then, it holds that*

$$\sup_{t\in\mathbb{R}}\left|P_n\Big((-\infty,t]\Big) - P\Big((-\infty,t]\Big)\right| = o_P(1).$$

The next lemma is an expansion of the Gaussian likelihood.

Lemma 6.2. *Assume $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$. For any symmetric matrix Φ and the perturbed precision matrix*

$$\Omega_t = \Omega + \frac{\sqrt{2}t}{\sqrt{n}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F}\Phi,$$

the following equation holds for all $\Omega \in A_n$ with A_n satisfying (2.1) or (2.3).

$$\begin{aligned} l_n(\Omega_t) - l_n(\Omega) \quad = \quad & \frac{t\sqrt{n}}{\sqrt{2}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F} tr\Big((\Sigma - \hat{\Sigma})\Phi\Big) \\ & -\frac{1}{2}t^2\frac{\left\|\Sigma^{1/2}\Phi\Sigma^{1/2}\right\|_F^2}{\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^2} - \frac{n}{2}\sum_{j=1}^{p}\frac{(h_j-s)^2}{(1-s)^3}ds, \end{aligned} \tag{6.1}$$

where $\{h_j\}_{j=1}^p$ are eigenvalues of $\Sigma^{1/2}(\Omega - \Omega_t)\Sigma^{1/2}$.

The following lemma is Proposition D.1 in the supplementary material of [26], which is rooted in [16].

Lemma 6.3. *Let $Y_l \sim N(0, I_{p\times p})$. Then, for any $t > 0$,*

$$P_I^n\left(\left\|\frac{1}{n}\sum_{l=1}^{n} Y_l Y_l^T - I\right\| \le 2\left(\sqrt{\frac{p}{n}}+t\right) + \left(\sqrt{\frac{p}{n}}+t\right)^2\right) \ge 1 - 2e^{-nt^2/2}.$$

Proof of Theorem 2.1. We are going to use Lemma 6.1 and establish the convergence of moment generating function. We claim that

$$l_n(\Omega_t) - l_n(\Omega) = \frac{t\sqrt{n}}{\sqrt{2}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F}\left(\phi(\Sigma) - \phi(\hat{\Sigma})\right) - \frac{1}{2}t^2 + o_P(1), \quad (6.2)$$

uniformly over A_n. The derivation of (6.2) will be given at the end of the proof. Define the posterior distribution conditioning on A_n by

$$\Pi^{A_n}(B|X^n) = \frac{\Pi(A_n \cap B|X^n)}{\Pi(A_n|X^n)}, \quad \text{for any } B.$$

It is easy to see

$$\sup_B \left|\Pi^{A_n}(B|X^n) - \Pi(B|X^n)\right| = o_P(1), \quad (6.3)$$

by the first condition of Theorem 2.1. Now we calculation the moment generating function of $\frac{\sqrt{n}(\phi(\Sigma)-\phi(\hat{\Sigma}))}{\sqrt{2}\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\|_F}$ under the distribution $\Pi^{A_n}(\cdot|X^n)$, which is

$$\begin{aligned}
&\int \exp\left(\frac{t\sqrt{n}\left(\phi(\Sigma) - \phi(\hat{\Sigma})\right)}{\sqrt{2}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F}\right) d\Pi^{A_n}(\Omega|X^n) \\
= &\frac{\int_{A_n} \exp\left(\frac{t\sqrt{n}\left(\phi(\Sigma)-\phi(\hat{\Sigma})\right)}{\sqrt{2}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F} + l_n(\Omega)\right) d\Pi(\Omega)}{\int_{A_n} \exp\left(l_n(\Omega)\right) d\Pi(\Omega)} \\
= &\left(1 + o_P(1)\right)\exp\left(t^2/2\right)\frac{\int_{A_n} \exp\left(l_n(\Omega_t)\right) d\Pi(\Omega)}{\int_{A_n} \exp\left(l_n(\Omega)\right) d\Pi(\Omega)} \\
= &\left(1 + o_P(1)\right)\exp\left(t^2/2\right),
\end{aligned}$$

where the second equality is because of (6.2) and the last inequality is because of the second condition of Theorem 2.1. We have shown that the moment generating function of $\frac{\sqrt{n}(\phi(\Sigma)-\phi(\hat{\Sigma}))}{\sqrt{2}\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\|_F}$ under the distribution $\Pi^{A_n}(\cdot|X^n)$ converges to the moment generating function of $N(0,1)$ in probability. By Lemma 6.1 and (6.3), we have established the desired result.

To finish the proof, let us derive (6.2). Using the result of the likelihood expansion in Lemma 6.2, we will first show

$$l_n(\Omega_t) - l_n(\Omega) = \frac{t\sqrt{n}}{\sqrt{2}\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F}\mathrm{tr}\left((\Sigma - \hat{\Sigma})\Phi\right) - \frac{t^2}{2} + o(1), \quad (6.4)$$

where the $o(1)$ above is uniform on A_n. Compare (6.4) with (6.1) in Lemma 6.2, it is sufficient to bound

$$R_1 = \left| \frac{\left\|\Sigma^{1/2}\Phi\Sigma^{1/2}\right\|_F^2}{\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^2} - 1 \right| \quad \text{and} \quad R_2 = \left| \frac{n}{2}\sum_{j=1}^{p} \frac{(h_j - s)^2}{(1-s)^3} ds \right|.$$

We use the following argument to bound R_1 on A_n.

$$\begin{aligned}
& \left| ||\Sigma^{1/2}\Phi\Sigma^{1/2}||_F^2 - ||\Sigma^{1/2*}\Phi\Sigma^{*1/2}||_F^2 \right| \\
= \; & |\mathrm{tr}(\Sigma\Phi\Sigma\Phi) - \mathrm{tr}(\Sigma^*\Phi\Sigma^*\Phi)| \\
\le \; & \left| \mathrm{tr}\Big(\Sigma\Phi(\Sigma - \Sigma^*)\Phi\Big) \right| + \left| \mathrm{tr}\Big((\Sigma - \Sigma^*)\Phi\Sigma^*\Phi\Big) \right| \\
= \; & \left| \mathrm{tr}\Big(\Sigma^{1/2}\Phi\Sigma\Phi\Sigma^{1/2}(I - \Sigma^{-1/2}\Sigma^*\Sigma^{-1/2})\Big) \right| \\
& + \left| \mathrm{tr}\Big(\Sigma^{*1/2}\Phi\Sigma^*\Phi\Sigma^{*1/2}(\Sigma^{*-1/2}\Sigma\Sigma^{*-1/2} - I)\Big) \right| \\
\le \; & \mathrm{tr}\Big(\Sigma^{1/2}\Phi\Sigma\Phi\Sigma^{1/2}\Big) ||I - \Sigma^{-1/2}\Sigma^*\Sigma^{-1/2}|| \qquad (6.5) \\
& + \mathrm{tr}\Big(\Sigma^{*1/2}\Phi\Sigma^*\Phi\Sigma^{*1/2}\Big) ||I - \Sigma^{*-1/2}\Sigma\Sigma^{*-1/2}|| \\
= \; & ||\Sigma^{1/2}\Phi\Sigma^{1/2}||_F^2 ||I - \Sigma^{-1/2}\Sigma^*\Sigma^{-1/2}|| \\
& + ||\Sigma^{1/2*}\Phi\Sigma^{*1/2}||_F^2 ||I - \Sigma^{*-1/2}\Sigma\Sigma^{*-1/2}|| \\
\le \; & o(1)||\Sigma^{1/2}\Phi\Sigma^{1/2}||_F^2 + o(1)||\Sigma^{1/2*}\Phi\Sigma^{*1/2}||_F^2, \qquad (6.6)
\end{aligned}$$

where the inequality (6.5) is by von Neumann's trace inequality and the inequality (6.6) is due to the fact that $||\Sigma - \Sigma^*|| = o(1)$ on A_n. Rearranging the above argument, we get $R_1 = o(1)$ uniformly on A_n. To bound R_2, we first use Weyl's theorem to get

$$\max_{1\le j\le p} |h_j| \le \frac{\sqrt{2}t}{\sqrt{n}} \frac{\left\|\Sigma^{1/2}\Phi\Sigma^{1/2}\right\|}{\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F} = O(n^{-1/2}),$$

on A_n. Thus, on A_n, we have

$$\begin{aligned}
R_2 \; & \le \; Cn \left| \sum_{j=1}^{p} \int_0^{h_j} (h_j - s)^2 ds \right| \\
& \le \; Cn \sum_{j=1}^{p} |h_j|^3 \\
& \le \; Cn \max_{1\le j\le p} |h_j| \sum_{j=1}^{p} |h_j|^2 \\
& \le \; Cn \times O(n^{-1/2}) \times O\left(\frac{\left\|\Sigma^{1/2}\Phi\Sigma^{1/2}\right\|_F^2}{n\left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^2} \right) \\
& = \; O(n^{-1/2}).
\end{aligned}$$

Hence, (6.4) is proved. Together with the approximate linearity condition (2.2) of the functional $\phi(\Sigma)$, (6.2) is proved. Thus, the proof is complete. □

Proof of Theorem 2.2. We follow the reasoning in the proof of Theorem 2.1 and omit some similar steps. Define

$$\Phi = -\Omega^* \Psi \Omega^*.$$

It is easy to see that

$$\left\| \Omega^{*1/2} \Psi \Omega^{*1/2} \right\|_F = \left\| \Sigma^{*1/2} \Phi \Sigma^{*1/2} \right\|_F.$$

Then by Lemma 6.2 and the similar arguments in the proof of Theorem 2.1, we obtain

$$l_n(\Omega_t) - l_n(\Omega) = \frac{t\sqrt{n}\mathrm{tr}\Big((\Sigma - \hat{\Sigma})\Phi\Big)}{\sqrt{2}\left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|_F} - \frac{1}{2}t^2 + o(1),$$

uniformly on A_n, which is analogous to (6.4). We are going to approximate $\sqrt{n}\mathrm{tr}((\Sigma - \hat{\Sigma})\Phi)$ by $\sqrt{n}(\psi(\Omega) - \psi(\hat{\Sigma}^{-1}))$ on A_n. Define $\hat{\Omega} = \hat{\Sigma}^{-1}$. The assumption $rp^2/n = o(1)$ implies that $p/n = o(1)$. Thus, $\hat{\Omega}$ is well defined. By Lemma 6.3,

$$||\hat{\Sigma} - \Sigma^*|| = O\Big(\sqrt{\frac{p}{n}}\Big), \quad ||\hat{\Omega} - \Omega^*|| = O\Big(\sqrt{\frac{p}{n}}\Big). \tag{6.7}$$

Using notation $V = 2\left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|_F^2$, the approximation error on A_n is

$$\begin{aligned}
& \sqrt{n}V^{-1/2}\left|\psi(\Omega) - \psi(\hat{\Omega}) - \mathrm{tr}\Big((\Sigma - \hat{\Sigma})\Phi\Big)\right| \\
= \; & \sqrt{n}V^{-1/2}\left|\mathrm{tr}\Big((\Omega - \hat{\Omega})\Psi\Big) + \mathrm{tr}\Big((\Sigma - \hat{\Sigma})\Omega^*\Psi\Omega^*\Big)\right| \\
= \; & \sqrt{n}V^{-1/2}\left|\mathrm{tr}\Big((\Sigma - \hat{\Sigma})(\Omega^*\Psi\Omega^* - \Omega\Psi\hat{\Omega})\Big)\right| \\
\leq \; & \sqrt{n}V^{-1/2}\left|\mathrm{tr}\Big((\Sigma - \hat{\Sigma})\Omega^*\Psi(\Omega^* - \hat{\Omega})\Big)\right| \\
& + \sqrt{n}V^{-1/2}\left|\mathrm{tr}\Big((\Sigma - \hat{\Sigma})(\Omega^* - \Omega)\Psi\hat{\Omega}\Big)\right|.
\end{aligned}$$

Let the singular value decomposition of Ψ be $\Psi = \sum_{l=1}^r d_l q_l q_l^T$. Then,

$$\begin{aligned}
& \left|\mathrm{tr}\Big((\Sigma - \hat{\Sigma})\Omega^*\Psi(\Omega^* - \hat{\Omega})\Big)\right| \\
\leq \; & \sum_{l=1}^r |d_l|\left|\mathrm{tr}\Big((\Sigma - \hat{\Sigma})\Omega^* q_l q_l^T(\Omega^* - \hat{\Omega})\Big)\right| \\
\leq \; & \sum_{l=1}^r |d_l|||(\Sigma - \hat{\Sigma})\Omega^* q_l||||q_l^T(\Omega^* - \hat{\Omega})\Big)|| \\
\leq \; & O_P\left(||\hat{\Sigma} - \Sigma||||\hat{\Sigma} - \Sigma^*||\sum_{l=1}^r |d_l|\right).
\end{aligned}$$

Similarly,

$$\begin{aligned} & \left|\mathrm{tr}\Big((\Sigma-\hat{\Sigma})(\Omega^*-\Omega)\Psi\hat{\Omega}\Big)\right| \\ \leq \quad & O_P\left(||\hat{\Sigma}-\Sigma||||\hat{\Sigma}-\Sigma^*||\sum_{l=1}^{r}|d_l|\right). \end{aligned}$$

Since

$$V^{-1/2} \leq C||\Psi||_F^{-1} = \frac{C}{\sqrt{\sum_{l=1}^{r} d_l^2}},$$

we have

$$\begin{aligned} & \sqrt{n}V^{-1/2}\left|\psi(\Omega)-\psi(\hat{\Omega})-\mathrm{tr}\Big((\Sigma-\hat{\Sigma})\Phi\Big)\right| \\ \leq \quad & O_P\Big(\sqrt{nr}||\hat{\Sigma}-\Sigma||||\hat{\Sigma}-\Sigma^*||\Big) \\ \leq \quad & O_P\Big(\sqrt{nr}||\hat{\Sigma}-\Sigma^*||^2+\sqrt{nr}||\hat{\Sigma}-\Sigma^*||||\Sigma-\Sigma^*||\Big) \\ \leq \quad & O_P\left(\sqrt{\frac{rp^2}{n}}+\sqrt{rp}||\Sigma-\Sigma^*||\right) \\ = \quad & o_P(1) \end{aligned}$$

uniformly on A_n, where we have used (6.7) in the second last inequality above. Hence,

$$l_n(\Omega_t)-l_n(\Omega) = \frac{t\sqrt{n}\Big(\psi(\Omega)-\psi(\hat{\Sigma}^{-1})\Big)}{\sqrt{2}\left\|\Omega^{*1/2}\Psi\Omega^{*1/2}\right\|_F} - \frac{1}{2}t^2 + o_P(1), \tag{6.8}$$

uniformly on A_n. The remaining part of the proof are the same as the corresponding steps in the proof of Theorem 2.1. Thus, the proof is complete. □

6.2. Proof of Lemma 2.1

Proof of Lemma 2.1. The proof has two parts. In the first part, we establish the first condition of the two theorems by proving a posterior contraction rate. In the second part, we establish the second condition of the two theorems by showing that a change of variable is negligible under Wishart density.

