Supplementary Material for "Bayesian Sparse Spiked Covariance Model With a Continuous Matrix Shrinkage Prior"

Fangzheng Xie[*], Joshua Cape[†], Carey E. Priebe[‡], and Yanxun Xu[‡]

Abstract. This supplementary material contains the proofs of the theoretical results, additional technical results, and additional numerical results.

Contents

A Proof of Lemma 2.1 2

B Proofs of Results in Section 3.1 3

B.1 Proof of Lemma 3.1 3

B.2 Proof of Lemma 3.2 6

B.3 Proof of Lemma 3.3 8

C Proofs of Theorems 3.1, 3.2, and 3.3 12

C.1 Proof sketch of Theorem 3.1 and auxiliary results 12

C.2 Proofs of Theorems 3.1 and 3.2 18

C.3 Proof of Theorem 3.3 23

D Proofs of Lemma 3.4 and Theorem 3.4 24

D.1 An oracle testing lemma for the matrix infinity norm 24

D.2 Existence of tests under the matrix infinity norm 28

D.3 Proofs of Lemma 3.4 and Theorem 3.4 31

E Additional Numerical Results 37

E.1 MCMC convergence diagnostics 37

E.2 Hyperparameter selection and sensitivity 37

References 40

---

arXiv: 1808.07433

[*]Department of Statistics, Indiana University fxie@iu.edu

[†]Department of Statistics, University of Pittsburgh joshua.cape@pitt.edu

[‡]Department of Applied Mathematics and Statistics, Johns Hopkins University cep@jhu.edu yanxun.xu@jhu.edu

A Proof of Lemma 2.1

We will need the following CS matrix decomposition of a partitioned orthonormal matrix to prove Lemma 2.1.

Theorem A.1 (Theorem 5.1 in Stewart and Sun, 1990)**.** *Let the orthonormal matrix* $\mathbf{W} \in \mathbb{O}(p,p)$ *be partitioned in the form*

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix},$$

where $\mathbf{W}_{11} \in \mathbb{R}^{r\times r}$*,* $\mathbf{W}_{22} \in \mathbb{R}^{(p-r)\times(p-r)}$*, and* $2r \le p$*. Then there exists orthonormal matrices* $\mathbf{U} = \operatorname{diag}(\mathbf{U}_{11}, \mathbf{U}_{22})$ *and* $\mathbf{V} = \operatorname{diag}(\mathbf{V}_{11}, \mathbf{V}_{22})$ *with* $\mathbf{U}_{11}, \mathbf{V}_{11} \in \mathbb{O}(r)$*, such that*

$$\mathbf{W} = \mathbf{U} \begin{bmatrix} \mathbf{C} & -\mathbf{S} & \mathbf{0} \\ \mathbf{S} & \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{(p-2r)} \end{bmatrix} \mathbf{V}^{\mathrm{T}},$$

where $\mathbf{C} = \operatorname{diag}(c_1, \ldots, c_r)$ *and* $\mathbf{S} = \operatorname{diag}(s_1, \ldots, s_r)$ *are diagonal with non-negative entries, and* $\mathbf{C}^2 + \mathbf{S}^2 = \mathbf{I}_r$*.*

Proof of Lemma 2.1. Let $\mathbf{U}_\perp$ and $\mathbf{U}_{0\perp} \in \mathbb{O}(p, p-r)$ be such that $[\mathbf{U}, \mathbf{U}_\perp]$ and $[\mathbf{U}_0, \mathbf{U}_{0\perp}] \in \mathbb{O}(p)$. By the CS decomposition, there exists $\mathbf{U}_{11}, \mathbf{V}_{11} \in \mathbb{O}(r)$ and $\mathbf{U}_{22}, \mathbf{V}_{22} \in \mathbb{O}(p-r)$, such that

$$\begin{bmatrix} \mathbf{U}_0^{\mathrm{T}}\mathbf{U} & \mathbf{U}_0^{\mathrm{T}}\mathbf{U}_\perp \\ \mathbf{U}_{0\perp}^{\mathrm{T}}\mathbf{U} & \mathbf{U}_{0\perp}^{\mathrm{T}}\mathbf{U}_\perp \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{C} & -\mathbf{S} & \mathbf{0} \\ \mathbf{S} & \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{(p-2r)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{11}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{22}^{\mathrm{T}} \end{bmatrix}$$

where $\mathbf{C} = \operatorname{diag}(c_1, \ldots, c_r)$ and $\mathbf{S} = \operatorname{diag}(s_1, \ldots, s_r)$ are diagonal with non-negative entries, and $\mathbf{C}^2 + \mathbf{S}^2 = \mathbf{I}_r$. Write $\mathbf{U}_{22}$ into two blocks $\mathbf{U}_{22} = [\mathbf{U}_{221}, \mathbf{U}_{222}]$ with $\mathbf{U}_{221} \in \mathbb{O}(p-r, r)$. Take $\mathbf{Q} = [\mathbf{U}_0\mathbf{U}_{11}, \mathbf{U}_{0\perp}\mathbf{U}_{22}]$. Clearly, we have

$$\mathbf{Q}^{\mathrm{T}}\mathbf{U}_0\mathbf{U}_{11} = \begin{bmatrix} \mathbf{U}_{11}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{22}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_0^{\mathrm{T}} \\ \mathbf{U}_{0\perp}^{\mathrm{T}} \end{bmatrix} \mathbf{U}_0\mathbf{U}_{11} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{0}_r \\ \mathbf{0}_{p-2r} \end{bmatrix}$$

and

$$\mathbf{Q}^{\mathrm{T}}\mathbf{U}\mathbf{V}_{11} = \begin{bmatrix} \mathbf{U}_{11}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{22}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_0^{\mathrm{T}}\mathbf{U} \\ \mathbf{U}_{0\perp}^{\mathrm{T}}\mathbf{U} \end{bmatrix} \mathbf{V}_{11} = \begin{bmatrix} \mathbf{C} \\ \mathbf{S} \\ \mathbf{0}_{p-2r} \end{bmatrix}$$

Observe that $\|\mathbf{U}\mathbf{U}^{\mathrm{T}} - \mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\|_2 = \|\mathbf{S}\|_2$, and that $\mathbf{U}_0^{\mathrm{T}}\mathbf{U} = \mathbf{U}_{11}\mathbf{C}\mathbf{V}_{11}^{\mathrm{T}}$ is the singular value decomposition of $\mathbf{U}_0^{\mathrm{T}}\mathbf{U}$, implying that $\mathbf{W}_{\mathbf{U}} = \mathbf{U}_{11}\mathbf{V}_{11}^{\mathrm{T}}$. We proceed to compute

$$\begin{aligned} \|\mathbf{U} - \mathbf{U}_0\mathbf{W}_{\mathbf{U}}\|_{2\to\infty} &\le \|\mathbf{U} - \mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\mathbf{U}\|_{2\to\infty} + \|\mathbf{U}_0(\mathbf{U}_0^{\mathrm{T}}\mathbf{U} - \mathbf{W}_{\mathbf{U}})\|_{2\to\infty} \\ &= \left\| \mathbf{Q} \begin{bmatrix} \mathbf{S}^2 \\ -\mathbf{S}\mathbf{C} \\ \mathbf{0} \end{bmatrix} \right\|_{2\to\infty} + \left\| \mathbf{Q} \begin{bmatrix} \mathbf{C}(\mathbf{C} - \mathbf{I}_r) \\ -\mathbf{S}(\mathbf{C} - \mathbf{I}_r) \\ \mathbf{0} \end{bmatrix} \right\|_{2\to\infty} \end{aligned}$$