Part I. The posterior distribution $\Omega|X^n$ is $\mathcal{W}_p((n\hat{\Sigma}+I)^{-1}, n+p+b-1)$. Conditioning on X^n, let $Z_l|X^n \sim P_{(n\hat{\Sigma}+I)^{-1}}$ i.i.d. for each $l = 1, 2, \ldots, n+p+b-1$. Then the posterior distribution of Ω is identical to the distribution of $\sum_{l=1}^{n+p+b-1} Z_l Z_l^T \big| X^n$. Define the set

$$G_n = \left\{||\Omega^{*1/2}\hat{\Sigma}\Omega^{*1/2}-I|| \leq C\sqrt{\frac{p}{n}}\right\},$$

and we have $P^n_{\Sigma^*}(G_n^c) \le \exp(-cp)$ by Lemma 6.3, for some $c, C > 0$. The event G_n implies $||\hat{\Sigma} - \Sigma^*|| \le C||\Sigma^*||\sqrt{\frac{p}{n}}$, by which we can deduce

$$\begin{aligned}\left\|\left(\hat{\Sigma} + \frac{1}{n}I\right)^{-1}\right\| &= \frac{1}{\lambda_{\min}\left(\hat{\Sigma} + \frac{1}{n}I\right)} \\ &\le \frac{1}{\lambda_{\min}(\Sigma^*) - \frac{1}{n} - ||\hat{\Sigma} - \Sigma^*||} \\ &\le 2||\Omega^*||.\end{aligned}$$

Using the obtained results, we can bound the deviation of the sample covariance by

$$\begin{aligned}&\left\|(n+p+b-1)(n\hat{\Sigma} + I)^{-1} - \Omega^*\right\| \\ &\le \left\|(n+p+b-1)(n\hat{\Sigma} + I)^{-1}\left(\frac{n}{n+b+p-1}(\hat{\Sigma} - \Sigma^*)\right.\right. \\ &\quad \left.\left. - \frac{b+p-1}{n+b+p-1}\Sigma^* + \frac{1}{n+p+b-1}I\right)\Omega^*\right\| \\ &\le \left\|\left(\hat{\Sigma} + \frac{1}{n}I\right)^{-1}\right\| ||(\hat{\Sigma} - \Sigma^*)\Omega^*|| + ||(p+b-1)(n\hat{\Sigma} + I)^{-1}|| + ||(n\hat{\Sigma} + I)^{-1}\Omega^*|| \\ &\le 2C||\Sigma^*||^{1/2}||\Omega^*||^{3/2}\sqrt{\frac{p}{n}} + \frac{2(p+b-1)}{n}||\Omega^*|| + \frac{2}{n}||\Omega^*||^2 \\ &\le C'||\Sigma^*||^{1/2}||\Omega^*||^{3/2}\sqrt{\frac{p}{n}},\end{aligned}$$

and the posterior deviation can be bounded by

$$\begin{aligned}&P^n_{\Sigma^*}\Pi\left(||\Omega - \Omega^*|| > 2C'||\Sigma^*||^{1/2}||\Omega^*||^{3/2}\sqrt{\frac{p}{n}}|X^n\right) \\ &\le P^n_{\Sigma^*}\Pi\left(||\Omega - \Omega^*|| > 2C'||\Sigma^*||^{1/2}||\Omega^*||^{3/2}\sqrt{\frac{p}{n}}|X^n\right)\mathbb{I}_{G_n} + P^n_{\Sigma^*}(G_n^c) \\ &\le P^n_{\Sigma^*}\Pi\left(\left\|\Omega - (n+p+b-1)(n\hat{\Sigma} + I)^{-1}\right\| > C'||\Sigma^*||^{1/2}||\Omega^*||^{3/2}\sqrt{\frac{p}{n}}|X^n\right)\mathbb{I}_{G_n} \\ &\quad + P^n_{\Sigma^*}(G_n^c) \\ &= P^n_{\Sigma^*}\mathbb{P}\left(\left\|\sum_{l=1}^{n+p+b-1} Z_l Z_l^T - (n+p+b-1)(n\hat{\Sigma} + I)^{-1}\right\| > \right. \\ &\quad \left. C'||\Sigma^*||^{1/2}||\Omega^*||^{3/2}\sqrt{\frac{p}{n}}\Big|X^n\right) + P^n_{\Sigma^*}(G_n^c)\end{aligned}$$

$$
\begin{aligned}
&\le P_{\Sigma^*}^n \mathbb{P}\Bigg(\Bigg\| (n+p+b-1)^{-1} \sum_{l=1}^{n+p+b-1} W_l W_l^T - I \Bigg\| > \\
&\quad \frac{1}{2} \frac{1}{||(\hat{\Sigma} + n^{-1} I)^{-1}||} C' ||\Sigma^*||^{1/2} ||\Omega^*||^{3/2} \sqrt{\frac{p}{n}} \Big| X^n \Bigg) + P_{\Sigma^*}^n (G_n^c) \\
&\le \mathbb{P}\Bigg(\Bigg\| (n+p+b-1)^{-1} \sum_{l=1}^{n+p+b-1} W_l W_l^T - I \Bigg\| > \frac{1}{4} C' ||\Sigma^*||^{1/2} ||\Omega^*||^{1/2} \sqrt{\frac{p}{n}} \Bigg) \\
&\quad + P_{\Sigma^*}^n (G_n^c) \\
&\le \exp\big(-c'p\big),
\end{aligned}
$$

where we use $W_l \sim N(0, I)$ in the above equations. In summary, we have proved

$$
P_{\Sigma^*}^n \Pi\Bigg(||\Omega - \Omega^*|| \le 2C' ||\Sigma^*||^{1/2} ||\Omega^*||^{3/2} \sqrt{\frac{p}{n}} \Big| X^n \Bigg) \to 1, \tag{6.9}
$$

which implies

$$
P_{\Sigma^*}^n \Pi\Big(||\Sigma - \Sigma^*|| \le M \sqrt{\frac{p}{n}} \Big| X^n \Big) \to 1,
$$

with some sufficiently large $M > 0$. We choose

$$
A_n = \Big\{ ||\Sigma - \Sigma^*|| \le M \sqrt{\frac{p}{n}} \Big\},
$$

so that $\Pi(A_n | X^n) = 1 - o_P(1)$ is true. For Theorem 2.1, let $\delta_n = M\sqrt{\frac{p}{n}}$. Then $\delta_n = o(1)$ by assumption, and $A_n \subset \{||\Sigma - \Sigma^*|| \le \delta_n\}$. For Theorem 2.2, let $\delta_n = M\sqrt{\frac{r^2 p}{n}}$, then we have $\delta_n = o(1)$ and $A_n \subset \{\sqrt{rp}||\Sigma - \Sigma^*|| \le \delta_n\}$.

Part II. Note that the proof for this part is the same for both Theorem 2.1 and Theorem 2.2 by letting $\Phi = -\Omega^* \Psi \Omega^*$. We introduce the notation

$$
\tilde{\Phi} = \Big(\sqrt{2} \Big\| \Sigma^{*1/2} \Phi \Sigma^{*1/2} \Big\|_F \Big)^{-1} \Phi.
$$

Now we study the integral $\int_{A_n} \exp(l_n(\Omega_t)) d\Pi(\Omega)$. Let $\mathcal{N}(p, b)$ be the normalizing constant of $\mathcal{W}_p(I, p+b-1)$. We have

$$
\begin{aligned}
&\int_{A_n} \exp\Big(l_n(\Omega_t) \Big) d\Pi(\Omega) \\
&= \mathcal{N}^{-1}(p,b) \int_{A_n} \exp\Big(l_n(\Omega + 2tn^{-1/2}\tilde{\Phi}) + \frac{b-2}{2} \log\det(\Omega) - \frac{1}{2}\mathrm{tr}(\Omega) \Big) d\Omega \\
&= \mathcal{N}^{-1}(p,b) \int_{A_n + 2tn^{-1/2}\tilde{\Phi}} \exp\Big(l_n(\Gamma) + \frac{b-2}{2} \log\det(\Gamma - 2tn^{-1/2}\tilde{\Phi}) \\
&\quad - \frac{1}{2}\mathrm{tr}(\Gamma - 2tn^{-1/2}\tilde{\Phi}) \Big) d\Gamma
\end{aligned}
$$

$$= \int_{A_n + 2tn^{-1/2}\tilde{\Phi}} \exp\Big(l_n(\Omega)\Big) \exp\Big(\frac{b-2}{2}\log\det(I - 2tn^{-1/2}\Omega^{-1}\tilde{\Phi}) + \frac{1}{2}\mathrm{tr}(2tn^{-1/2}\tilde{\Phi})\Big) d\Pi(\Omega).$$

The above integrals are meaningful because $A_n \cup (A_n + 2tn^{-1/2}\tilde{\Phi}) \subset \{\Omega : \Omega > 0, \Omega = \Omega^T\}$. Note that

$$A_n + 2tn^{-1/2}\tilde{\Phi} = \left\{ \left\| (\Omega - 2tn^{-1/2}\tilde{\Phi})^{-1} - \Sigma^* \right\| \le M\sqrt{\frac{p}{n}} \right\}.$$

Since $||2tn^{-1/2}\tilde{\Phi}|| = o(\sqrt{\frac{p}{n}})$, there exist M', M'' arbitrarily close to M such that $M' < M < M''$ and

$$A'_n \subset A_n + 2tn^{-1/2}\tilde{\Phi} \subset A''_n$$

for $A'_n = \{||\Sigma - \Sigma^*|| \le M'\sqrt{\frac{p}{n}}\}$ and $A''_n = \{||\Sigma - \Sigma^*|| \le M''\sqrt{\frac{p}{n}}\}$. The result (6.9) implies $\Pi(A'_n|X^n) = 1 - o_P(1)$ and $\Pi(A''_n|X^n) = 1 - o_P(1)$ are also true when M', M, M'' are large enough. Let $||\tilde{\Phi}||_N$ be the nuclear norm of $\tilde{\Phi}$, defined as the sum of its absolute eigenvalues. Note that on A''_n,

$$||\tilde{\Phi}||_N \le C\frac{||\Phi||_N}{||\Phi||_F} \le C\sqrt{p}.$$

Since

$$\begin{aligned} &\sup_{A''_n} \left| \frac{b-2}{2}\log\det(I - 2tn^{-1/2}\Omega^{-1}\tilde{\Phi}) + \frac{1}{2}\mathrm{tr}(2tn^{-1/2}\tilde{\Phi}) \right| \\ &\le tn^{-1/2} \sup_{A''_n} \left| |b-2| ||\Omega^{-1/2}\tilde{\Phi}\Omega^{-1/2}||_N + ||\tilde{\Phi}||_N \right| \\ &\le O(\sqrt{p/n}) \\ &= o(1), \end{aligned}$$

we have

$$\int_{A_n} \exp\Big(l_n(\Omega_t)\Big) d\Pi(\Omega) \le \big(1 + o(1)\big) \int_{A''_n} \exp\Big(l_n(\Omega)\Big) d\Pi(\Omega),$$

and

$$\int_{A_n} \exp\Big(l_n(\Omega_t)\Big) d\Pi(\Omega) \le \big(1 - o(1)\big) \int_{A'_n} \exp\Big(l_n(\Omega)\Big) d\Pi(\Omega).$$

The facts that $\Pi(A'_n|X^n) = 1 - o_P(1)$ and $\Pi(A''_n|X^n) = 1 - o_P(1)$ lead to

$$\frac{\int_{A_n} \exp\Big(l_n(\Omega_t)\Big) d\Pi(\Omega)}{\int_{A_n} \exp\Big(l_n(\Omega)\Big) d\Pi(\Omega)} = 1 + o_P(1). \qquad \square$$

6.3. Proof of Lemma 2.2

Now we are going to prove Lemma 2.2. Like the proof of Lemma 2.1, it has two parts. The first part is to show posterior contraction on some appropriate set A_n. Note that Wishart prior is a conjugate prior. The posterior contraction can be directly calculated. For the Gaussian prior, its non-conjugacy requires to apply some general result from nonparametric Bayes theory. To be specific, we follow the testing approach in [21]. The outline of using testing approach to prove posterior contraction for Bayesian matrix estimation is referred to Section 5 in [17].

We first state some lemmas.

Lemma 6.4. *Assume* $p^2 = o(n/\log n)$ *and* $||\Sigma^*|| \vee ||\Omega^*|| \leq \Lambda = O(1)$. *For the Gaussian prior* Π, *we have*

$$\Pi\Big(||\Omega||^2||\Sigma - \Sigma^*||_F^2 \leq \frac{p^2 \log n}{n}\Big) \geq \exp\Big(-Cp^2 \log n\Big),$$

for some constant $C > 0$.

The next lemma is Lemma 5.1 in [17].

Lemma 6.5. *Let* $K_n = \{||\Omega||^2||\Sigma - \Sigma^*||_F^2 \leq \frac{p^2 \log n}{n}\}$. *Then for any* $b > 0$, *we have*

$$\begin{aligned} & P_{\Sigma^*}^n\left(\int \exp\Big(l_n(\Omega) - l_n(\Omega^*)\Big) d\Pi(\Omega) \leq \Pi(K_n) \exp\big(-(b+1)p^2 \log n\big)\right) \\ \leq \quad & \exp\Big(-Cb^2p^2 \log n\Big), \end{aligned}$$

for some constant $C > 0$.

The next lemma is Lemma 5.9 in [17].

Lemma 6.6. *For* $||\Sigma^*|| \vee ||\Omega^*|| \leq \Lambda = O(1)$ *and* $||\Sigma_1|| \vee ||\Omega_1|| \leq 2\Lambda$, *there exist small* $\delta, \delta' > 0$ *only depending on* Λ, *and a testing function* ϕ *such that*

$$P_{\Sigma^*}^n \phi \leq 2\exp\Big(-C\delta'||\Sigma^* - \Sigma_1||_F^2\Big),$$

$$\sup_{\{\Sigma \in supp(\Pi): ||\Sigma - \Sigma_1||_F \leq \delta ||\Sigma^* - \Sigma_1||_F\}} P_\Sigma^n(1-\phi) \leq 2\exp\Big(-C\delta'||\Sigma^* - \Sigma_1||_F^2\Big),$$

for some constant $C > 0$.

Proof of Lemma 2.2. Like what we have done in the Wishart case, the proof has two parts. In the first part, we establish the first condition of the two theorems by proving a posterior contraction rate. In the second part, we establish the second condition of the two theorems by showing that a change of variable is negligible under Gaussian density.

Part I. Define

$$A_n = \left\{ ||\Sigma - \Sigma^*||_F \leq M\sqrt{\frac{p^2 \log n}{n}} \right\},$$

for some M sufficiently large. Then, we may write

$$\Pi(A_n^c|X^n) = \frac{\int_{A_n^c} \exp\Big(l_n(\Omega) - l_n(\Omega^*)\Big) d\Pi(\Omega)}{\int \exp\Big(l_n(\Omega) - l_n(\Omega^*)\Big) d\Pi(\Omega)} = \frac{N_n}{D_n}.$$

Let us establish a testing between the following hypotheses:

$$H_0 : \Sigma = \Sigma^* \quad \text{vs} \quad H_1 : \Sigma \in A_n^c \cap \text{supp}(\Pi).$$

There exists $\{\Sigma_j\}_{j=1}^N \subset A_n^c \cap \text{supp}(\Pi)$, such that

$$A_n^c \cap \text{supp}(\Pi) \subset \text{supp}(\Pi) \cap \left(\cup_{j=1}^N \left\{ ||\Sigma - \Sigma_j||_F \leq \sqrt{\frac{p^2 \log n}{n}} \right\} \right).$$

We choose the smallest N, which is determined by the covering number. Since $A_n^c \cap \text{supp}(\Pi) \subset \{||\Omega||_F \leq 2\Lambda\sqrt{p}\}$, we have

$$\log N \leq C' p^2 \log \left(\frac{2\Lambda\sqrt{n}}{\sqrt{p \log n}} \right) \leq C p^2 \log n.$$

By Lemma 6.6, there exists ϕ_j such that

$$P_{\Sigma^*}^n \phi_j \leq 2 \exp\Big(- CM^2 p^2 \log n \Big),$$

$$\sup_{\{\Sigma \in \text{supp}(\Pi) : ||\Sigma - \Sigma_j||_F \leq \sqrt{p^2 \log n / n}\}} P_\Sigma^n (1 - \phi_j) \leq 2 \exp\Big(- CM^2 p^2 \log n \Big).$$

Define $\phi = \max_{1 \leq j \leq N} \phi_j$. Using union bound to control the testing error, we have

$$P_{\Sigma^*}^n \phi \leq \exp\Big(- C_1 M^2 p^2 \log n \Big),$$

$$\sup_{\{\Sigma \in A_n^c \cap \text{supp}(\Pi)\}} P_\Sigma^n (1 - \phi) \leq \exp\Big(- C_1 M^2 p^2 \log n \Big),$$

for sufficiently large M. We bound $\Pi(A_n^c|X^n)$ by

$$\begin{aligned} P_{\Sigma^*}^n \Pi(A_n^c|X^n) &\leq P_{\Sigma^*}^n \Pi(A_n^c|X^n)(1-\phi) \mathbb{I}\left\{ D_n > \exp(-2p^2 \log n) \right\} \\ &\quad + P_{\Sigma^*}^n \phi + P_{\Sigma^*}^n \Big(D_n \leq \exp(-2p^2 \log n) \Big) \\ &\leq \exp\big(2p^2 \log n\big) P_{\Sigma^*}^n \int_{A_n^c} \exp\Big(l_n(\Omega) - l_n(\Omega^*)\Big)(1-\phi) d\Pi(\Omega) \\ &\quad + P_{\Sigma^*}^n \phi + P_{\Sigma^*}^n \Big(D_n \leq \exp(-2p^2 \log n) \Big) \end{aligned}$$

$$\begin{aligned}
&\leq \exp\left(2p^2\log n\right)\int_{A_n^c} P_\Sigma^n(1-\phi)d\Pi(\Omega)\\
&\quad + P_{\Sigma^*}^n\phi + P_{\Sigma^*}^n\Big(D_n \leq \exp(-2p^2\log n)\Big)\\
&\leq \exp\left(2p^2\log n\right)\sup_{\Sigma\in A_n^c\cap \mathrm{supp}(\Pi)} P_\Sigma^n(1-\phi)\\
&\quad + P_{\Sigma^*}^n\phi + P_{\Sigma^*}^n\Big(D_n \leq \exp(-2p^2\log n)\Big).
\end{aligned}$$

In the upper bound above, the first two terms are bounded by the testing error we have established. The last term can be bounded by combining the results of Lemma 6.4 and Lemma 6.5. Hence, we have proved that

$$\Pi(A_n^c|X^n) = 1 - o_P(1).$$

For Theorem 2.1, let $\delta_n = M\sqrt{\frac{p^2\log n}{n}}$. Then $\delta_n = o(1)$ by assumption, and $A_n \subset \{||\Sigma - \Sigma^*|| \leq \delta_n\}$. For Theorem 2.2, let $\delta_n = M\sqrt{\frac{rp^3\log n}{n}}$, then we have $\delta_n = o(1)$ and $A_n \subset \{\sqrt{rp}||\Sigma - \Sigma^*|| \leq \delta_n\}$.

Part II. Let Π_G induce a prior distribution on symmetric Ω with each of the upper triangular element independently following $N(0,1)$. The density of Π_G is

$$\frac{d\Pi_G(\Omega)}{d\Omega} = \xi_p^{-1}\exp\Big(-\frac{1}{2}||\bar{\Omega}||_F^2\Big),$$

where we use $\bar{\Omega}$ to zero out the lower triangular elements of Ω except the diagonal part and ξ_p is the normalizing constant. Write

$$\begin{aligned}
&\int_{A_n}\exp\Big(l_n(\Omega_t)\Big)d\Pi(\Omega)\\
=\ &\xi_p^{-1}\int_{A_n}\exp\Big(l_n(\Omega_t) - \frac{1}{2}||\bar{\Omega}||_F^2\Big)d\bar{\Omega}.
\end{aligned}$$

Remembering the notation $\tilde{\Phi}$ defined in the proof of Lemma 2.1, we have

$$\begin{aligned}
&\int_{A_n}\exp\Big(l_n(\Omega_t) - \frac{1}{2}||\bar{\Omega}||_F^2\Big)d\bar{\Omega}\\
=\ &\int_{A_n+2tn^{-1/2}\tilde{\Phi}}\exp\Big(l_n(\Gamma) - \frac{1}{2}||\bar{\Gamma} - 2tn^{-1/2}\bar{\tilde{\Phi}}||_F^2\Big)d\bar{\Gamma}\\
=\ &\int_{A_n+2tn^{-1/2}\tilde{\Phi}}\exp\Big(l_n(\Gamma) - \frac{1}{2}||\bar{\Gamma}||_F^2 + 2tn^{-1/2}\mathrm{tr}\big(\bar{\Gamma}\bar{\tilde{\Phi}}\big) - 2t^2n^{-1}||\bar{\tilde{\Phi}}||_F^2\Big)d\bar{\Gamma}.
\end{aligned}$$

We may choose M', M'' arbitrarily close to M such that $M' < M < M''$ and $A_n' \subset A_n + 2tn^{-1/2}\tilde{\Phi} \subset A_n''$ for

$$A_n' = \left\{||\Sigma - \Sigma^*||_F \leq M'\sqrt{\frac{p^2\log n}{n}}\right\},\quad A_n'' = \left\{||\Sigma - \Sigma^*||_F \leq M''\sqrt{\frac{p^2\log n}{n}}\right\}.$$

This can always be done because $||2tn^{-1/2}\tilde{\Phi}||_F = O(n^{-1/2}) = o(\sqrt{\frac{p^2 \log n}{n}})$. Moreover, we have $\Pi(A_n'|X^n) = 1 - o_P(1)$, $\Pi(A_n''|X^n) = 1 - o_P(1)$ and

$$\begin{aligned} & \sup_{A_n''} \left| 2tn^{-1/2} \mathrm{tr}\big(\bar{\Gamma}\bar{\tilde{\Phi}}\big) - 2t^2 n^{-1} ||\bar{\tilde{\Phi}}||_F^2 \right| \\ \leq \; & C \sup_{A_n''} \left| n^{-1/2} ||\Gamma||_F ||\tilde{\Phi}||_F + n^{-1} ||\tilde{\Phi}||_F^2 \right| \\ = \; & O\left(\sqrt{\frac{p}{n}} + \frac{1}{n} \right) \\ = \; & o(1). \end{aligned}$$

Therefore, using the same argument in the proof of Lemma 2.1, we have

$$\frac{\int_{A_n} \exp\Big(l_n(\Omega_t)\Big) d\Pi(\Omega)}{\int_{A_n} \exp\Big(l_n(\Omega)\Big) d\Pi(\Omega)} = 1 + o_P(1).$$

This completes the proof. □

6.4. *Proof of technical Lemmas*

Proof of Lemma 6.2. First, we show Ω_t is a valid precision matrix under the event A_n, i.e., $\Omega_t > 0$. Using Weyl's theorem, we have

$$|\lambda_{\min}(\Omega_t) - \lambda_{\min}(\Omega^*)| \leq ||\Omega_t - \Omega|| + ||\Omega - \Omega^*||,$$

where the first term is bounded by

$$||\Omega_t - \Omega|| \leq \frac{\sqrt{2}t}{\sqrt{n}} \frac{||\Phi||}{\big|\big|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\big|\big|_F} = O(n^{-1/2}).$$

Hence,

$$|\lambda_{\min}(\Omega_t) - \lambda_{\min}(\Omega^*)| \leq O(n^{-1/2}) + ||\Omega - \Omega^*||.$$

Under the current assumption, $O(n^{-1/2}) + ||\Omega - \Omega^*|| = o(\lambda_{\min}(\Omega^*))$. Hence, $\lambda_{\min}(\Omega_t) > 0$. Knowing the fact that $l_n(\Omega_t)$ is well-defined, we study $l_n(\Omega_t) - l_n(\Omega)$,

$$\begin{aligned} l_n(\Omega_t) - l_n(\Omega) &= \frac{n}{2}\mathrm{tr}\Big(\hat{\Sigma}(\Omega - \Omega_t)\Big) + \frac{n}{2}\log\det\Big(I - (\Omega - \Omega_t)\Sigma\Big) \\ &= \frac{n}{2}\mathrm{tr}\Big((\hat{\Sigma} - \Sigma)(\Omega - \Omega_t)\Big) + \frac{n}{2}\mathrm{tr}\Big(\Sigma^{1/2}(\Omega - \Omega_t)\Sigma^{1/2}\Big) \\ &\quad + \frac{n}{2}\log\det\Big(I - \Sigma^{1/2}(\Omega - \Omega_t)\Sigma^{1/2}\Big). \end{aligned}$$