$$
\begin{aligned}
&= \left\| \begin{bmatrix} \mathbf{U}_0 & \mathbf{U}_{0\perp} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{221} & \mathbf{U}_{222} \end{bmatrix} \begin{bmatrix} \mathbf{S}^2 \\ -\mathbf{SC} \\ \mathbf{0} \end{bmatrix} \right\|_{2\to\infty} \\
&\quad + \left\| \begin{bmatrix} \mathbf{U}_0 & \mathbf{U}_{0\perp} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{221} & \mathbf{U}_{222} \end{bmatrix} \begin{bmatrix} \mathbf{C}(\mathbf{C}-\mathbf{I}_r) \\ -\mathbf{S}(\mathbf{C}-\mathbf{I}_r) \\ \mathbf{0} \end{bmatrix} \right\|_{2\to\infty} \\
&= \left\| \begin{bmatrix} \mathbf{U}_0 & \mathbf{U}_{0\perp} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{221} \end{bmatrix} \begin{bmatrix} \mathbf{S}^2 \\ -\mathbf{SC} \end{bmatrix} \right\|_{2\to\infty} \\
&\quad + \left\| \begin{bmatrix} \mathbf{U}_0 & \mathbf{U}_{0\perp} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{221} \end{bmatrix} \begin{bmatrix} \mathbf{C}(\mathbf{C}-\mathbf{I}_r) \\ -\mathbf{S}(\mathbf{C}-\mathbf{I}_r) \end{bmatrix} \right\|_{2\to\infty} \\
&= \left\| \begin{bmatrix} \mathbf{U}_0\mathbf{U}_{11} & \mathbf{U}_{0\perp}\mathbf{U}_{221} \end{bmatrix} \begin{bmatrix} \mathbf{S}^2 \\ -\mathbf{SC} \end{bmatrix} \right\|_{2\to\infty} \\
&\quad + \left\| \begin{bmatrix} \mathbf{U}_0\mathbf{U}_{11} & \mathbf{U}_{0\perp}\mathbf{U}_{221} \end{bmatrix} \begin{bmatrix} \mathbf{C}(\mathbf{C}-\mathbf{I}_r) \\ -\mathbf{S}(\mathbf{C}-\mathbf{I}_r) \end{bmatrix} \right\|_{2\to\infty}.
\end{aligned}
$$

Denote $\mathbf{V}_{\mathbf{U}} = [\mathbf{U}_0\mathbf{U}_{11}, \mathbf{U}_{0\perp}\mathbf{U}_{221}]$. Clearly, $\mathbf{V}_{\mathbf{U}} \in \mathbb{O}(p, 2r)$:

$$
\mathbf{V}_{\mathbf{U}}^{\mathrm{T}}\mathbf{V}_{\mathbf{U}} = \begin{bmatrix} \mathbf{U}_{11}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{221}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_0^{\mathrm{T}} \\ \mathbf{U}_{0\perp}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_0 & \mathbf{U}_{0\perp} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{221} \end{bmatrix} = \mathbf{I}_{2r}.
$$

Furthermore, by the previous derivation and the fact that $\|\mathbf{AB}\|_{2\to\infty} \le \|\mathbf{A}\|_{2\to\infty}\|\mathbf{B}\|_2$, we have

$$
\begin{aligned}
\|\mathbf{U} - \mathbf{U}_0\mathbf{W}_{\mathbf{U}}\|_{2\to\infty} &\le \|\mathbf{V}_{\mathbf{U}}\|_{2\to\infty} \left(\left\| \begin{bmatrix} \mathbf{S}^2 \\ -\mathbf{SC} \end{bmatrix} \right\|_2 + \left\| \begin{bmatrix} \mathbf{C}(\mathbf{C}-\mathbf{I}_r) \\ -\mathbf{S}(\mathbf{C}-\mathbf{I}_r) \end{bmatrix} \right\|_2 \right) \\
&= \|\mathbf{V}_{\mathbf{U}}\|_{2\to\infty} \left(\left\|\mathbf{S}^4 + \mathbf{SC}^2\mathbf{S}\right\|_2^{1/2} + \left\|(\mathbf{C}-\mathbf{I}_r)^2\right\|_2^{1/2} \right) \\
&= \|\mathbf{V}_{\mathbf{U}}\|_{2\to\infty} \left(\|\mathbf{S}\|_2 + \|\mathbf{I}_r - \mathbf{C}\|_2 \right) \\
&\le \|\mathbf{V}_{\mathbf{U}}\|_{2\to\infty} \left(\|\mathbf{S}\|_2 + \left\|\mathbf{I}_r - \mathbf{C}^2\right\|_2 \right) \\
&= \|\mathbf{V}_{\mathbf{U}}\|_{2\to\infty} \left(\left\|\mathbf{U}\mathbf{U}^{\mathrm{T}} - \mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\right\|_2 + \left\|\mathbf{U}\mathbf{U}^{\mathrm{T}} - \mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\right\|_2^2 \right),
\end{aligned}
$$

and the proof is thus completed. □

B Proofs of Results in Section 3.1

B.1 Proof of Lemma 3.1

We first introduce the group assignment indicators $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_p]^{\mathrm{T}} \in \{0,1\}^p$ for each row of $\mathbf{B}$, such that

$$
\begin{aligned}
(\mathbf{B}_{j*} \mid \xi_j = 0) &\sim \prod_{k=1}^{r} \psi_r(b_{jk} \mid \lambda + \lambda_0), \\
(\mathbf{B}_{j*} \mid \xi_j = 1) &\sim \prod_{k=1}^{r} \psi_1(b_{jk} \mid \lambda).
\end{aligned}
$$

Therefore, $\pi(b_{jk} \mid \xi_j = 1) = (\lambda/2)\mathrm{e}^{-\lambda|b_{jk}|}$ follows the Laplace distribution with scale parameter $1/\lambda$. Recall that the Laplace distribution can be alternatively represented as a normal-variance mixture distribution as follows:

$$(b_{jk} \mid \xi_j = 1, \phi_{jk}) \sim \mathrm{N}\left(0, \frac{\phi_{jk}}{\lambda^2}\right), \quad \text{and} \quad \phi_{jk} \sim \mathrm{Exp}(1/2).$$

On the other hand, by the prior construction $(|b_{jk}| \mid \xi_j = 0, \lambda_0) \sim \mathrm{Gamma}(1/r, \lambda_0 + \lambda)$, it follows that $(\|\mathbf{B}_{j*}\|_1 \mid \xi_j = 0, \lambda_0) \sim \mathrm{Exp}(\lambda_0 + \lambda)$. Denote $S_0 = \mathrm{supp}(\mathbf{B}_0)$. Now we construct the following event

$$\mathcal{B} = \bigcap_{j \in S_0} \{\xi_j = 1, 1 \le \phi_{jk} \le 2, k \in [r]\} \cap \bigcap_{j \in S_0^c} \{\xi_j = 0\} \cap \left\{\lambda_0 + \lambda \ge \frac{\sqrt{2p}}{\eta}\left(\log \frac{p}{s}\right)\right\}$$

and denote $\boldsymbol{\phi} = [\phi_{jk} : j \in S_0, k \in [r]]_{s \times r}$.

Step 1: Conditioning on the event $\mathcal{B}$. For any $(\boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0) \in \mathcal{B}$, we use a union bound to derive

$$\begin{aligned}
&\Pi\left(\|\mathbf{B} - \mathbf{B}_0\|_{\mathrm{F}} < \eta \;\middle|\; \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right) \\
&\quad \ge \Pi\left(\sum_{j \in S_0} \|\mathbf{B}_{j*} - \mathbf{B}_{0j*}\|_2^2 < \frac{\eta^2}{2} \;\middle|\; \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right) \prod_{j \in S_0^c} \Pi\left(\|\mathbf{B}_{j*}\|_1 \le \frac{\eta}{\sqrt{2p}} \;\middle|\; \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right) \\
&\quad \ge \Pi\left(\sum_{j \in S_0} \|\mathbf{B}_{j*} - \mathbf{B}_{0j*}\|_2^2 < \frac{\eta^2}{2} \;\middle|\; \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right) \prod_{j \in S_0^c} \left[1 - \exp\left\{-\frac{(\lambda_0 + \lambda)\eta}{\sqrt{2p}}\right\}\right] \\
&\quad \ge \Pi\left(\sum_{j \in S_0} \|\mathbf{B}_{j*} - \mathbf{B}_{0j*}\|_2^2 < \frac{\eta^2}{2} \;\middle|\; \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right) \left\{\left(1 - \frac{s}{p}\right)^{p/s}\right\}^s \\
&\quad \ge \Pi\left(\sum_{j \in S_0} \|\mathbf{B}_{j*} - \mathbf{B}_{0j*}\|_2^2 < \frac{\eta^2}{2} \;\middle|\; \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right) \exp\{-\log(2\mathrm{e})s\},
\end{aligned}$$

where the last inequality is due to the fact that $(1 - x)^{1/x} \ge \exp\{-\log(2\mathrm{e})\}$ when $x \in [0, 1/2]$. It then suffices to provide a lower bound for the first factor. We take advantage of the fact that $(b_{jk} \mid \xi_j = 1, \phi_{jk}) \sim \mathrm{N}(0, \phi_{jk}/\lambda^2)$ and apply Anderson's lemma (see, for example, Lemma 1.4 in the supporting document of Pati et al., 2014) together with the union bound to derive