Let $\{h_j\}_{j=1}^p$ be eigenvalues of $\Sigma^{1/2}(\Omega - \Omega_t)\Sigma^{1/2}$. Then, we have

$$\frac{n}{2}\mathrm{tr}\Big(\Sigma^{1/2}(\Omega - \Omega_t)\Sigma^{1/2}\Big) + \frac{n}{2}\log\det\Big(I - \Sigma^{1/2}(\Omega - \Omega_t)\Sigma^{1/2}\Big)$$

$$= \frac{n}{2}\sum_{j=1}^{p}\Big(h_j + \log(1-h_j)\Big)$$
$$= -\frac{n}{4}\sum_{j=1}^{p} h_j^2 - \frac{n}{2}\sum_{j=1}^{p}\int_0^{h_j}\frac{(h_j-s)^2}{(1-s)^3}ds$$
$$= -\frac{n}{4}||\Sigma^{1/2}(\Omega-\Omega_t)\Sigma^{1/2}||_F^2 - \frac{n}{2}\sum_{j=1}^{p}\int_0^{h_j}\frac{(h_j-s)^2}{(1-s)^3}ds,$$

where $\int_0^{h_j}\frac{(h_j-s)^2}{(1-s)^3}ds$ is the remainder of the Taylor expansion. Therefore, we have obtained the expansion

$$\begin{aligned} & l_n(\Omega_t) - l_n(\Omega) \\ &= \frac{n}{2}\mathrm{tr}\Big((\hat{\Sigma}-\Sigma)(\Omega-\Omega_t)\Big) - \frac{n}{4}||\Sigma^{1/2}(\Omega-\Omega_t)\Sigma^{1/2}||_F^2 - \frac{n}{2}\sum_{j=1}^{p}\int_0^{h_j}\frac{(h_j-s)^2}{(1-s)^3}ds \\ &= \frac{t\sqrt{n}}{\sqrt{2}||\Sigma^{*1/2}\Phi\Sigma^{*1/2}||_F}\mathrm{tr}\Big((\Sigma-\hat{\Sigma})\Phi\Big) - \frac{t^2}{2}\frac{||\Sigma^{1/2}\Phi\Sigma^{1/2}||_F^2}{||\Sigma^{*1/2}\Phi\Sigma^{*1/2}||_F^2} \\ &\quad - \frac{n}{2}\sum_{j=1}^{p}\int_0^{h_j}\frac{(h_j-s)^2}{(1-s)^3}ds. \end{aligned}$$

The proof is complete. □

Proof of Lemma 6.4. Define Π_G to be the distribution which specifies i.i.d. $N(0,1)$ on the upper triangular part of Ω and then take the lower triangular part to satisfy $\Omega^T = \Omega$. Define

$$D = \{||\Omega|| \vee ||\Sigma|| \le 2\Lambda\}.$$

Then according to the definition of Π, we have

$$\Pi(B) = \frac{\Pi_G(B\cap D)}{\Pi_G(D)}, \quad \text{for any } B.$$

Since $\Pi_G(D) \le 1$, we have

$$\Pi(B) \ge \Pi_G(B\cap D), \quad \text{for any } B.$$

In particular, we have

$$\begin{aligned} &\Pi\Big(||\Omega||^2||\Sigma-\Sigma^*||_F^2 \le p^2\log n/n\Big) \\ &\ge \Pi_G\Big(||\Omega||^2||\Sigma-\Sigma^*||_F^2 \le p^2\log n/n, ||\Omega||\vee||\Sigma|| \le 2\Lambda\Big). \end{aligned}$$

Since $p^2/n = o(1)$, we have

$$\left\{||\Omega-\Omega^*||_F \le \frac{p\sqrt{\log n}}{(2\Lambda)^3\sqrt{n}}\right\} \subset \left\{||\Omega||^2||\Sigma-\Sigma^*||_F^2 \le \frac{p^2\log n}{n}, ||\Omega||\vee||\Sigma|| \le 2\Lambda\right\}.$$

Thus,

$$\Pi\Big(||\Omega||^2||\Sigma-\Sigma^*||_F^2 \le p^2\log n/n\Big) \ge \Pi_G\Big(||\Omega-\Omega^*||_F \le \frac{p\sqrt{\log n}}{(2\Lambda)^3\sqrt{n}}\Big).$$

Calculate using Gaussian density directly, for example, according to Lemma E.1 in [17], and we have

$$\begin{aligned}
&\Pi_G\Big(||\Omega-\Omega^*||_F \le \frac{p\sqrt{\log n}}{(2\Lambda)^3\sqrt{n}}\Big)\\
\ge\ & e^{-||\Omega^*||_F^2}\left(\mathbb{P}\Big(|Z|^2 \le \frac{\log n}{cn}\Big)\right)^{p(p+1)/2}\\
\ge\ & \exp\Big(-||\Omega^*||_F^2 - Cp^2\log n\Big),
\end{aligned}$$

where $Z\sim N(0,1)$. The proof is complete by observing that $||\Omega^*||_F^2 = o(p^2\log n)$ under the assumption. □

Appendix A: Proof of Theorem 4.1 & Theorem 4.3

Lemma A.1. *Under the setting of Theorem 4.1, assume* $p^2/n = o(1)$ *and* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$*, then we have*

$$\begin{aligned}
&l_n(\mu_{X,t},\mu_{Y,t},\Omega_t) - l_n(\mu_X,\mu_Y,\Omega)\\
=\ & \frac{2t\sqrt{n}}{V}tr\Big((\Sigma-\hat\Sigma)\Phi\Big) - \frac{t\sqrt{n}}{V}(\bar X-\mu_X)^T\Omega^*\xi_X\\
& -\frac{t\sqrt{n}}{V}(\bar Y-\mu_Y)^T\Omega^*\xi_Y - \frac{1}{2}t^2 + o_P(1),
\end{aligned}$$

uniformly on A_n.

Proof. Since

$$\begin{aligned}
&l_n(\mu_{X,t},\mu_{Y,t},\Omega_t) - l_n(\mu_X,\mu_Y,\Omega)\\
=\ & \Big(l_X(\mu_{X,t},\Omega_t) - l_X(\mu_X,\Omega)\Big) + \Big(l_Y(\mu_{Y,t},\Omega_t) - l_Y(\mu_Y,\Omega)\Big),
\end{aligned}$$

we expand both quantities in the brackets using the general notation $l(\mu_t,\Omega_t) - l(\mu,\Omega)$. Using Taylor expansion as in the proof of Lemma 6.2 and the notation $\tilde\Sigma = \frac{1}{n}\sum_{i=1}^n(X_i-\mu)(X_i-\mu)^T$, we have

$$\begin{aligned}
&l(\mu_t,\Omega_t) - l(\mu,\Omega)\\
=\ & \frac{n}{2}\mathrm{tr}\Big((\tilde\Sigma-\Sigma)(\Omega-\Omega_t)\Big) - n(\mu-\mu_t)^T\Omega_t(\bar X-\mu) - \frac{n}{2}\Big\|\Sigma^{1/2}(\Omega-\Omega_t)\Sigma^{1/2}\Big\|_F^2\\
& -\frac{n}{2}(\mu-\mu_t)^T\Omega_t(\mu-\mu_t) - \frac{n}{2}\sum_{j=1}^p\int_0^{h_j}\frac{(h_j-s)^2}{(1-s)^3}ds
\end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{t\sqrt{n}}{V}\mathrm{tr}\Big((\Sigma-\tilde{\Sigma})\Phi\Big)+\frac{t\sqrt{n}}{V}\xi^T\Omega_t(\bar{X}-\mu) \\
& -\frac{t^2}{V^2}\left\|\Sigma^{1/2}\Phi\Sigma^{1/2}\right\|_F^2-\frac{t^2}{2V^2}\xi^T\Omega_t\xi-\frac{n}{2}\sum_{j=1}^{p}\int_0^{h_j}\frac{(h_j-s)^2}{(1-s)^3}ds,
\end{aligned}
$$

where $\{h_j\}_{j=1}^p$ are eigenvalues of $\Sigma^{1/2}(\Omega-\Omega_t)\Sigma^{1/2}$. The same proof in Lemma 6.2 implies

$$
\left|\frac{n}{2}\sum_{j=1}^{p}\int_0^{h_j}\frac{(h_j-s)^2}{(1-s)^3}ds\right|=o(1),\quad \text{on } A_n.
$$

Therefore,

$$
\begin{aligned}
& l_n(\mu_{X,t},\mu_{Y,t},\Omega_t)-l_n(\mu_X,\mu_Y,\Omega) \\
= \quad & \frac{2t\sqrt{n}}{V}\mathrm{tr}\left(\left(\Sigma-\frac{1}{2}\big(\tilde{\Sigma}_X+\tilde{\Sigma}_Y\big)\right)\Phi\right)+\frac{t\sqrt{n}}{V}(\bar{X}-\mu_X)^T\Omega_t\xi_X \\
& +\frac{t\sqrt{n}}{V}(\bar{Y}-\mu_Y)^T\Omega_t\xi_Y-\frac{t^2}{2V^2}\Big(4\left\|\Sigma^{1/2}\Phi\Sigma^{1/2}\right\|_F^2+\xi_X^T\Omega_t\xi_X+\xi_Y^T\Omega_t\xi_Y\Big) \\
& +o(1).
\end{aligned}
$$

We approximate $\frac{2t\sqrt{n}}{V}\mathrm{tr}((\Sigma-\frac{1}{2}\big(\tilde{\Sigma}_X+\tilde{\Sigma}_Y\big))\Phi)$ by $\frac{2t\sqrt{n}}{V}\mathrm{tr}((\Sigma-\hat{\Sigma})\Phi)$, and the approximation error is bounded by

$$
\begin{aligned}
& \frac{C\sqrt{n}}{V}\left|\mathrm{tr}\Big((\hat{\Sigma}_X-\tilde{\Sigma}_X)\Phi\Big)\right|+\frac{C\sqrt{n}}{V}\left|\mathrm{tr}\Big((\hat{\Sigma}_Y-\tilde{\Sigma}_Y)\Phi\Big)\right| \\
\le \quad & CV^{-1}\sqrt{n}\Big((\bar{X}-\mu_X)^T\Phi(\bar{X}-\mu_X)+(\bar{Y}-\mu_Y)^T\Phi(\bar{Y}-\mu_Y)\Big) \\
\le \quad & C||\Phi||V^{-1}\sqrt{n}\Big(||\bar{X}-\mu_X||^2+||\bar{Y}-\mu_Y||^2\Big) \\
\le \quad & C||\Phi||V^{-1}\sqrt{n}\Big(||\mu_X-\mu_X^*||^2+||\mu_Y-\mu_Y^*||^2+O_P(p/n)\Big) \\
= \quad & o_P(1),
\end{aligned}
$$

under A_n and the assumption $p^2/n=o(1)$, where we have used the fact that $||\Phi||/V\le C$. We approximate $\frac{t\sqrt{n}}{V}(\bar{X}-\mu_X)^T\Omega_t\xi_X$ by $\frac{t\sqrt{n}}{V}(\bar{X}-\mu_X)^T\Omega^*\xi_X$, and the difference is bounded by

$$
\begin{aligned}
& C\sqrt{n}V^{-1}\left|\xi_X^T(\Omega_t-\Omega)(\bar{X}-\mu_X)\right|+C\sqrt{n}V^{-1}\left|\xi_X^T(\Omega-\Omega^*)(\bar{X}-\mu_X)\right| \\
\le \quad & CV^{-2}||\xi_X||||\bar{X}-\mu_X||||\Phi||+C\sqrt{n}V^{-1}||\xi_X||||\Omega-\Omega^*||||\bar{X}-\mu_X|| \\
\le \quad & C||\bar{X}-\mu_X||+C\sqrt{n}||\Omega-\Omega^*||||\bar{X}-\mu_X|| \\
\le \quad & C\Big(||\mu_X-\mu_X^*||+\sqrt{n}||\mu_X-\mu_X^*||||\Omega-\Omega^*||+\sqrt{p}||\Omega-\Omega^*||+O_P(\sqrt{p/n})\Big) \\
= \quad & o_P(1),
\end{aligned}
$$

under A_n and the fact that $||\xi_X||/V\le C$. Using the same argument, we can also approximate $\frac{t\sqrt{n}}{V}(\bar{Y}-\mu_Y)^T\Omega_t\xi_Y$ by $\frac{t\sqrt{n}}{V}(\bar{Y}-\mu_Y)^T\Omega^*\xi_Y$. Now we approximate

the quadratic terms. Using the same argument in the proof of Lemma 6.2, we have

$$\frac{\left|\left\|\Sigma^{1/2}\Phi\Sigma^{1/2}\right\|_F^2 - \left\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\right\|_F^2\right|}{V^2} = o(1).$$

We also have

$$\frac{|\xi_X^T(\Omega_t - \Omega^*)\xi_X|}{V^2} \le C\Big(||\Omega - \Omega^*|| + ||\Omega_t - \Omega||\Big) = o(1),$$

and the same bound for $\frac{|\xi_Y^T(\Omega_t-\Omega^*)\xi_Y|}{V^2}$. Therefore,

$$\frac{t^2}{2V^2}\Big(4\left\|\Sigma^{1/2}\Phi\Sigma^{1/2}\right\|_F^2 + \xi_X^T\Omega_t\xi_X + \xi_Y^T\Omega_t\xi_Y\Big) = \frac{t^2}{2} + o(1),$$

on A_n. The proof is complete by considering all the approximations above. □

Lemma A.2. *Under the same setting of Lemma A.1 and further assume* $V^{-1} = O(1)$*, we have*

$$\begin{aligned}&\frac{t\sqrt{n}}{V}\Big(\Delta(\mu_X,\mu_Y,\Omega) - \Delta(\bar{X},\bar{Y},\hat{\Sigma}^{-1})\Big)\\ =\ &\frac{2t\sqrt{n}}{V}tr\Big((\Sigma-\hat{\Sigma})\Phi\Big) - \frac{t\sqrt{n}}{V}(\bar{X}-\mu_X)^T\Omega^*\xi_X - \frac{t\sqrt{n}}{V}(\bar{Y}-\mu_Y)^T\Omega^*\xi_Y\\ &+o_P(1),\end{aligned}$$

uniformly on A_n.