$$\begin{aligned}
&\Pi\left(\sum_{j \in S_0} \|\mathbf{B}_{j*} - \mathbf{B}_{0j*}\|_2^2 < \frac{\eta^2}{2} \;\middle|\; \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right) \\
&\quad \ge \exp\left(-\frac{1}{2} \sum_{j \in S_0} \sum_{k=1}^r \frac{\lambda^2 b_{0jk}^2}{\phi_{jk}}\right) \Pi\left(\sum_{j \in S_0} \sum_{k=1}^r b_{jk}^2 < \frac{\eta^2}{2} \;\middle|\; \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right)
\end{aligned}$$

$$
\begin{aligned}
&\geq \exp\left(-\frac{1}{2}\sum_{j\in S_0}\sum_{k=1}^{r}\lambda^2 b_{0jk}^2\right)\prod_{j\in S_0}\prod_{k=1}^{r}\Pi\left(\frac{\lambda^2 b_{jk}^2}{\phi_{jk}}<\frac{\lambda^2\eta^2}{4rs}\,\middle|\,\boldsymbol{\phi},\boldsymbol{\xi},\lambda_0\right)\\
&\geq \exp\left(-\frac{\lambda^2}{2}\sum_{j\in S_0}\|\mathbf{B}_0\|_{2\to\infty}^2\right)\prod_{j\in S_0}\prod_{k=1}^{r}\left\{2\Phi\left(\frac{\lambda\eta}{2\sqrt{rs}}\right)-1\right\}\\
&\geq \exp\left(-\frac{\lambda^2}{2}s\|\mathbf{B}_0\|_{2\to\infty}^2 - sr - sr\left|\log\frac{\lambda\eta}{2\sqrt{rs}}\right|\right),
\end{aligned}
$$

where the fact that $\log\{2\Phi(x)-1\}\geq -1-\log(x)$ for small $x>0$ is applied in the last inequality.

Step 2: Control the prior probability of the event $\mathcal{B}$. Recall that

$$
\mathcal{B}=\bigcap_{j\in S_0}\{\xi_j=1, 1\leq\phi_{jk}\leq 2, k\in[r]\}\cap\bigcap_{j\in S_0^c}\{\xi_j=0\}\cap\left\{\lambda_0+\lambda\geq\frac{\sqrt{2p}}{\eta}\left(\log\frac{p}{s}\right)\right\}.
$$

Then conditioning on θ, we obtain by construction

$$
\begin{aligned}
\Pi(\mathcal{B}) &= \left\{\prod_{j\in S_0}\prod_{k=1}^{r}\Pi(1\leq\phi_{jk}\leq 2)\right\}\left\{\int_0^1\theta^s(1-\theta)^{p-s}\Pi(\mathrm{d}\theta)\right\}\\
&\quad\times\Pi\left\{\lambda_0+\lambda\geq\frac{\sqrt{2p}}{\eta}\left(\log\frac{p}{s}\right)\right\}\\
&\geq \exp(-3sr)\left\{\int_0^1\theta^s(1-\theta)^{p-s}\Pi(\mathrm{d}\theta)\right\}\Pi\left\{\lambda_0\geq\frac{\sqrt{2p}}{\eta}\left(\log\frac{p}{s}\right)\right\}.
\end{aligned}
$$

We first focus on the third factor. By assumption $\eta>1/p^\gamma$ for some $\gamma>0$, implying that

$$
\Pi\left\{\lambda_0>\frac{\sqrt{2p}}{\eta}\left(\log\frac{p}{s}\right)\right\}\geq\Pi(\lambda_0>p^\gamma)=1-\frac{1}{\Gamma(1/p^2)}\int_{1/p^\gamma}^{\infty}x^{1/p^2-1}\mathrm{e}^{-x}\mathrm{d}x.
$$

Using an inequality for the incomplete Gamma function (Alzer, 1997):

$$
\int_{4\delta}^{\infty}x^{-1}\mathrm{e}^{-x/2}\mathrm{d}x=\int_{2\delta}^{\infty}x^{-1}\mathrm{e}^{-x}\mathrm{d}x\leq\log\frac{1}{\delta}
$$

for small values of $\delta>0$, and the fact that $\Gamma(x)\geq 1$ when $0<x\leq 1$, we have:

$$
1-\frac{1}{\Gamma(1/p^2)}\int_{1/p^\gamma}^{\infty}x^{1/p^2-1}\mathrm{e}^{-x}\mathrm{d}x\geq 1-\int_{4/(4p^\gamma)}^{\infty}x^{-1}\mathrm{e}^{-x/2}\mathrm{d}x\geq 1-\log\frac{1}{4p^\gamma}\geq\mathrm{e}^{-1}
$$

for sufficiently large p (sufficiently small η). Next we consider the second factor. Write

$$
\int_0^1\theta^s(1-\theta)^{p-s}\Pi(\mathrm{d}\theta)
$$

$$
\begin{aligned}
&\geq \int_{s/p^{1+\kappa}}^{2s/p^{1+\kappa}} \exp\left\{-s\log\left(\frac{1-\theta}{\theta}\right) - p\log\left(\frac{1}{1-\theta}\right)\right\}\Pi(\mathrm{d}\theta)\\
&\geq \int_{s/p^{1+\kappa}}^{2s/p^{1+\kappa}} \exp\left\{-s\log\left(\frac{p^{1+\kappa}}{s}\right) - p\log\left(1+\frac{s}{p^{1+\kappa}-s}\right)\right\}\Pi(\mathrm{d}\theta)\\
&\geq \exp\left\{-(\kappa+1)s\log p - 2s\right\}\Pi\left(\frac{s}{p^{1+\kappa}} \leq \theta \leq \frac{2s}{p^{1+\kappa}}\right)
\end{aligned}
$$

for sufficiently large p. Observe that

$$
\begin{aligned}
\Pi\left(\frac{s}{p^{1+\kappa}} \leq \theta \leq \frac{2s}{p^{1+\kappa}}\right) &= \Pi\left(\frac{p^{1+\kappa}-2s}{p^{1+\kappa}} \leq 1-\theta \leq \frac{p^{1+\kappa}-s}{p^{1+\kappa}}\right)\\
&\geq \frac{1}{4p^{1+\kappa}}\left(1-\frac{2s}{p^{1+\kappa}}\right)^{4p^{1+\kappa}-1}\left(\frac{s}{p^{1+\kappa}}\right)\\
&\geq \frac{1}{4p^{1+\kappa}}\left\{\left(1-\frac{2s}{p^{1+\kappa}}\right)^{p^{1+\kappa}/(2s)}\right\}^{8s}\left(\frac{s}{p^{1+\kappa}}\right)\\
&\geq \exp\left\{-(2\kappa+19)s\log p\right\},
\end{aligned}
$$

we conclude that $\Pi(\mathcal{B}) \geq \exp\left\{-3sr - 1 - (3\kappa+22)s\log p\right\}$.

Step 3: Lower bound prior concentration by restricting over $\mathcal{B}$: We complete the proof by restricting over the event $\mathcal{B}$ as follows:

$$
\begin{aligned}
\Pi\left(\|\mathbf{B}-\mathbf{B}_0\|_{\mathrm{F}} < \eta\right) &\geq \mathbb{E}_\Pi\left\{\Pi\left(\|\mathbf{B}-\mathbf{B}_0\|_{\mathrm{F}} < \eta \,\middle|\, \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right)\mathbb{1}(\mathcal{B})\right\}\\
&\geq \left\{\inf_{(\boldsymbol{\phi},\boldsymbol{\xi},\lambda_0)\in\mathcal{B}} \Pi\left(\|\mathbf{B}-\mathbf{B}_0\|_{\mathrm{F}} < \eta \,\middle|\, \boldsymbol{\phi}, \boldsymbol{\xi}, \lambda_0\right)\right\}\Pi(\mathcal{B})\\
&\geq \exp\left[-C_1\max\left\{\lambda^2 s\|\mathbf{B}_0\|_{2\to\infty}^2, sr\left|\log\frac{\lambda\eta}{\sqrt{rs}}\right|, s\log p\right\}\right],
\end{aligned}
$$

where $C_1 > 0$ is some absolute constant.