Proof of Theorem 4.1. Combining Lemma A.1 and Lemma A.2, we have

$$\begin{aligned}&l_n(\mu_{X,t},\mu_{Y,t},\Omega_t) - l_n(\mu_X,\mu_Y,\Omega)\\ =\ &\frac{t\sqrt{n}}{V}\Big(\Delta(\mu_X,\mu_Y,\Omega) - \Delta(\bar{X},\bar{Y},\hat{\Sigma}^{-1})\Big) - \frac{1}{2}t^2 + o_P(1),\end{aligned}$$

uniformly in A_n. The remaining of the proof is the same as the proof of Theorem 2.1. □

The proof of Theorem 4.3, is very similar to the proof of Theorem 4.1. We simply state the technical steps in the following lemmas and omit the details of the proof.

Lemma A.3. *Under the setting of Theorem 4.3, assume* $p^2/n = o(1)$ *and* $||\Sigma^*|| \vee ||\Omega^*|| = O(1)$*, then we have*

$$\begin{aligned}&l_n(\mu_{X,t},\mu_{Y,t},\Omega_{X,t},\Omega_{Y,t}) - l_n(\mu_X,\mu_Y,\Omega_X,\Omega_Y)\\ =\ &\frac{t\sqrt{n}}{V}tr\Big((\Sigma_X-\hat{\Sigma}_X)\Phi_X\Big) + \frac{t\sqrt{n}}{V}tr\Big((\Sigma_Y-\hat{\Sigma}_Y)\Phi_Y\Big)\\ &-\frac{t\sqrt{n}}{V}(\bar{X}-\mu_X)^T\Omega^*\xi_X - \frac{t\sqrt{n}}{V}(\bar{Y}-\mu_Y)^T\Omega^*\xi_Y - \frac{1}{2}t^2 + o_P(1),\end{aligned}$$

uniformly on A_n.

Lemma A.4. *Under the same setting of Lemma A.3 and further assume* $V^{-1} = O(1)$ *and* $p^3/n = o(1)$*, then*

$$\begin{aligned} & \frac{t\sqrt{n}}{V}\Big(\Delta(\mu_X,\mu_Y,\Omega_X,\Omega_Y) - \Delta(\bar{X},\bar{Y},\hat{\Sigma}_X^{-1},\hat{\Sigma}_Y^{-1})\Big) \\ = \; & \frac{t\sqrt{n}}{V} tr\Big((\Sigma_X - \hat{\Sigma}_X)\Phi_X\Big) + \frac{t\sqrt{n}}{V} tr\Big((\Sigma_Y - \hat{\Sigma}_Y)\Phi_Y\Big) \\ & - \frac{t\sqrt{n}}{V}(\bar{X}-\mu_X)^T\Omega^*\xi_X - \frac{t\sqrt{n}}{V}(\bar{Y}-\mu_Y)^T\Omega^*\xi_Y + o_P(1), \end{aligned}$$

uniformly on A_n*.*

Proof of Theorem 4.3. Combining Lemma A.3 and Lemma A.4, we have

$$\begin{aligned} & l_n(\mu_{X,t},\mu_{Y,t},\Omega_{X,t},\Omega_{Y,t}) - l_n(\mu_X,\mu_Y,\Omega_X,\Omega_Y) \\ = \; & \frac{t\sqrt{n}}{V}\Big(\Delta(\mu_X,\mu_Y,\Omega_X,\Omega_Y) - \Delta(\bar{X},\bar{Y},\hat{\Sigma}_X^{-1},\hat{\Sigma}_Y^{-1})\Big) - \frac{1}{2}t^2 + o_P(1), \end{aligned}$$

uniformly in A_n. The remaining of the proof is the same as the proof of Theorem 2.1. □

Appendix B: Proof of Theorem 4.2 & Theorem 4.4

In this section, we are going to prove Theorem 4.2 and Theorem 4.4. Due to the similarity of the two theorems, we only present the details of the proof of Theorem 4.4. The proof of Theorem 4.2 will be outlined. By the remark after Theorem 4.4, it is sufficient to check the two conditions in Theorem 4.4 for X and Y separately. Therefore, we only prove for the X part and omit the subscript X from now on.

Denote the prior for (Ω,μ) as $\Pi = \Pi_\Omega \times \Pi_\mu$. The following lemma is a generalization of Lemma 6.5 to the nonzero mean case.

Lemma B.1. *Let* ϵ *be any sequence such that* $\epsilon \to 0$*. Define*

$$K_n = \left\{ ||\Omega||^2 ||\Sigma - \Sigma^*||_F^2 + 2||\Omega|| ||\mu - \mu^*||^2 \le \epsilon^2 \right\}.$$

Then for any $b > 0$*, we have*

$$\begin{aligned} & P_{\Sigma^*}^n\left(\int \exp\Big(l_n(\Omega) - l_n(\Omega^*)\Big) d\Pi(\Omega) \le \Pi(K_n)\exp\big(-(b+1)n\epsilon^2\big)\right) \\ \le \; & \exp\Big(-Cb^2 n\epsilon^2\Big), \end{aligned}$$

for some constant $C > 0$*.*

Proof. We renormalize the prior Π as $\tilde{\Pi} = \Pi(K_n)^{-1}\tilde{\Pi}$ so that $\tilde{\Pi}$ is a distribution with support within K_n. Write $\mathbb{E}_{\tilde{\Pi}}$ to be the expectation using probability $\tilde{\Pi}$. Define the random variable

$$\begin{aligned} Y_i &= \int \log \frac{dP_\Sigma}{dP_{\Sigma^*}}(X_i) d\tilde{\Pi}(\Omega) \\ &= c + \frac{1}{2}(X_i - \mu^*)^T(\Omega^* - \mathbb{E}_{\tilde{\Pi}}\Omega)(X_i - \mu^*) + (X_i - \mu^*)^T \mathbb{E}_{\tilde{\Pi}}\Big(\Omega(\mu - \mu^*)\Big), \end{aligned}$$

for $i = 1, \ldots, n$, where c is a constant independent of $X_1, \ldots, X_n$. Then, Y_i is a sub-exponential random variable with mean

$$\begin{aligned} -P_{\Sigma^*} Y_i &= \int D(P_{\Sigma^*} || P_\Sigma) d\tilde{\Pi}(\Omega) \\ &\le \int \left(\frac{1}{4} ||\Omega||^2 ||\Sigma - \Sigma^*||_F^2 + \frac{1}{2} ||\Omega|| ||\mu - \mu^*||^2 \right) d\tilde{\Pi}(\Omega) \\ &\le \epsilon^2/4. \end{aligned}$$

Thus, by Jensen's inequality, we have

$$\begin{aligned} & P^n_{\Sigma^*}\left(\int \frac{dP^n_\Sigma}{dP^n_{\Sigma^*}}(X^n) d\tilde{\Pi}(\Omega) \le \exp\Big(-(b+1)n\epsilon^2 \Big) \right) \\ \le\ & P^n_{\Sigma^*}\left(\frac{1}{n}\sum_{i=1}^n Y_i \le -(b+1)\epsilon^2 \right) \\ \le\ & P^n_{\Sigma^*}\left(\frac{1}{n}\sum_{i=1}^n (Y_i - P_{\Sigma^*} Y_i) \le -b\epsilon^2 \right) \\ =\ & P^n_{\Sigma}\left(\frac{1}{n}\sum_{i=1}^n (Y_{1i} - P_{\Sigma^*} Y_{1i}) + \frac{1}{n}\sum_{i=1}^n (Y_{2i} - P_{\Sigma^*} Y_{2i}) \le -b\epsilon^2 \right), \end{aligned}$$

where in the last equality we defined

$$Y_{1i} = \frac{1}{2}(X_i - \mu^*)^T(\Omega^* - \mathbb{E}_{\tilde{\Pi}}\Omega)(X_i - \mu^*), \quad Y_{2i} = (X_i - \mu^*)^T \mathbb{E}_{\tilde{\Pi}}\Big(\Omega(\mu - \mu^*)\Big),$$

for $i = 1, \ldots, n$. By union bound, we have

$$\begin{aligned} & P^n_{\Sigma}\left(\frac{1}{n}\sum_{i=1}^n (Y_{1i} - P_{\Sigma^*} Y_{1i}) + \frac{1}{n}\sum_{i=1}^n (Y_{2i} - P_{\Sigma^*} Y_{2i}) \le -b\epsilon^2 \right) \\ \le\ & P^n_{\Sigma}\left(\frac{1}{n}\sum_{i=1}^n (Y_{1i} - P_{\Sigma^*} Y_{1i}) \le -\frac{b\epsilon^2}{2} \right) \\ & + P^n_{\Sigma}\left(\frac{1}{n}\sum_{i=1}^n (Y_{2i} - P_{\Sigma^*} Y_{2i}) \le -\frac{b\epsilon^2}{2} \right). \end{aligned}$$

In the proof of Lemma 5.1 of [17], we have shown that

$$P^n_{\Sigma}\left(\frac{1}{n}\sum_{i=1}^n (Y_{1i} - P_{\Sigma^*} Y_{1i}) \le -\frac{b\epsilon^2}{2} \right) \le \exp\Big(-Cb^2 n\epsilon^2 \Big).$$

Hence, it is sufficient to bound the second term. Define $Z_i = \Omega^{*1/2}(X_i - \mu^*)$, and then we have

$$Y_{2i} - P_{\Sigma^*} Y_{2i} = Z_i^T a \quad \text{and} \quad Z_i \sim N(0, I_{p\times p})$$

with $a = \Sigma^{*1/2}\mathbb{E}_{\tilde{\Pi}}(\Omega(\mu - \mu^*))$. By Bernstein's inequality (see, for example, Proposition 5.16 of [29]), we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n a^T Z_i \le -\frac{b\epsilon^2}{2}\right) \le \exp\left(-C\min\left(\frac{(nb\epsilon^2)^2}{n||a||^2}, \frac{nb\epsilon^2}{||a||_\infty}\right)\right).$$

Since

$$\begin{aligned} ||a||^2 &\le ||\Sigma^*|| \left\|\mathbb{E}_{\tilde{\Pi}}\Big(\Omega(\mu - \mu^*)\Big)\right\|^2 \\ &\le ||\Sigma^*||\mathbb{E}_{\tilde{\Pi}}||\Omega(\mu - \mu^*)||^2 \\ &\le C'\epsilon^2, \end{aligned}$$

and $||a||_\infty \le ||a|| \le \sqrt{C'\epsilon^2}$, then

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n a^T Z_i \le -\frac{b\epsilon^2}{2}\right) \le \exp\Big(-C\min\big(b^2 n\epsilon^2, bn\epsilon\big)\Big) = \exp\Big(-Cb^2 n\epsilon^2\Big),$$

because $\epsilon \to 0$. The conclusion follows the fact that

$$\begin{aligned} & P_{\Sigma^*}\left(\int \frac{dP_\Sigma^n}{dP_{\Sigma^*}^n}(X^n) d\Pi(\Omega) \le \Pi(K_n)\exp\Big(-(b+1)n\epsilon^2\Big)\right) \\ \le\ & P_{\Sigma^*}^n\left(\int \frac{dP_\Sigma^n}{dP_{\Sigma^*}^n}(X^n) d\tilde{\Pi}(\Omega) \le \exp\Big(-(b+1)n\epsilon^2\Big)\right). \end{aligned}$$ □

The following lemma proves prior concentration.

Lemma B.2. *Assume $p^2 = o(n/\log n)$, $||\mu^*|| = O(1)$ and $||\Sigma^*|| \vee ||\Omega^*|| \le \Lambda = O(1)$. For the prior $\Pi = \Pi_\Omega \times \Pi_\mu$, we have*

$$\Pi\Big(||\Omega||^2||\Sigma - \Sigma^*||_F^2 + 2||\Omega||||\mu - \mu^*||^2 \le \frac{p^2\log n}{n}\Big) \ge \exp\Big(-Cp^2\log n\Big),$$

for some constant $C > 0$.

Proof. We have

$$\begin{aligned} & \Pi\Big(||\Omega||^2||\Sigma - \Sigma^*||_F^2 + 2||\Omega||||\mu - \mu^*||^2 \le \frac{p^2\log n}{n}\Big) \\ \ge\ & \Pi\Big(||\Omega||^2||\Sigma - \Sigma^*||_F^2 + 4\Lambda||\mu - \mu^*||^2 \le \frac{p^2\log n}{n}\Big) \\ \ge\ & \Pi\Big(||\Omega||^2||\Sigma - \Sigma^*||_F^2 \le \frac{p^2\log n}{2n}, 4\Lambda||\mu - \mu^*||^2 \le \frac{p^2\log n}{2n}\Big) \\ =\ & \Pi_\Omega\Big(||\Omega||^2||\Sigma - \Sigma^*||_F^2 \le \frac{p^2\log n}{2n}\Big)\Pi_\mu\Big(4\Lambda||\mu - \mu^*||^2 \le \frac{p^2\log n}{2n}\Big), \end{aligned}$$

where the first term is lower bounded in Lemma 6.4. It is sufficient to lower bound $\Pi_\mu(4\Lambda||\mu-\mu^*||^2 \le \frac{p^2\log n}{2n})$. By the definition of Gaussian density,

$$\begin{aligned}
&\Pi_\mu\Big(4\Lambda||\mu-\mu^*||^2 \le \frac{p^2\log n}{2n}\Big)\\
\ge\ & e^{-||\mu^*||^2/2}\left(\mathbb{P}\Big(|Z|^2 \le \frac{p\log n}{cn}\Big)\right)^p\\
\ge\ & \exp\Big(-||\mu^*||^2/2 - Cp\log n\Big).
\end{aligned}$$

The proof is complete by noticing $||\mu^*|| = O(1)$. □

Lemma B.3. *Assume $||\Sigma^*||\vee||\Omega^*|| \le \Lambda = O(1)$. Then for any constant $M > 0$, there exists a testing function ϕ such that*

$$P^n_{(\mu^*,\Omega^*)}\phi \le \exp\Big(-CM^2p^2\log n\Big),$$

$$\sup_{\{(\mu,\Omega)\in supp(\Pi):||\mu-\mu^*||>M\sqrt{\frac{p^2\log n}{n}}\}} P^n_{(\mu,\Omega)}(1-\phi) \le \exp\Big(-CM^2p^2\log n\Big),$$

for some constant $C > 0$.

Proof. Use notation $\epsilon^2 = p^2\log n/n$. Consider the testing function

$$\phi = \left\{||\bar{X}-\mu^*|| > \frac{M\epsilon}{2}\right\}.$$

Then we have

$$P^n_{(\mu^*,\Omega^*)}\phi = \mathbb{P}\left(\frac{1}{\sqrt{n}}||\Sigma^{*1/2}Z\Sigma^{*1/2}|| > \frac{M\epsilon}{2}\right) \le \mathbb{P}\Big(||Z||^2 \ge CM^2n\epsilon^2\Big),$$

where $Z \sim N(0, I_{p\times p})$. We also have for any (μ,Ω) in the alternative set,

$$\begin{aligned}
P^n_{(\mu,\Omega)}(1-\phi) &\le P^n_{(\mu,\Omega)}\left(||\mu-\mu^*|| - ||\bar{X}-\mu|| \le \frac{M}{2}\epsilon\right)\\
&\le P^n_{(\mu,\Omega)}\left(||\bar{X}-\mu|| > \frac{M\epsilon}{2}\right)\\
&= \mathbb{P}\left(\frac{1}{\sqrt{n}}||\Sigma^{1/2}Z\Sigma^{1/2}|| > \frac{M\epsilon}{2}\right)\\
&\le \mathbb{P}\Big(||Z||^2 \ge CM^2n\epsilon^2\Big).
\end{aligned}$$

Finally, it is sufficient to bound $\mathbb{P}(||Z||^2 \ge CM^2n\epsilon^2)$. We have

$$\begin{aligned}
& \mathbb{P}\Big(||Z||^2 \geq CM^2n\epsilon^2\Big) \\
= \;& \mathbb{P}\bigg(\sum_{j=1}^{p}(Z_j^2-1) \geq CM^2n\epsilon^2 - p\bigg) \\
\leq \;& \mathbb{P}\bigg(\sum_{j=1}^{p}(Z_j^2-1) \geq CM^2n\epsilon^2/2\bigg) \\
\leq \;& \exp\Big(-C\min\big((M^2n\epsilon^2)^2/p, M^2n\epsilon^2\big)\Big) \\
= \;& \exp\Big(-CM^2n\epsilon^2\Big),
\end{aligned}$$

where we have used Bernstein's inequality. The proof is complete. □

Lemma B.4. *Assume* $||\Sigma^*|| \vee ||\Omega^*|| \leq \Lambda = O(1)$ *and* $||\Sigma_1|| \vee ||\Omega_1|| \leq 2\Lambda$. *There exist small* $\delta, \delta', \bar{\delta} > 0$ *only depending on* Λ *such that for any* $M > 0$, *there exists a testing function* ϕ *such that*

$$P^n_{(\mu^*,\Omega^*)}\phi \leq 2\exp\Big(-C\delta'||\Sigma^* - \Sigma_1||_F^2\Big),$$

$$\sup_{\left\{\substack{(\mu,\Omega)\in supp(\Pi):||\Sigma-\Sigma_1||_F\leq\delta||\Sigma^*-\Sigma_1||_F,\\ ||\mu-\mu^*||\leq M\epsilon}\right\}} P^n_{\Sigma}(1-\phi) \leq 2\exp\Big(-C\delta'||\Sigma^* - \Sigma_1||_F^2\Big),$$

for some constant $C > 0$, *whenever* $6\Lambda M^2\epsilon^2 \leq \bar{\delta}||\Sigma_1 - \Sigma^*||_F^2$.

Proof. Since the lemma is a slight variation of Lemma 5.9 in [17]. We do not write the proof in full details. We choose to highlight the part where the current form is different from that in [17], and omit the similar part where the readers may find its full details in the proof of Lemma 5.9 in Gao and Zhou. We use the testing function

$$\phi = \left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu^*)^T(\Omega^*-\Omega_1)(X_i-\mu^*) > \log\det(\Omega\Sigma_1)\right\}.$$

We immediately have

$$P^n_{(\mu^*,\Omega^*)}\phi \leq 2\exp\Big(-C\delta'||\Sigma_1 - \Sigma^*||_F^2\Big),$$

as is proved in [17]. Now we are going to bound $P^n_{(\mu,\Omega)}(1-\phi)$ for every (μ,Ω) in the alternative set. Note that we have

$$\begin{aligned}
& 1-\phi \\
= \;& \left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^T(\Omega^*-\Omega_1)(X_i-\mu) + 2(\bar{X}-\mu)^T(\Omega^*-\Omega)(\mu-\mu^*)\right. \\
& \left.+(\mu-\mu^*)^T(\Omega^*-\Omega_1)(\mu-\mu^*) < \log\det(\Omega\Sigma_1)\right\}
\end{aligned}$$

$$
\begin{aligned}
= \quad & \Bigg\{ \frac{1}{n} \sum_{i=1}^{n} \Bigg((X_i - \mu)^T (\Omega^* - \Omega_1)(X_i - \mu) \\
& - P_{(\mu,\Omega)} (X_i - \mu)^T (\Omega^* - \Omega_1)(X_i - \mu) \Bigg) \\
& + 2(\bar{X} - \mu)^T (\Omega^* - \Omega_1)(\mu - \mu^*) + (\mu - \mu^*)^T (\Omega_1 - \Omega^*)(\mu - \mu^*) > \bar{\rho} \Bigg\},
\end{aligned}
$$

where we have proved in [17] that

$$
\bar{\rho} \geq \bar{\delta} ||\Sigma_1 - \Sigma^*||_F^2,
$$

for some $\bar{\delta}$ only depending on Λ. Using union bound, we have

$$
\begin{aligned}
& P^n_{(\mu,\Omega)} (1 - \phi) \\
\leq \quad & P^n_{(\mu,\Omega)} \Bigg\{ \frac{1}{n} \sum_{i=1}^{n} \Bigg((X_i - \mu)^T (\Omega^* - \Omega_1)(X_i - \mu) \\
& - P_{(\mu,\Omega)} (X_i - \mu)^T (\Omega^* - \Omega_1)(X_i - \mu) \Bigg) > \frac{\bar{\rho}}{2} \Bigg\} \\
& + P^n_{(\mu,\Omega)} \Big\{ 2(\bar{X} - \mu)^T (\Omega^* - \Omega_1)(\mu - \mu^*) \\
& + (\mu - \mu^*)^T (\Omega_1 - \Omega^*)(\mu - \mu^*) > \frac{\bar{\rho}}{2} \Big\}.
\end{aligned}
$$

[17] showed that the first term above is bounded by $2\exp(-C\delta' ||\Sigma_1 - \Sigma^*||_F^2)$. It is sufficient to bound the second term to close the proof. Actually, this is the only difference between this proof and the one in [17]. Note that

$$
|(\mu - \mu^*)^T (\Omega_1 - \Omega^*)(\mu - \mu^*)| \leq 3\Lambda ||\mu - \mu^*||^2 \leq 3\Lambda M^2 \epsilon^2.
$$

By assumption,

$$
3\Lambda M^2 \epsilon^2 \leq \frac{1}{2} \bar{\delta} ||\Sigma_1 - \Sigma^*||_F^2 \leq \frac{\bar{\rho}}{4}.
$$

Hence,

$$
\begin{aligned}
& P^n_{(\mu,\Omega)} \Big\{ 2(\bar{X} - \mu)^T (\Omega^* - \Omega_1)(\mu - \mu^*) + (\mu - \mu^*)^T (\Omega_1 - \Omega^*)(\mu - \mu^*) > \frac{\bar{\rho}}{2} \Big\} \\
& \leq P^n_{(\mu,\Omega)} \Big\{ 2(\bar{X} - \mu)^T (\Omega^* - \Omega_1)(\mu - \mu^*) > \frac{\bar{\rho}}{4} \Big\} \\
& = \mathbb{P} \Bigg(Z^T a > \frac{\bar{\rho}}{8} \Bigg),
\end{aligned}
$$

where $Z \sim N(0,1)$ and $a = \Sigma^{1/2} (\Omega^* - \Omega)(\mu - \mu^*)$. Using Hoeffding's inequality (see, for example, Proposition 5.10 of [29]), we have

$$
\mathbb{P} \Bigg(Z^T a > \frac{\bar{\rho}}{8} \Bigg) \leq \exp \Big(- \frac{C\bar{\rho}^2}{||a||^2} \Big),
$$

where

$$||a||^2 \leq 9\Lambda^3||\mu-\mu^*||^2 \leq 9\Lambda^3 M^2\epsilon^2 \leq \frac{3\Lambda^2}{4}\bar{\rho},$$

according to the assumption. Thus,

$$\mathbb{P}\left(Z^T a > \frac{\bar{\rho}}{8}\right) \leq \exp\Big(-C\delta'||\Sigma_1-\Sigma^*||_F^2\Big),$$

for some δ' only depending on Λ. Therefore, $P^n_{(\mu,\Omega)}(1-\phi) \leq \exp(-C\delta'||\Sigma_1-\Sigma^*||_F^2)$ for all (μ,Ω) in the alternative set and the proof is complete. □

Proof of Theorem 4.2 and Theorem 4.4. According to the remark after Theorem 4.3,

$$\Pi(A_n|X^n,Y^n) = \Pi_X(A_{X,n}|X^n)\Pi_Y(A_{Y,n}|Y^n).$$

Thus, it is sufficient to show both $\Pi_X(A_{X,n}|X^n)$ and $\Pi_Y(A_{Y,n}|Y^n)$ converge to 1 in probability. Since they have the same form, we treat them together by omitting the subscript X and Y. The posterior distribution is defined as

$$\Pi(A_n^c|X^n) = \frac{\int_{A_n^c}\exp\Big(l_n(\mu,\Omega)-l_n(\mu^*,\Omega^*)\Big)d\Pi(\mu,\Omega)}{\int\exp\Big(l_n(\mu,\Omega)-l_n(\mu^*,\Omega^*)\Big)d\Pi(\mu,\Omega)} = \frac{N_n}{D_n},$$

where we consider

$$A_n = \left\{||\mu-\mu^*|| \leq M\sqrt{\frac{p^2\log n}{n}}, ||\Sigma-\Sigma^*||_F \leq \bar{M}\sqrt{\frac{p^2\log n}{n}}\right\},$$

for some M and $\bar{M}$ sufficiently large. We are going to establish a test between the following hypotheses:

$$H_0 : (\mu,\Omega) = (\mu^*,\Omega^*) \quad \text{vs} \quad H_1 : (\mu,\Omega) \in A_n^c \cap \text{supp}(\Pi).$$

Decompose A_n^c as

$$A_n^c = B_{1n} \cup B_{2n},$$

where

$$B_{1n} = \left\{||\mu-\mu^*|| > M\sqrt{\frac{p^2\log n}{n}}\right\},$$

and

$$B_{2n} = \left\{||\mu-\mu^*|| \leq M\sqrt{\frac{p^2\log n}{n}}, ||\Sigma-\Sigma^*||_F > \bar{M}\sqrt{\frac{p^2\log n}{n}}\right\}.$$

By Lemma B.3, there exists ϕ_1 such that

$$P^n_{(\mu^*,\Omega^*)}\phi_1 \vee \sup_{\text{supp}(\Pi)\cap B_{1n}} P^n_{(\mu,\Omega)}(1-\phi_1) \leq \exp\Big(-CM^2p^2\log n\Big).$$

For B_{2n}, we pick a covering set $\{\Sigma_j\}_{j=1}^N \subset B_{2n} \cap \text{supp}(\Pi)$, such that

$$B_{2n} \subset \cup_{j=1}^N B_{2nj},$$

where

$$B_{2nj} = \left\{ ||\mu - \mu^*|| \leq M\sqrt{\frac{p^2 \log n}{n}}, ||\Sigma - \Sigma_j||_F \leq \sqrt{\frac{p^2 \log n}{n}} \right\},$$

and the covering number N can be chosen to satisfy

$$\log N \leq Cp^2 \log n,$$

as is shown in detail in the proof of Lemma 2.2. We may choose $\bar{M}$ large enough so that the assumption of Lemma B.4 is satisfied, which implies the existence of ϕ_{2j} such that

$$P^n_{(\mu^*,\Omega^*)}\phi_{2j} \vee \sup_{\text{supp}(\Pi)\cap B_{2nj}} P^n_{(\mu,\Omega)}(1-\phi_{2j}) \leq \exp\Big(- C\bar{M}^2 p^2 \log n\Big).$$

Define the final test as $\phi = \max(\phi_1, \vee_{j=1}^N \phi_{2j})$. Then using union bound, we have

$$P^n_{(\mu^*,\Omega^*)}\phi \vee \sup_{\text{supp}(\Pi)\cap A_n^c} P^n_{(\mu,\Omega)}(1-\phi) \leq \exp\Big(- C(\bar{M}^2 \wedge M^2)p^2 \log n\Big),$$

for large M and $\bar{M}$. Combining the testing result and the conclusions from Lemma B.1 and Lemma B.2, we have

$$P^n_{(\mu^*,\Omega^*)}\Pi(A_n^c|X^n) = o_P(1),$$

by using the same argument in the proof of Lemma 2.2. For QDA, as long as $p^2 = o(\frac{\sqrt{n}}{\log n})$, A_n satisfies the requirement. For LDA, we use a A_n defined as

$$A_n = \left\{ ||\mu_X - \mu_X^*|| \vee ||\mu_Y - \mu_Y^*|| \leq M\sqrt{\frac{p^2 \log n}{n}}, ||\Sigma - \Sigma^*||_F \leq \bar{M}\sqrt{\frac{p^2 \log n}{n}} \right\}.$$

The proof needs some slight modification (including the previous lemmas) which is not essential and we choose to omit here. When $p^2 = o(\frac{\sqrt{n}}{\log n})$ is true, A_n also satisfies the requirement there.