B.2 Proof of Lemma 3.2

Recall that by construction, $(\|\mathbf{B}_{j*}\|_1 \mid \xi_j = 1) \sim \mathrm{Gamma}(r,\lambda)$ and $(\|\mathbf{B}_{j*}\|_1 \mid \xi_j = 0, \lambda_0) \sim \mathrm{Exp}(\lambda_0+\lambda)$, and $(\xi_j \mid \theta) \sim \mathrm{Bernoulli}(\theta)$ independently for each $j \in [p]$. Then with $\boldsymbol{\xi}$ integrated out, we have, independently for each $j \in [p]$,

$$
\pi(\|\mathbf{B}_{j*}\|_1 \mid \theta, \lambda_0) = (1-\theta)(\lambda_0+\lambda)\mathrm{e}^{-(\lambda_0+\lambda)\|\mathbf{B}_{j*}\|_1} + \theta\frac{\lambda^r}{\Gamma(r)}\|\mathbf{B}_{j*}\|_1^{r-1}\mathrm{e}^{-\lambda\|\mathbf{B}_{j*}\|_1}.
$$

Therefore, with λ_0 integrated out, for any $\delta > 1/p^\gamma$, we obtain

$$
\Pi(\|\mathbf{B}_{j*}\|_1 > \delta \mid \theta) \leq (1-\theta)\frac{1}{\Gamma(1/p^2)}\int_0^\infty \lambda_0^{-1/p^2-1}\mathrm{e}^{-1/\lambda_0}\mathrm{e}^{-(\lambda_0+\lambda)\delta}\mathrm{d}\lambda_0 + \theta
$$

$$\leq \frac{2}{\mathrm{e}^p}\int_0^\infty u^{1/p^2-1}\exp\left(-\frac{\delta}{u}-u\right)\mathrm{d}u+\theta,$$

where the last inequality is due to the change of variable $u = 1/\lambda_0$ and the fact that $\Gamma(1/p^2) \geq \mathrm{e}^p/2$ for sufficiently large p. Now we break down the integral in the preceding display as follows:

$$\int_0^\infty u^{1/p^2-1}\exp\left(-\frac{\delta}{u}-u\right)\mathrm{d}u \leq \int_0^{4\delta} u^{1/p^2-1}\exp\left(-\frac{\delta}{u}\right)\mathrm{d}u + \int_{4\delta}^\infty u^{1/p^2-1}\exp\left(-u\right)\mathrm{d}u.$$

For the first term, we observe that the function $u \mapsto (1/p^2-1)\log u - \delta/u$ achieves the maximum at $u = \delta/(1-p^{-2})$, and therefore, for sufficiently large p (small δ)

$$\int_0^{4\delta} u^{1/p^2-1}\exp\left(-\frac{\delta}{u}\right)\mathrm{d}u \leq 4\delta\exp\left\{\left(1-\frac{1}{\mathrm{e}^p}\right)\left(\log\frac{1-p^{-2}}{\delta}\right)\right\} \leq 4\delta^{1/p^2} \leq \log\frac{1}{\delta}.$$

For the second term, we apply the technique developed by Bhattacharya et al. (2015) to derive

$$\int_{4\delta}^\infty u^{1/p^2-1}\exp\left(-u\right)\mathrm{d}u \leq \int_{4\delta}^\infty u^{-1}\mathrm{e}^{-u/2}\mathrm{d}u \leq \log\frac{1}{\delta},$$

where the inequality for incomplete Gamma function due to Alzer (1997) is applied. Therefore, for any θ in the event $\{\theta < A_1 s\log p/p^{1+\kappa}\}$ for some constant A_1 to be determined later, we obtain

$$\Pi(\|\mathbf{B}_{j*}\|_1 > \delta \mid \theta) \leq \frac{4}{\mathrm{e}^p}\left(\log\frac{1}{\delta}\right) + \theta \leq \frac{4\gamma\log p + A_1 s\log p}{p^{1+\kappa}} \leq \frac{s\log p}{p}\left(\frac{A_1+4\gamma}{p^\kappa}\right).$$

A version of the Chernoff's inequality for binomial distributions states that (Hagerup and Rüb, 1990)

$$\mathbb{P}(X > ap) \leq \left\{\left(\frac{q}{a}\right)^a\exp(a)\right\}^p \quad \text{if } X \sim \text{Binomial}(p,q) \text{ and } q \leq a < 1.$$

Then over the event $\{\theta < A_1 s\log p/p^{1+\kappa}\}$, we have

$$\begin{aligned}\Pi(|\mathrm{supp}_\delta(\mathbf{B})| > \beta s \mid \theta) &\leq \exp\left[-\beta s\left\{\log\frac{\beta}{\mathrm{e}(A_1+4\gamma)\log p} + \kappa\log p\right\}\right] \\ &= \exp\left(-\frac{1}{2}\beta\kappa s\log p\right)\end{aligned}$$

by taking $A_1 = \beta/\mathrm{e} - 4\gamma$, $q = \Pi(\|\mathbf{B}_{j*}\|_1 > \delta \mid \theta) \leq (A_1+4\gamma)s\log p/p^{1+\kappa}$, and $a = \beta s/p$. Observe that for sufficiently small x, $(1-x)^{1/x} \leq \mathrm{e}^{-1/2}$. Then we integrate with respect to $\Pi(\mathrm{d}\theta)$ and proceed to compute

$$\begin{aligned}\Pi(|\mathrm{supp}_\delta(\mathbf{B})| > \beta s) &= \int_0^1 \Pi(|\mathrm{supp}_\delta(\mathbf{B})| > \beta s \mid \theta)\Pi(\mathrm{d}\theta) \\ &\leq \int_0^{A_1 s\log p/p^{1+\kappa}} \Pi(|\mathrm{supp}_\delta(\mathbf{B})| > \beta s \mid \theta)\Pi(\mathrm{d}\theta) + \Pi\left(\theta > \frac{A_1 s\log p}{p^{1+\kappa}}\right)\end{aligned}$$

$$
\begin{aligned}
&\leq \exp\left(-\frac{1}{2}\beta\kappa s\log p\right) + \left\{\left(1-\frac{A_1 s\log p}{p^{1+\kappa}}\right)^{p^{1+\kappa}/(A_1 s\log p)}\right\}^{A_1 s\log p} \\
&\leq \exp\left(-\frac{1}{2}\beta\kappa s\log p\right) + \exp\left(-\frac{A_1}{2}s\log p\right) \\
&\leq 2\exp\left\{-\min\left(\frac{\beta\kappa}{2},\frac{\beta}{2\mathrm{e}}-2\gamma\right)s\log p\right\},
\end{aligned}
$$

and the proof is thus completed.

B.3 Proof of Lemma 3.3

To prove Lemma 3.3, we need the following technical results regarding moments of Gamma mixture distributions.

Lemma B.1. *Suppose that w follows a mixture of an exponential distribution* $\mathrm{Exp}(\lambda_0)$ *and a Gamma distribution* $\mathrm{Gamma}(r,\lambda)$, *with mixing weights $1-\theta$ and θ, respectively. Let $\xi = \mathbb{1}(w>\delta)$, where δ is some sufficiently small constant such that $\Gamma(r)\leq 2\Gamma(r,\lambda\delta)$, and $\Gamma(r,\delta)=\int_\delta^\infty w^{r-1}\mathrm{e}^{-w}\mathrm{d}w$ is the (upper) incomplete Gamma function. Then the moments of w satisfy*

$$
\sup_{m\geq 1}\frac{1}{m}\left\{E(w^m\mid\xi=1)\right\}^{1/m}\leq 2\delta+\frac{2}{\lambda_0}+\frac{2(r+1)}{\lambda}
$$

and

$$
\sup_{m\geq 1}\frac{1}{m}\{E(w^m)\}^{1/m}\leq\frac{1}{\lambda_0}+\frac{r+1}{\lambda}.
$$

Furthermore, if $\theta\leq\mathrm{e}^{-r}$, then the moments of w satisfies

$$
\sup_{m\geq 1}\frac{1}{m}\{E(w^m)\}^{1/m}\leq\frac{1}{\lambda_0}+\frac{1}{\lambda}.
$$

Proof of Lemma B.1. Since $p(w)=(1-\theta)\lambda_0\mathrm{e}^{-\lambda_0 w}+\theta\{\lambda_0^r/\Gamma(r)\}w^{r-1}\mathrm{e}^{-\lambda w}$, then

$$
\begin{aligned}
\mathbb{P}(\xi=1) &= (1-\theta)\int_\delta^\infty\lambda_0\mathrm{e}^{-\lambda_0 w}\mathrm{d}w+\theta\int_\delta^\infty\frac{\lambda^r}{\Gamma(r)}w^{r-1}\mathrm{e}^{-\lambda w}\mathrm{d}w \\
&= (1-\theta)\mathrm{e}^{-\lambda_0\delta}+\theta\frac{\Gamma(r,\lambda\delta)}{\Gamma(r)}.
\end{aligned}
$$