Now we are going to check the second conditions of Theorem 4.1 and Theorem 4.3. We mainly sketch the QDA case. Using the notation in Lemma 2.2,

$$\int_{A_n} \exp\Big(l_n(\Omega_t,\mu_t)\Big) d\Pi(\Omega,\mu) = \xi_p^{-1} \int_{A_n} \exp\Big(l_n(\Omega_t,\mu_t) - \frac{1}{2}||\bar{\Omega}||_F^2 - \frac{1}{2}||\mu||^2\Big) d\bar{\Omega} d\mu,$$

where ξ_p is a normalizing constant and

$$\begin{aligned}
&\int_{A_n} \exp\Big(l_n(\Omega_t,\mu_t) - \frac{1}{2}||\bar{\Omega}||_F^2 - \frac{1}{2}||\mu||^2\Big) d\bar{\Omega} d\mu \\
= &\int_{A_n + (2tn^{-1/2}\Phi, tn^{-1/2}\xi)} \\
&\exp\Big(l_n(\Gamma,\theta) - \frac{1}{2}||\bar{\Gamma} - 2tn^{1/2}\bar{\Phi}||_F^2 - \frac{1}{2}||\theta - tn^{-1/2}\xi||^2\Big) d\bar{\Gamma} d\theta.
\end{aligned}$$

Proceeding as in Lemma 2.2, the result is proved. □

Appendix C: Proof of Lemma 3.1

Let $\hat{\Omega} = \hat{\Sigma}^{-1}$. Note that

$$\begin{aligned}
&\Big|\log\det(\Sigma) - \log\det(\hat{\Sigma}) - \mathrm{tr}\Big((\Sigma - \hat{\Sigma})\Omega^*\Big)\Big| \\
\leq\ &\Big|\log\det\Big(\hat{\Omega}^{1/2}\Sigma\hat{\Omega}^{1/2} - I + I\Big) - \mathrm{tr}\Big(\hat{\Omega}^{1/2}\Sigma\hat{\Omega}^{1/2} - I\Big)\Big| \\
&+ \Big|\mathrm{tr}\Big((\Sigma - \hat{\Sigma})(\Omega^* - \hat{\Omega})\Big)\Big|.
\end{aligned}$$

For the second term,

$$\begin{aligned}
&\sqrt{\frac{n}{p}}\Big|\mathrm{tr}\Big((\Sigma - \hat{\Sigma})(\Omega^* - \hat{\Omega})\Big)\Big| \\
\leq\ &C\sqrt{\frac{n}{p}}||\Sigma - \hat{\Sigma}||_F||\hat{\Sigma} - \Sigma^*||_F \\
\leq\ &C\sqrt{\frac{n}{p}}\Big(||\hat{\Sigma} - \Sigma^*||_F^2 + ||\Sigma - \Sigma^*||_F||\hat{\Sigma} - \Sigma^*||_F\Big) \\
\leq\ &O_P\left(\sqrt{\frac{p^3}{n}} + \sqrt{p}||\Sigma - \Sigma^*||_F\right),
\end{aligned}$$

which converges to zero whenever $\sqrt{p}||\Sigma - \Sigma^*||_F \leq \delta_n = o(1)$. For the first term,

$$\begin{aligned}
&\sqrt{\frac{n}{p}}\Big|\log\det\Big(\hat{\Omega}^{1/2}\Sigma\hat{\Omega}^{1/2} - I + I\Big) - \mathrm{tr}\Big(\hat{\Omega}^{1/2}\Sigma\hat{\Omega}^{1/2} - I\Big)\Big| \\
\leq\ &C\sqrt{\frac{n}{p}}\Big\|\hat{\Omega}^{1/2}\Sigma\hat{\Omega}^{1/2} - I\Big\|_F^2 \\
\leq\ &C\sqrt{\frac{n}{p}}\Big(||\Sigma - \Sigma^*||_F^2 + ||\hat{\Sigma} - \Sigma^*||_F^2\Big) \\
\leq\ &O_P\left(\sqrt{\frac{p^3}{n}} + \sqrt{\frac{n}{p}}||\Sigma - \Sigma^*||_F^2\right),
\end{aligned}$$

which converges to zero whenver $\sqrt{\frac{n}{p}}||\Sigma - \Sigma^*||_F^2 \leq \delta_n = o(1)$. Thus, the proof is complete.

Appendix D: Proof of Lemma 3.2, Lemma 3.3 & Proposition 5.1

Due to the similarity between Lemma 3.2 and Lemma 3.3, we only give the proof of Lemma 3.2. Let us study the linear approximation of eigenvalue perturbation. In particular, we are going to find the first-order Taylor expansion of $\lambda_m(\Sigma) - \lambda_m(\hat{\Sigma})$ and control the error term in some set A_n. We have the following spectral decomposition for the three covariance matrices $\Sigma, \hat{\Sigma}, \Sigma^*$.

$$\Sigma = UDU^T, \quad \hat{\Sigma} = \hat{U}\hat{D}\hat{U}^T, \quad \Sigma^* = U^*D^*U^{*T}.$$

Denote the m-th column of $U, \hat{U}, U^*$ by $u_m, \hat{u}_m, u_m^*$. Then,

$$\begin{aligned}\lambda_m(\Sigma) - \lambda_m(\hat{\Sigma}) &= \lambda_m(\hat{U}^T\Sigma\hat{U}) - \lambda_m(\hat{U}^T\hat{\Sigma}\hat{U}) \\ &= \lambda_m(\hat{D} + \hat{U}^T(\hat{\Sigma} - \Sigma)\hat{U}) - \lambda_m(\hat{D}).\end{aligned}$$

Write $A = \hat{D}$ and $\Delta = \hat{U}^T(\hat{\Sigma} - \Sigma)\hat{U}$. The problem is reduced to the eigenvalue perturbation of a diagonal matrix. According to the expansion formula in [23] and [1], we have

$$\lambda_m(A + \Delta) - \lambda_m(A) = \lambda'_m(A, \Delta) + \sum_{k=2}^{\infty} \lambda_m^{(k)}(A, \Delta), \tag{D.1}$$

where the first-order term is

$$\lambda'_m(A, \Delta) = \Delta_{mm} = \mathrm{tr}\Big(\hat{U}^T(\hat{\Sigma} - \Sigma)\hat{U}E_{mm}\Big)$$

$$= \mathrm{tr}\Big((\hat{\Sigma} - \Sigma)\hat{U}E_{mm}\hat{U}^T\Big) = \mathrm{tr}\Big((\hat{\Sigma} - \Sigma)\hat{u}_m\hat{u}_m^T\Big).$$

In the remainder term, we have

$$\lambda_m^{(k+1)}(A, \Delta) = -\frac{1}{k+1} \sum_{v_1 + \cdots + v_{k+1} = k, v_1, \ldots, v_{k+1} \geq 0} \mathrm{tr}\Big(\Delta\tilde{A}^{v_1} \ldots \Delta\tilde{A}^{v_{k+1}}\Big),$$

where $\tilde{A}^v$ is the matrix power when $v \geq 1$ with an exception that $\tilde{A}^0 = -e_m e_m^T$, where e_m is the m-th vector of the canonical basis of $\mathbb{R}^p$. The matrix $\tilde{A}$ is defined as

$$\tilde{A} = \sum_{1 \leq j \leq p, j \neq m} \frac{e_j e_j^T}{a_m - a_j},$$

where $a_j = \lambda_j(\hat{\Sigma})$ is the (j, j)-th entry of A. Therefore, for any integer $v \geq 1$,

$$||\tilde{A}^v|| = ||\tilde{A}||^v \leq \max\left\{\frac{1}{|a_m - a_{m-1}|}, \frac{1}{|a_m - a_{m+1}|}\right\}.$$

We are going to show that the first term in (D.1) is a good enough approximation of $\lambda_m(A + \Delta) - \lambda_m(A)$ by bounding the higher-order terms. Let us provide a bound for $|\lambda_m^{(k+1)}(A, \Delta)|$. Let $\mathbb{N} = \{0, 1, 2, \ldots\}$. Consider the set

$$\{(v_1, \ldots, v_{k+1}) \in \mathbb{N} : v_1 + \cdots + v_{k+1} = k\}.$$

From its definition, there must be some l, such that $v_l = 0$ to satisfy $v_1 + \cdots + v_{k+1} = k$. Thus, the set can be decomposed into a union of disjoint subsets as follows,

$$\begin{aligned} & \big\{(v_1, \ldots, v_{k+1}) \in \mathbb{N}^{k+1} : v_1 + \cdots + v_{k+1} = k\big\} \\ = & \bigcup_{l=1}^{k+1} \Bigg\{ (v_1, \ldots, v_{l-1}, v_{l+1}, \ldots, v_{k+1}) \in \mathbb{N}^k : \\ & v_1 + \cdots + v_{l-1} + v_{l+1} + \cdots + v_{k+1} = k \Bigg\}. \end{aligned}$$

Clearly, each the cardinality of each subset is

$$\binom{2k-1}{k-1} \leq (3e)^k.$$

We give names to the sets we have mentioned by

$$V_{k+1} = \cup_{l=1}^{k+1} V_{k+1,l}.$$

For $l = 1$, we have

$$\begin{aligned} \left| \sum_{V_{k+1,1}} \mathrm{tr}\Big(\Delta\tilde{A}^{v_1} \ldots \Delta\tilde{A}^{v_{k+1}}\Big) \right| &\leq \sum_{V_{k+1,1}} \left| \mathrm{tr}\Big(\Delta\tilde{A}^{v_1} \ldots \Delta\tilde{A}^{v_{k+1}}\Big) \right| \\ &= \sum_{V_{k+1,1}} \left| \mathrm{tr}\Big(\Delta\tilde{A}^{0}\Delta\tilde{A}^{v_2} \ldots \Delta\tilde{A}^{v_{k+1}}\Big) \right| \\ &= \sum_{V_{k+1,1}} \left| \mathrm{tr}\Big(\Delta e_m e_m^T \Delta\tilde{A}^{v_2} \ldots \Delta\tilde{A}^{v_{k+1}}\Big) \right| \\ &\leq \sum_{V_{k+1,1}} \|\Delta e_m\| \left\| e_m^T \Delta\tilde{A}^{v_2} \ldots \Delta\tilde{A}^{v_{k+1}} \right\| \\ &\leq \sum_{V_{k+1,1}} ||\Delta||^{k+1} ||\tilde{A}||^{v_2 + \cdots + v_{k+1}} \\ &= \sum_{V_{k+1,1}} ||\Delta||^{k+1} ||\tilde{A}||^{k} \\ &= ||\Delta|| \Big(3e||\Delta||||\tilde{A}||\Big)^k \end{aligned}$$

In the same way, the bound also holds for other l. Therefore,

$$\begin{aligned} |\lambda_m^{(k+1)}(A, \Delta)| &= \left| \frac{1}{k+1} \sum_{l=1}^{k+1} \sum_{V_{k+1,l}} \mathrm{tr}\Big(\Delta\tilde{A}^{v_1} \ldots \Delta\tilde{A}^{v_{k+1}}\Big) \right| \\ &\leq \frac{1}{k+1} \sum_{l=1}^{k+1} ||\Delta|| \Big(3e||\Delta||||\tilde{A}||\Big)^k \\ &= ||\Delta|| \Big(3e||\Delta||||\tilde{A}||\Big)^k. \end{aligned}$$

When $3e||\Delta||||\tilde{A}|| < 1$, we may sum over k, and obtain

$$\begin{aligned}\left|\sum_{k=2}^{\infty}\lambda_m^{(k)}(A,\Delta)\right| &\leq ||\Delta||\sum_{k=1}^{\infty}\Big(3e||\Delta||||\tilde{A}||\Big)^k \\ &\leq \frac{3e||\Delta||^2||\tilde{A}||}{1-3e||\Delta||||\tilde{A}||}.\end{aligned}$$

Note that

$$||\Delta|| = ||\hat{\Sigma}-\Sigma|| \leq ||\hat{\Sigma}-\Sigma^*|| + ||\Sigma-\Sigma^*|| \leq O_P\Big(\sqrt{\frac{p}{n}}\Big) + ||\Sigma-\Sigma^*||,$$

and

$$\begin{aligned}||\tilde{A}|| &\leq C\min\Big\{\max\{|\lambda_m-\lambda_{m-1}|^{-1},|\lambda_m-\lambda_{m+1}|^{-1}\},||\hat{\Sigma}-\Sigma||^{-1}\Big\} \\ &= O_P\Big(\min\big(\delta^{-1},\sqrt{n/p}\big)\Big).\end{aligned}$$

Therefore,

$$3e||\Delta||||\tilde{A}|| = O_P\left(\frac{\sqrt{p/n}+||\Sigma-\Sigma^*||}{\delta}\right) = o_P(1),$$

holds under the assumption $\delta^{-1}\sqrt{p/n} = o(1)$ and when $\delta^{-1}||\Sigma-\Sigma^*|| = o_P(1)$. The remainders are controlled by

$$\sqrt{n}\left|\sum_{k=2}^{\infty}\lambda_m^{(k)}(A,\Delta)\right| = O_P\left(\frac{p}{\delta\sqrt{n}}+\frac{\sqrt{n}||\Sigma-\Sigma^*||^2}{\delta}\right) = o_P(1),$$

under the assumption $\frac{p}{\delta\sqrt{n}} = o(1)$ and when $\delta^{-1}\sqrt{n}||\Sigma-\Sigma^*||^2 = o(1)$. Hence, by (D.1), we have proved

$$\sup_{\{\delta^{-1}||\Sigma-\Sigma^*||\vee\delta^{-1}\sqrt{n}||\Sigma-\Sigma^*||^2\leq\delta_n\}}\sqrt{n}\,|\lambda_m(A+\Delta)-\lambda_m(A)-\lambda_m'(A,\Delta)| = o_P(1),$$

for any $\delta_n = o(1)$.