Then for any measurable $A\subset\mathbb{R}$, we have

$$
\begin{aligned}
\mathbb{P}(w\in A\mid\xi=1) &= \frac{1}{\mathbb{P}(\xi=1)}(1-\theta)\int_A\mathbb{1}(w>\delta)\lambda_0\mathrm{e}^{-\lambda_0 w}\mathrm{d}w \\
&\quad+\frac{1}{\mathbb{P}(\xi=1)}\theta\int_A\mathbb{1}(w>\delta)\frac{\lambda^r}{\Gamma(r)}w^{r-1}\mathrm{e}^{-\lambda w}\mathrm{d}w \\
&= \int_A\mathbb{1}(w>\delta)\left\{(1-\theta')\lambda_0\mathrm{e}^{-\lambda_0(w-\delta)}\mathrm{d}w+\theta'\frac{\lambda^r}{\Gamma(r,\lambda\delta)}w^{r-1}\mathrm{e}^{-\lambda w}\right\}\mathrm{d}w,
\end{aligned}
$$

where

$$\theta' = \frac{\theta\Gamma(r,\lambda\delta)/\Gamma(r)}{(1-\theta)\mathrm{e}^{-\lambda_0\delta} + \theta\Gamma(r,\lambda\delta)/\Gamma(r)} \in (0,1).$$

Therefore,

$$p(w \mid \xi = 1) = \left\{(1-\theta')\lambda_0 \mathrm{e}^{-\lambda_0(w-\delta)}\mathrm{d}w + \theta' \frac{\lambda^r}{\Gamma(r,\lambda\delta)} w^{r-1}\mathrm{e}^{-\lambda w}\right\} \mathbb{1}(w > \delta).$$

Hence we proceed and compute

$$\begin{aligned}
&\{E(w^m \mid \xi = 1)\}^{1/m} \\
&\quad = \left\{(1-\theta') \int_\delta^\infty w^m \lambda_0 \mathrm{e}^{-\lambda_0(w-\delta)}\mathrm{d}w + \theta' \frac{\Gamma(r)}{\Gamma(r,\lambda\delta)} \int_\delta^\infty w^m \frac{\lambda^r}{\Gamma(r)} w^{r-1}\mathrm{e}^{-\lambda w}\mathrm{d}w\right\}^{1/m} \\
&\quad \le \left\{\int_0^\infty (w+\delta)^m \lambda_0 \mathrm{e}^{-\lambda_0 w}\mathrm{d}w + \frac{\Gamma(r)}{\Gamma(r,\lambda\delta)} \int_0^\infty w^m \frac{\lambda^r}{\Gamma(r)} w^{r-1}\mathrm{e}^{-\lambda w}\mathrm{d}w\right\}^{1/m} \\
&\quad = \left\{\int_0^\delta (w+\delta)^m \lambda_0 \mathrm{e}^{-\lambda_0 w}\mathrm{d}w + \int_\delta^\infty (w+\delta)^m \lambda_0 \mathrm{e}^{-\lambda_0 w}\mathrm{d}w + \frac{\Gamma(r)}{\Gamma(r,\lambda\delta)} \frac{(r+m-1)!}{(r-1)!\lambda^m}\right\}^{1/m} \\
&\quad \le \left\{\int_0^\infty (2\delta)^m \lambda_0 \mathrm{e}^{-\lambda_0 w}\mathrm{d}w + \int_0^\infty (2w)^m \lambda_0 \mathrm{e}^{-\lambda_0 w}\mathrm{d}w + \frac{\Gamma(r)}{\Gamma(r,\lambda\delta)} \frac{(r+m-1)!}{(r-1)!\lambda^m}\right\}^{1/m} \\
&\quad = \left\{(2\delta)^m + 2^m \frac{m!}{\lambda_0^m} + \frac{\Gamma(r)}{\Gamma(r,\lambda\delta)} \frac{(r+m-1)!}{(r-1)!\lambda^m}\right\}^{1/m} \le 2\delta + \frac{2m}{\lambda_0} + \frac{2(r+m)}{\lambda}.
\end{aligned}$$

Hence

$$\sup_{m\ge 1} \{E(w^m)\}^{1/m} \le \sup_{m\ge 1} \frac{1}{m}\left\{2\delta + \frac{2m}{\lambda_0} + \frac{2(r+m)}{\lambda}\right\} = 2\delta + \frac{2}{\lambda_0} + \frac{2(r+1)}{\lambda}.$$

Now we compute the sub-exponential norm. Write

$$\begin{aligned}
&\sup_{m\ge 1} \frac{1}{m}\{E(w^m)\}^{1/m} \\
&\quad = \sup_{m\ge 1} \frac{1}{m}\left\{(1-\theta)\int_0^\infty w^m \lambda_0 \mathrm{e}^{-\lambda_0 w}\mathrm{d}w + \theta \int_0^\infty w^m \frac{\lambda^r}{\Gamma(r)} w^{r-1}\mathrm{e}^{-\lambda w}\mathrm{d}w\right\}^{1/m} \\
&\quad = \sup_{m\ge 1} \frac{1}{m}\left\{(1-\theta)\frac{m!}{\lambda_0^m} + \theta \frac{(m+r-1)!}{\lambda^m (r-1)!}\right\}^{1/m} \\
&\quad \le \frac{1}{\lambda_0} + \frac{1}{\lambda}\sup_{m\ge 1} \theta^{1/m}\left(1 + \frac{r}{m}\right) \le \frac{1}{\lambda_0} + \frac{r+1}{\lambda}.
\end{aligned}$$

If $\theta \le \mathrm{e}^{-r}$, we can further derive the following result using the fact that $\log(1+ru) \le ru$ for $u \in (0,1]$:

$$\sup_{m\ge 1} \frac{1}{m}\{E(w^m)\}^{1/m} \le \frac{1}{\lambda_0} + \frac{1}{\lambda}\sup_{m\ge 1} \theta^{1/m}\left(1 + \frac{r}{m}\right)$$

$$\leq \frac{1}{\lambda_0} + \frac{1}{\lambda} \exp\left[\sup_{u\in(0,1]} \{-ru + \log(1+ru)\}\right] \leq \frac{1}{\lambda_0} + \frac{1}{\lambda}.$$

□

Proof of Lemma 3.3. Denote $S_0 = \mathrm{supp}(\mathbf{B}_0)$. We first use the union bound to derive

$$\begin{aligned}
&\Pi\left[\sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \mathbb{1}\{j \in \mathrm{supp}_{\delta_n}(\mathbf{B}) \cup \mathrm{supp}(\mathbf{B}_0)\} \geq t_n\right] \\
&\quad \leq \Pi\left\{\sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \mathbb{1}(\|\mathbf{B}_{j*}\|_1 > \delta_n) \geq t_n/2\right\} + \Pi\left(\sum_{j\in S_0} \|\mathbf{B}_{j*}\|_1 \geq t_n/2\right),
\end{aligned}$$

and then analyze the above two terms separately.

Step 1: Upper bounding the second term. Recall that

$$\pi(\|\mathbf{B}_{j*}\|_1 \mid \lambda_0, \theta) = (1-\theta)(\lambda_0+\lambda)\mathrm{e}^{-(\lambda_0+\lambda)\|\mathbf{B}_{j*}\|_1} + \theta\frac{\lambda^r}{\Gamma(r)}\|\mathbf{B}_{j*}\|^{r-1}\mathrm{e}^{-\lambda\|\mathbf{B}_{j*}\|_1}.$$

Denote $\beta' = \beta/\mathrm{e} - 4\gamma > 0$. Over the event $\{\theta \leq (\beta' s\log p)/p^{1+\kappa}\}$, it holds that

$$\mathbb{E}_\Pi(\|\mathbf{B}_{j*}\|_1 \mid \theta, \lambda_0) \leq \frac{1}{\lambda_0} + \frac{\beta' s\log p}{\lambda p^{1+\kappa}} \leq \frac{2}{\lambda}.$$