Finally, for the first order term $\lambda_m'(A,\Delta)$, we approximate it by $\mathrm{tr}((\hat{\Sigma}-\Sigma)u_m^*u_m^{*T})$, and the approximation error is

$$\Big|\mathrm{tr}\Big((\hat{\Sigma}-\Sigma)(\hat{u}_m\hat{u}_m^T-u_m^*u_m^{*T})\Big)\Big| = ||u_m^*u_m^{*T}-\hat{u}_m\hat{u}_m^T||_F\left|\mathrm{tr}\Big((\hat{\Sigma}-\Sigma)K\Big)\right|,$$

where K is a rank-two unit Frobenius norm matrix. It has SVD $K = c_1d_1d_1^T + c_2d_2d_2^T$, with $c_1\vee c_2\leq 1$. Therefore,

$$\begin{aligned}&||u_m^*u_m^{*T}-\hat{u}_m\hat{u}_m^T||_F\left|\mathrm{tr}\Big((\hat{\Sigma}-\Sigma)K\Big)\right| \leq C||\hat{\Sigma}-\Sigma||||\hat{\Sigma}-\Sigma^*|| \\ \leq\ & O_P\Big(\frac{p}{n}\Big)+O_P\Big(\sqrt{\frac{p}{n}}||\Sigma-\Sigma^*||\Big).\end{aligned}$$

Under the assumption $p = o(n)$, when $\sqrt{p}||\Sigma - \Sigma^*|| = o(1)$, we have

$$\sqrt{n}\left|\mathrm{tr}\Big((\hat{\Sigma} - \Sigma)(\hat{u}_m\hat{u}_m^T - u_m^* u_m^{*T})\Big)\right| = o_P(1).$$

Therefore, the proof of Lemma 3.2 is complete.

Now we prove Proposition 5.1. We redefine $A = D^*$ and $\Delta = U^{*T}(\Sigma^* - \hat{\Sigma})U^*$ and correspondingly $\tilde{A}^v$. In the case where δ is a constant, we have

$$||\tilde{A}|| \le C, \quad ||\Delta|| = O_P\left(\sqrt{\frac{p}{n}}\right).$$

Similar to (D.1), we have

$$\lambda_1(\hat{\Sigma}) - \lambda_1(\Sigma^*) = \lambda_1'(A,\Delta) + \sum_{k=2}^{\infty}\lambda_1^{(k)}(A,\Delta),$$

where $\lambda_1'(A,\Delta) = \mathrm{tr}((\Sigma^* - \hat{\Sigma})u_1^* u_1^{*T})$ and thus $\sqrt{n}\lambda_1'(A,\Delta)$ is asymptotically normal. For the remainder term, we decompose it as

$$\sum_{k=2}^{\infty}\lambda_1^{(k)}(A,\Delta) = \lambda_1^{(2)}(A,\Delta) + \sum_{k=3}^{\infty}\lambda_1^{(k)}(A,\Delta).$$

Using similar techniques in proving Lemma 3.2, we have

$$\begin{aligned}\sqrt{n}\left|\sum_{k=3}^{\infty}\lambda_1^{(k)}(A,\Delta)\right| &\le \sqrt{n}\sum_{k=2}^{\infty}||\Delta||\Big(3e||\Delta||||\tilde{A}||\Big)^k \\ &\le C\sqrt{n}||\Delta||^3 \\ &= O_P\left(\sqrt{\frac{p^3}{n^2}}\right),\end{aligned}$$

which is $o_P(1)$ under the assumption. Therefore,

$$\sqrt{n}\Big(\lambda_1(\hat{\Sigma}) - \lambda_1(\Sigma^*)\Big) = \sqrt{n}\mathrm{tr}\Big((\Sigma^* - \hat{\Sigma})u_1^* u_1^{*T}\Big) + \sqrt{n}\lambda_1^{(2)}(A,\Delta) + o_P(1).$$

It remains to show that $\sqrt{n}\lambda_1^{(2)}(A,\Delta)$ is not $o_P(1)$. Note that $\tilde{A} = \sum_{j\ge 2}\frac{e_j e_j^T}{a_1 - a_j}$, with $a_j = \lambda_j(\Sigma^*)$. Hence,

$$\begin{aligned}\sqrt{n}\left|\lambda_1^{(2)}(A,\Delta)\right| &= \sqrt{n}\left|\mathrm{tr}\Big(\Delta\tilde{A}\Delta e_1 e_1^T\Big)\right| \\ &= \sqrt{n}\left|\sum_{j\ge 2}\frac{1}{a_1 - a_j}|e_1^T U^{*T}(\Sigma^* - \hat{\Sigma})U^* e_j|^2\right| \\ &\ge \frac{\sqrt{n}}{a_1 - a_2}\sum_{j\ge 2}|e_1^T U^{*T}(\Sigma^* - \hat{\Sigma})U^* e_j|^2.\end{aligned}$$

For fixed eigengap, $a_1 - a_2$ is a constant. When Σ^* is diagonal, $U^* = I$, and we have $|e_1^T U^{*T}(\Sigma^* - \hat{\Sigma})U^* e_j|^2 = \hat{\sigma}_{1j}^2$. Moreover, the fact that Σ^* is diagonal implies $\hat{\sigma}_{1j}^2$ are independent for $j = 2, \ldots, p$. Hence,

$$\sqrt{n}\left|\lambda_1^{(2)}(A, \Delta)\right| \geq C\sqrt{n}\sum_{j=2}^{p}\hat{\sigma}_{1j}^2,$$

where $\sqrt{n}\sum_{j=2}^{p}\hat{\sigma}_{1j}^2$ is at the level of $p/\sqrt{n}$, which diverges to ∞ under the assumption. The proof is complete.

Appendix E: Proof of Corollary 5.1

The proof follows closely to that of Lemma 2.1. First, Theorem 5.2 implies that $\Pi(A_n|X^n) = 1 - o_P(1)$ with A_n defined by (5.9). To check the second condition, let

$$\Omega_t = \Omega + 2tn^{-1/2}\tilde{\Phi},$$

with $\tilde{\Phi} = (\sqrt{2}\|\Sigma^{*1/2}\Phi\Sigma^{*1/2}\|_F)^{-1}\Phi$. Then, using the same argument in the proof of Lemma 2.1, we get

$$\begin{aligned}
&\int_{A_n} \exp\left(l_n(\Omega_t)\right) d\Pi(\Omega) \\
= &\int_{A_n + 2tn^{-1/2}\tilde{\Phi}} \exp\left(l_n(\Omega)\right) \\
&\exp\left(\frac{b-2}{2}\log\det(I - 2tn^{-1/2}\Omega^{-1}\tilde{\Phi}) + \frac{1}{2}\mathrm{tr}(2tn^{-1/2}\tilde{\Phi})\right) d\Pi(\Omega).
\end{aligned}$$

Note that

$$A_n + 2tn^{-1/2}\tilde{\Phi} = \left\{\Omega \in \mathcal{F}_k : \left\|\Omega - 2tn^{-1/2}\tilde{\Phi} - \Omega^*\right\| \leq M\sqrt{\frac{k^5 \log p}{n}}\right\}.$$

Using the fact that $\tilde{\Phi} \in \mathcal{F}_k$ and $||2tn^{-1/2}\tilde{\Phi}|| \leq 2tn^{-1/2}||\tilde{\Phi}||_F = O(n^{-1/2})$, we have

$$A_n' \subset A_n + 2tn^{-1/2}\tilde{\Phi} \subset A_n''$$

for

$$A_n' = \left\{\Omega \in \mathcal{F}_k : ||\Omega - \Omega^*||_{\ell_1} \leq M'\sqrt{\frac{k^5 \log p}{n}}\right\},$$

$$A_n'' = \left\{\Omega \in \mathcal{F}_k : ||\Omega - \Omega^*||_{\ell_1} \leq M''\sqrt{\frac{k^5 \log p}{n}}\right\}.$$

Moreover, $\Pi(A_n'|X^n) = 1 - o_P(1)$ and $\Pi(A_n''|X^n) = 1 - o_P(1)$ by Theorem 5.2 for large M, M', M''. Finally,

$$\sup_{A_n''}\left|\frac{b-2}{2}\log\det(I - 2tn^{-1/2}\Omega^{-1}\tilde{\Phi}) + \frac{1}{2}\mathrm{tr}(2tn^{-1/2}\tilde{\Phi})\right|$$

$$\begin{aligned} &\leq tn^{-1/2} \sup_{A_n''} \left| |b-2| \left\| \Omega^{-1/2} \tilde{\Phi} \Omega^{-1/2} \right\|_N + ||\tilde{\Phi}||_N \right| \\ &= O\left(\sqrt{\frac{r}{n}}\right) = o(1). \end{aligned}$$

Following the rest of the proof of Lemma 2.1, the proof is complete.

References

[1] Ames, B. P. and Sendov, H. S. (2012). A new derivation of a formula by Kato. *Linear Algebra and its Applications* **436** 722–730. MR2854902

[2] Banerjee, S. and Ghosal, S. (2014). Posterior convergence rates for estimating large precision matrices using graphical models. *Electronic Journal of Statistics* **8** 2111–2137. MR3273620

[3] Bernstein, S. (1927). Theory of probability.

[4] Bontemps, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *The Annals of Statistics* **39** 2557–2584. MR2906878

[5] Boucheron, S. and Gassiat, E. (2009). A Bernstein-von Mises theorem for discrete probability distributions. *Electronic Journal of Statistics* **3** 114–148. MR2471588

[6] Cai, T. T., Liang, T. and Zhou, H. H. (2013). Law of Log Determinant of Sample Covariance Matrix and Optimal Estimation of Differential Entropy for High-Dimensional Gaussian Distributions. *arXiv preprint arXiv:1309.0482*. MR3332804

[7] Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics* **38** 2118–2144. MR2676885

[8] Cai, T. T. and Zhou, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics* **40** 2389–2420. MR3097607

[9] Castillo, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probability Theory and Related Fields* **152** 53–99. MR2875753

[10] Castillo, I. and Rousseau, J. (2015). A General Bernstein–von Mises Theorem in semiparametric models. *The Annals of Statistics (to appear)*.

[11] Castillo, I., Schmidt-Hieber, J. and van der Vaart, A. W. (2015). Bayesian linear regression with sparse priors. *The Annals of Statistics (to appear)*.

[12] Cheng, G. and Kosorok, M. R. (2008a). Higher order semiparametric frequentist inference with the profile sampler. *The Annals of Statistics* 1786–1818. MR2435456

[13] Cheng, G. and Kosorok, M. R. (2008b). General frequentist properties of the posterior profile distribution. *The Annals of Statistics* **36** 1819–1853. MR2435457

[14] Cheng, G. and Kosorok, M. R. (2009). The penalized profile sampler. *Journal of multivariate analysis* **100** 345–362. MR2483424
[15] Clarke, B. and Ghosal, S. (2010). Posterior normality and reference priors for exponential families with increasing dimension. *Electronic Journal of Statistics* **4** 737–780. MR2678969
[16] Davidson, K. R. and Szarek, S. J. (2001). Local operator theory, random matrices and Banach spaces. *Handbook of the geometry of Banach spaces* **1** 317–366. MR1863696
[17] Gao, C. and Zhou, H. H. (2015). Rate-optimal posterior contraction for sparse PCA. *The Annals of Statistics* **43** 785–818. MR3325710
[18] Ghosal, S. (1997). Normal approximation to the posterior distribution for generalized linear models with many covariates. *Mathematical Methods of Statistics* **6** 332–348. MR1475901
[19] Ghosal, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli* **5** 315–331. MR1681701
[20] Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *Journal of Multivariate Analysis* **74** 49–68. MR1790613
[21] Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *The Annals of Statistics* **28** 500–531. MR1790007
[22] Hoffmann, M., Rousseau, J. and Schmidt-Hieber, J. (2013). On adaptive posterior concentration rates. *arXiv preprint arXiv:1305.5270*.
[23] Kato, T. (1995). *Perturbation theory for linear operators* **132**. Springer. MR1335452
[24] Laplace, P. (1774). Mémoire sur la probabilité des cause par les événments. *Eng. trans. in Statistic Sciences* **1** 359–378.
[25] Le Cam, L. and Yang, G. L. (2000). *Asymptotics in statistics: Some basic concepts*. Springer. MR1784901
[26] Ma, Z. (2013). Sparse principal component analysis and iterative thresholding. *The Annals of Statistics* **41** 772–801. MR3099121
[27] Rivoirard, V. and Rousseau, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *The Annals of Statistics* **40** 1489–1523. MR3015033
[28] van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge university press. MR1652247
[29] Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*. MR2963170
[30] Von Mises, R. (1928). *Wahrscheinlichkeit, statistik und wahrheit*. Berlin.