Since $(\beta' s\log p)/p^{1+\kappa} \leq 1/\sqrt{p} = \mathrm{e}^{-(\log p)/2} \leq \mathrm{e}^{-r}$ for sufficiently large n, we invoke Lemma B.1 to derive

$$\sup_{m\geq 1}\left\{\frac{1}{m}\mathbb{E}_\Pi\left(\|\mathbf{B}_{j*}\|_1^m \mid \theta, \lambda_0\right)\right\}^{1/m} \leq \frac{2}{\lambda}$$

over the event $\{\theta \leq (\beta' s\log p)/p^{1+\kappa}\}$, and proceed to apply the large deviation inequality for sub-exponential random variables (Proposition 5.16 in Vershynin, 2010) to obtain

$$\begin{aligned}
&\Pi\left(\sum_{j\in S_0}\|\mathbf{B}_{j*}\|_1 \geq t_n/2 \,\middle|\, \lambda_0, \theta\right) \\
&\quad \leq \Pi\left[\sum_{j\in S_0}\{\|\mathbf{B}_{j*}\|_1 - \mathbb{E}_\Pi(\|\mathbf{B}_{j*}\|_1 \mid \lambda_0, \theta)\} \geq t_n/2 - \frac{2s}{\lambda} \,\middle|\, \lambda_0, \theta\right] \\
&\quad \leq \Pi\left[\sum_{j\in S_0}\{\|\mathbf{B}_{j*}\|_1 - \mathbb{E}_\Pi(\|\mathbf{B}_{j*}\|_1 \mid \lambda_0, \theta)\} \geq t_n/4 \,\middle|\, \lambda_0, \theta\right] \\
&\quad \leq \exp\left\{-C\min\left(\frac{t_n^2}{s^2}, t_n\right)\right\}
\end{aligned}$$

for sufficiently large n, where C is some absolute constant. Observe that

$$\Pi\left(\theta > \frac{\beta' s \log p}{p^{1+\kappa}}\right) = \left\{\left(1 - \frac{\beta' s \log p}{p^{1+\kappa}}\right)^{p^{1+\kappa}/(\beta' s \log p)}\right\}^{\beta' s \log p}$$
$$\leq \exp\left\{-\left(\frac{\beta}{2\mathrm{e}} - 4\gamma\right) s \log p\right\}$$

since $(1-x)^{1/x} \leq \mathrm{e}^{-1/2}$ for $x \leq 1$, and so we obtain

$$\Pi\left(\sum_{j\in S_0} \|\mathbf{B}_{j*}\|_1 \geq t_n/2\right) \leq \mathbb{E}_\Pi\left[\Pi\left(\sum_{j\in S_0} \|\mathbf{B}_{j*}\|_1 \geq t_n/2 \,\middle|\, \lambda_0, \theta\right) \mathbb{1}\left(\theta \leq \frac{\beta' s \log p}{p^{1+\kappa}}\right)\right]$$
$$+ \Pi\left(\theta > \frac{\beta' s \log p}{p^{1+\kappa}}\right)$$
$$\leq \exp\left\{-C \min\left(\frac{t_n^2}{s^2}, t_n\right)\right\} + \exp\left\{-\left(\frac{\beta}{2\mathrm{e}} - 4\gamma\right) s \log p\right\}$$

for sufficiently large n.

Step 2: Upper bounding the first term. Denote $\zeta_j = \mathbb{1}(\|\mathbf{B}_{j*}\|_1 > \delta_n)$ and $\zeta = [\zeta_1, \ldots, \zeta_p]^{\mathrm{T}}$. By Lemma B.1 we obtain the following bound for the conditional expected value and moments of $\|\mathbf{B}_{j*}\|_1$ given $\zeta_j = 1$ and θ for sufficiently large n:

$$\mathbb{E}_\Pi(\|\mathbf{B}_{j*}\|_1 \mid \zeta_j = 1, \theta) \leq \sup_{m\geq 1} \{\mathbb{E}_\Pi(\|\mathbf{B}_{j*}\|_1^m \mid \zeta_j = 1, \theta)\}^{1/m} \leq 2\delta_n + \frac{2}{\lambda_0} + \frac{2(r+1)}{\lambda} \leq \frac{8r}{\lambda}.$$

Since $|\mathrm{supp}_{\delta_n}(\mathbf{B})| = \sum_{j=1}^p \zeta_j$, then over the event $\{\zeta : |\mathrm{supp}_{\delta_n}(\mathbf{B})| \leq \beta s\}$, we invoke the large deviation inequality for sub-exponential random variables again to derive

$$\Pi\left(\sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \zeta_j > t_n/2 \,\middle|\, \zeta, \theta\right)$$
$$\leq \Pi\left[\sum_{j\in\mathrm{supp}_{\delta_n}(\mathbf{B})} \{\|\mathbf{B}_{j*}\|_1 - \mathbb{E}_\Pi(\|\mathbf{B}_{j*}\|_1 \mid \zeta_j = 1)\} > \frac{t_n}{2} - \frac{8sr}{\lambda} \,\middle|\, \zeta, \theta\right]$$
$$\leq \Pi\left[\sum_{j\in\mathrm{supp}_{\delta_n}(\mathbf{B})} \{\|\mathbf{B}_{j*}\|_1 - \mathbb{E}_\Pi(\|\mathbf{B}_{j*}\|_1 \mid \zeta_j = 1)\} > \frac{t_n}{4} \,\middle|\, \zeta, \theta\right]$$
$$\leq \exp\left[-C \min\left\{\left(\frac{t_n}{\beta s r}\right)^2, \frac{t_n}{r}\right\}\right]$$

for sufficiently large n. Invoking Lemma 3.2, we obtain

$$\Pi\left(\sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \zeta_j > t_n/2\right)$$

$$
\begin{aligned}
&\leq \mathbb{E}_\Pi \left\{ \Pi \left(\sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \zeta_j > t_n/2 \;\Big|\; \zeta, \theta \right) \mathbb{1}(\zeta : |\mathrm{supp}_{\delta_n}(\mathbf{B})| \leq \beta s) \right\} \\
&\quad + \Pi \left(|\mathrm{supp}_{\delta_n}(\mathbf{B})| > \beta s \right) \\
&\leq \exp \left[-C \min \left\{ \left(\frac{t_n}{\beta s r} \right)^2, \frac{t_n}{r} \right\} \right] + 2 \exp \left\{ - \min \left(\frac{\beta \kappa}{2}, \frac{\beta}{2\mathrm{e}} - 2\gamma \right) s \log p \right\}.
\end{aligned}
$$

Step 3: Combining upper bounds: Combining the previous two upper bounds, we obtain

$$
\begin{aligned}
&\Pi \left[\sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \mathbb{1}\{ j \in \mathrm{supp}_{\delta_n}(\mathbf{B}) \cup \mathrm{supp}(\mathbf{B}_0) \} \geq t_n \right] \\
&\quad \leq 2 \exp \left[-C \min \left\{ \left(\frac{t_n}{\beta s r} \right)^2, \left(\frac{t_n}{s} \right)^2, \frac{t_n}{r} \right\} \right] + 3 \exp \left\{ - \min \left(\frac{\beta \kappa}{2}, \frac{\beta}{2\mathrm{e}} - 2\gamma \right) s \log p \right\},
\end{aligned}
$$

and the proof is completed. □

C Proofs of Theorems 3.1, 3.2, and 3.3

C.1 Proof sketch of Theorem 3.1 and auxiliary results

Now we sketch the proof of Theorem 3.1 along with some important auxiliary results. The proof strategy is based on a modification of the standard testing-and-prior-concentration approach, which was originally developed in Ghosal et al. (2000) for proving convergence rates of posterior distributions, and later adopted to a variety of statistical contexts. The posterior contraction for $\boldsymbol{\Sigma}$ with respect to the infinity norm loss can be proved in a similar fashion. Denote

$$
\mathcal{U}_n = \{ \boldsymbol{\Sigma} : \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\|_2 \leq M \epsilon_n \}
$$

and write the posterior distribution as

$$
\Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n) = \frac{\int_{\mathcal{U}_n^c} \exp\{\ell_n(\boldsymbol{\Sigma}) - \ell_n(\boldsymbol{\Sigma}_0)\} \Pi(\mathrm{d}\boldsymbol{\Sigma})}{\int \exp\{\ell_n(\boldsymbol{\Sigma}) - \ell_n(\boldsymbol{\Sigma}_0)\} \Pi(\mathrm{d}\boldsymbol{\Sigma})} = \frac{N_n(\mathcal{U}_n)}{D_n}, \tag{C.1}
$$

where $\ell_n(\boldsymbol{\Sigma})$ is the log-likelihood function of $\boldsymbol{\Sigma}$ given by

$$
\ell_n(\boldsymbol{\Sigma}) = \sum_{i=1}^n \log p(\mathbf{y}_i \mid \boldsymbol{\Sigma}) = \sum_{i=1}^n \left\{ -\frac{1}{2} \log \det(2\pi \boldsymbol{\Sigma}) - \frac{1}{2} \mathbf{y}_i^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{y}_i \right\}.
$$

To provide a useful upper bound for $\mathbb{E}_0\{\Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n)\}$ (*e.g.*, $\exp(-C_0 s \log p)$ appearing in Theorem 3.4), we modify the original testing-and-prior-concentration approach and require that the following three conditions hold:

(I): Prior concentration condition. The prior distribution provides sufficient concentration around the true $\mathbf{\Sigma}_0$: There exists some constant $C_3 > 0$ such that

$$\Pi\{\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}}^2 \leq (s \log p)/n\} \geq \exp(-C_3 s \log p)$$

for sufficient large n.

(II): Existence of Tests. There exists a sequence of subsets $(\mathcal{F}_n)_{n=1}^{\infty}$ of $\Theta(p, r, s)$, such that $\Pi(\mathbf{\Sigma} \in \mathcal{F}_n^c) \leq \exp(-C_4 s \log p)$ for some sufficiently large constant $C_4 > 0$, and there exists a sequence of test functions $(\phi_n)_{n=1}^{\infty}$, such that

$$\mathbb{E}_0(\phi_n) \lesssim \exp\left(-C_{41}\sqrt{M} n \epsilon_n^2\right),$$
$$\sup_{\mathbf{\Sigma} \in \mathcal{U}_n^c \cap \mathcal{F}_n} \mathbb{E}_{\mathbf{\Sigma}}(1 - \phi_n) \lesssim \exp(-C_{42} M n \epsilon_n^2)$$

for some constants $C_{41}, C_{42} > 0$.

The prior concentration condition can be verified by invoking Lemma 3.1. This condition is useful, as it guarantees that the denominator D_n appearing in the right-hand side of (C.1) can be lower bounded with high probability. The following lemma formalizes this result.

Lemma C.1. *Let $\mathcal{K}_n(\eta) = \{\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}} \leq \eta\}$ and $\eta < \sigma_0^2/2$. Then there exists some event $\mathcal{A}_n$ such that*

$$\mathcal{A}_n \subset \left\{D_n \geq \Pi_n\{\mathbf{\Sigma} \in \mathcal{K}_n(\eta)\} \exp\left[-\left\{\frac{C_3 \log \rho}{2(\lambda_{0r} + \sigma_0^2)} + 1\right\} n\eta^2\right]\right\}$$

for some absolute constant $C_3 > 0$, and

$$\mathbb{P}_0(\mathcal{A}_n^c) \leq 2 \exp\left\{-\tilde{C}_3 \min\left(\frac{n\eta^2}{\|\mathbf{\Sigma}_0^{-1}\|_2^2}, n\eta^2\right)\right\},$$

where $\rho = 2(\lambda_{01} + \sigma_0^2)/(\lambda_{0r} + \sigma_0^2)$ depends on the spectra of $\mathbf{\Sigma}$ only, and $\tilde{C}_3 > 0$ is an absolute constant.

Proof of Lemma C.1. To prove Lemma C.1, we need the following auxiliary matrix inequality:

Lemma C.2 (Pati et al., 2014, Supplement Lemma 1.3)**.** *Let $\mathbf{\Sigma}, \mathbf{\Sigma}_0$ be $p \times p$ positive definite matrices and $\eta \in (0, 1)$. If $\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}} \leq \eta$ and $\eta < 2\lambda_r(\mathbf{\Sigma}_0)$, then*

$$\log\det\left(\mathbf{\Sigma}_0 \mathbf{\Sigma}^{-1}\right) - \operatorname{tr}\left(\mathbf{\Sigma}_0 \mathbf{\Sigma}^{-1} - \mathbf{I}\right) \geq -C_3 \frac{\eta^2 \log \rho}{\lambda_r(\mathbf{\Sigma}_0)}$$

for some absolute constant $C_3 > 0$, where $\rho = 2\lambda_1(\mathbf{\Sigma}_0)/\lambda_r(\mathbf{\Sigma}_0)$.

Denote $\Pi\{\cdot \mid \mathcal{K}_n(\eta)\} = \Pi\{\cdot \cap \mathcal{K}_n(\eta)\}/\Pi_n\{\mathcal{K}_n(\eta)\}$ to be the re-normalized restriction of Π on $\mathcal{K}_n(\eta)$. Define random variable

$$\begin{aligned}
w_{ni} &= \int \log\frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)}\Pi\{\mathrm{d}\boldsymbol{\Sigma} \mid \mathcal{K}_n(\eta)\} \\
&= \int \left\{\frac{1}{2}\log\det(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1})\right\}\Pi\{\mathrm{d}\boldsymbol{\Sigma} \mid \mathcal{K}_n(\eta)\} \\
&\quad + \frac{1}{2}\mathbf{y}_i^{\mathrm{T}}\left[\int\left(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}^{-1}\right)\Pi\{\mathrm{d}\boldsymbol{\Sigma} \mid \mathcal{K}_n(\eta)\}\right]\mathbf{y}_i.
\end{aligned} \tag{C.2}$$

Invoking Fubini's theorem and Lemma C.2, we derive

$$\begin{aligned}
\mathbb{E}_0(w_{ni}) &= \int \left\{\frac{1}{2}\log\det(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1})\right\}\Pi\{\mathrm{d}\boldsymbol{\Sigma} \mid \mathcal{K}_n(\eta)\} \\
&\quad + \frac{1}{2}\int \mathbb{E}_0\left\{\mathbf{y}_i^{\mathrm{T}}\left(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}^{-1}\right)\mathbf{y}_i\right\}\Pi\{\mathrm{d}\boldsymbol{\Sigma} \mid \mathcal{K}_n(\eta)\} \\
&= \frac{1}{2}\int\left\{\log\det\left(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\right) + \mathrm{tr}\left(\mathbf{I} - \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\right)\right\}\Pi\{\mathrm{d}\boldsymbol{\Sigma} \mid \mathcal{K}_n(\eta)\} \\
&\geq -\frac{C_3\log\rho}{2(\lambda_{0r} + \sigma_0^2)}\eta^2.
\end{aligned}$$

Hence by Jensen's inequality,

$$\begin{aligned}
\log D_n - \log\Pi\{\boldsymbol{\Sigma} \in \mathcal{K}_n(\eta)\} &\geq \log\left[\int_{\mathcal{K}_n(\eta)} \exp\left\{\ell_n(\boldsymbol{\Sigma}) - \ell_n(\boldsymbol{\Sigma}_0)\right\}\frac{\Pi(\mathrm{d}\boldsymbol{\Sigma})}{\Pi\{\mathcal{K}_n(\eta)\}}\right] \\
&= \log\left[\int \exp\left\{\ell_n(\boldsymbol{\Sigma}) - \ell_n(\boldsymbol{\Sigma}_0)\right\}\Pi\{\mathrm{d}\boldsymbol{\Sigma} \mid \mathcal{K}_n(\eta)\}\right] \\
&\geq \int\{\ell_n(\boldsymbol{\Sigma}) - \ell_n(\boldsymbol{\Sigma}_0)\}\Pi\{\mathrm{d}\boldsymbol{\Sigma} \mid \mathcal{K}_n(\eta)\} \\
&= n\mathbb{E}_0(w_{ni}) + \sum_{i=1}^n\{w_{ni} - \mathbb{E}_0(w_{ni})\} \\
&\geq -\frac{C_3\log\rho}{2(\lambda_{0r} + \sigma_0^2)}n\eta^2 + \sum_{i=1}^n\{w_{ni} - \mathbb{E}_0(w_{ni})\}.
\end{aligned}$$

Now let $\mathcal{A}_n = \{|\sum_{i=1}^n\{w_{ni} - \mathbb{E}_0(w_{ni})\}| \leq n\eta^2\}$. Clearly,

$$\begin{aligned}
\mathcal{A}_n &\subset \left\{\log D_n - \log\Pi\{\boldsymbol{\Sigma} \in \mathcal{K}_n(\eta)\} \geq -\left\{\frac{C_3\log\rho}{2(\lambda_{0r} + \sigma_0^2)} + 1\right\}n\eta^2\right\} \\
&= \left\{D_n \geq \Pi\{\boldsymbol{\Sigma} \in \mathcal{K}_n(\eta)\}\exp\left[-\left\{\frac{C_3\log\rho}{2(\lambda_{0r} + \sigma_0^2)} + 1\right\}n\eta^2\right]\right\}.
\end{aligned}$$

We now analyze the probabilistic bound of $\mathcal{A}_n^c$. Recall $\boldsymbol{\Sigma}_0 = \mathbf{U}_0\boldsymbol{\Lambda}_0\mathbf{U}_0^{\mathrm{T}} + \sigma_0^2\mathbf{I}_p$. Let $\mathbf{U}_{0\perp}$ to be the orthonormal $(p-r)$-frame in $\mathbb{R}^p$ such that $[\mathbf{U}_0, \mathbf{U}_{0\perp}] \in \mathbb{O}(p)$, and denote

$$\boldsymbol{\Sigma}_0^{1/2} = [\mathbf{U}_0, \mathbf{U}_{0\perp}]\mathrm{diag}\{\lambda_1(\boldsymbol{\Sigma}_0)^{1/2}, \ldots, \lambda_p(\boldsymbol{\Sigma}_0)^{1/2}\}[\mathbf{U}_0, \mathbf{U}_{0\perp}]^{\mathrm{T}}.$$

Clearly, $\mathbf{\Sigma}_0 = (\mathbf{\Sigma}_0^{1/2})^2$, and by denoting $\mathbf{v}_i = \mathbf{\Sigma}_0^{-1/2}\mathbf{y}_i$, we have $\mathbf{v}_i \sim \mathrm{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ under $\mathbb{P}_0$. Re-writing $w_{ni} - \mathbb{E}_0(w_{ni})$ in terms of $\mathbf{v}_i$, we have

$$w_{ni} - \mathbb{E}_0(w_{ni}) = \mathbf{v}_i^{\mathrm{T}}\mathbf{\Omega}\mathbf{v}_i - \mathbb{E}_0(\mathbf{v}_i\mathbf{\Omega}\mathbf{v}_i),$$

where

$$\mathbf{\Omega} = \frac{1}{2}\int \left(\mathbf{I}_p - \mathbf{\Sigma}_0^{1/2}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0^{1/2}\right)\Pi\{\mathrm{d}\mathbf{\Sigma} \mid \mathcal{K}_n(\eta)\}.$$

Let $\mathbf{\Omega} = U_\Omega D_\Omega U_\Omega^{\mathrm{T}}$ be the spectral decomposition of $\mathbf{\Omega}$, and let $\mathbf{x}_i = \mathbf{U}_{\mathbf{\Omega}}^{\mathrm{T}}\mathbf{v}_i$. Then we proceed to bound

$$\begin{aligned}
\mathbb{P}_0(\mathcal{A}_n^c) &\le \mathbb{P}_0\left(\left|\sum_{i=1}^n \{w_{ni} - \mathbb{E}_0(w_{ni})\}\right| \ge n\eta^2\right)\\
&= \mathbb{P}_0\left(\left|\sum_{i=1}^n \left\{\mathbf{x}_i^{\mathrm{T}}D_\Omega\mathbf{x}_i - \mathbb{E}_0\left(\mathbf{x}_i^{\mathrm{T}}D_\Omega\mathbf{x}_i\right)\right\}\right| \ge n\eta^2\right)\\
&= \mathbb{P}_0\left(\left|\sum_{i=1}^n\sum_{j=1}^p \lambda_j(\Omega)\left\{x_{ij}^2 - \mathbb{E}_0(x_{ij}^2)\right\}\right| \ge n\eta^2\right)\\
&\le 2\exp\left[-C_3'\min\left\{\frac{n^2\eta^4}{n\sum_{j=1}^p\lambda_j(\Omega)^2}, \frac{n\eta^2}{\max_{j\in[p]}\lambda_j(\Omega)}\right\}\right]
\end{aligned}$$

for some absolute constant $C_3' > 0$, where the large deviation inequality for sub-exponential random variables is applied again in the last inequality. Observe that over $\mathcal{K}_n(\eta)$ for $\eta \le \sigma_0^2/2$,

$$\begin{aligned}
\|\mathbf{\Sigma}^{-1}\|_2 &\le \|\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}_0^{-1}\|_2 + \|\mathbf{\Sigma}_0^{-1}\|_2 = \|\mathbf{\Sigma}^{-1}(\mathbf{\Sigma} - \mathbf{\Sigma}_0)\mathbf{\Sigma}_0^{-1}\|_2 + \|\mathbf{\Sigma}_0^{-1}\|_2\\
&\le \|\mathbf{\Sigma}^{-1}\|_2\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}}\|\mathbf{\Sigma}_0^{-1}\|_2 + \|\mathbf{\Sigma}_0^{-1}\|_2 \le \frac{\eta}{\sigma_0^2}\|\mathbf{\Sigma}^{-1}\|_2 + \|\mathbf{\Sigma}_0^{-1}\|_2,
\end{aligned}$$

implying that $\|\mathbf{\Sigma}^{-1}\|_2 \le 2\|\mathbf{\Sigma}_0^{-1}\|_2$. Also observe that

$$\begin{aligned}
\sum_{j=1}^p \lambda_j(\mathbf{\Omega})^2 = \|\mathbf{\Omega}\|_{\mathrm{F}}^2 &\le \frac{1}{4}\int \left\|\mathbf{I}_p - \mathbf{\Sigma}_0^{1/2}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0^{1/2}\right\|_{\mathrm{F}}^2 \Pi\{\mathrm{d}\mathbf{\Sigma} \mid \mathcal{K}_n(\eta)\}\\
&= \frac{1}{4}\int \left\|\mathbf{I}_p - \mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0\right\|_{\mathrm{F}}^2 \Pi\{\mathrm{d}\mathbf{\Sigma} \mid \mathcal{K}_n(\eta)\}\\
&\le \frac{1}{4}\int \|\mathbf{\Sigma}^{-1}\|_2^2\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}}^2\Pi\{\mathrm{d}\mathbf{\Sigma} \mid \mathcal{K}_n(\eta)\}\\
&\le \|\mathbf{\Sigma}_0^{-1}\|_2^2\int \|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}}^2\Pi\{\mathrm{d}\mathbf{\Sigma} \mid \mathcal{K}_n(\eta)\} \le \|\mathbf{\Sigma}_0^{-1}\|_2^2\eta^2.
\end{aligned}$$

We finally obtain

$$\mathbb{P}_0(\mathcal{A}_n^c) \le 2\exp\left\{-\tilde{C}_3\min\left(\frac{n\eta^2}{\|\mathbf{\Sigma}_0^{-1}\|_2^2}, n\eta^2\right)\right\}$$

for some absolute constant $\tilde{C}_3 > 0$. ∎

Verifying the existence of tests is slightly more involved. The following lemma establishes the existence of required test functions.

Lemma C.3. *Assume the data* $\mathbf{y}_1,\dots,\mathbf{y}_n$ *follow* $\mathrm{N}_p(\mathbf{0}_p,\mathbf{\Sigma})$, $1\le r\le p$. *Suppose* $\mathbf{U}_0\in\mathbb{O}(p,r)$ *satisfies* $|\mathrm{supp}(\mathbf{U}_0)|\le s$, *and* $r\le s\le p$. *For any positive* δ, t, *and* τ, *define*

$$\mathcal{F}(\delta,\tau,t)=\left\{\mathbf{B}\in\mathbb{R}^{p\times r}:|\mathrm{supp}_\delta(\mathbf{B})|\le\tau,\sum_{j=1}^p\|\mathbf{B}_{j*}\|_2^2\mathbb{1}\{j\in\mathrm{supp}_\delta(\mathbf{B})\cup\mathrm{supp}(\mathbf{U}_0)\}\le t^2\right\}.$$

Let the positive sequences $(\delta_n,\tau_n,t_n,\epsilon_n)_{n=1}^\infty$ *satisfy* $(\sqrt{p}\delta_n+2t_n)\sqrt{p}\delta_n\le M_1\epsilon_n$ *for some constant* $M_1>0$, *and* $\epsilon_n\le 1$. *Consider testing*

$$H_0:\mathbf{\Sigma}=\mathbf{\Sigma}_0=\mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}}+\sigma_0^2\mathbf{I}_p$$

versus

$$H_1:\mathbf{\Sigma}\in\left\{\mathbf{\Sigma}=\mathbf{B}\mathbf{B}^{\mathrm{T}}+\sigma^2\mathbf{I}_p:\|\mathbf{\Sigma}-\mathbf{\Sigma}_0\|_2>M\epsilon_n,\mathbf{B}\in\mathcal{F}(\delta_n,\tau_n,t_n)\right\}.$$

Then for each $M\ge\max\{M_1/2,(128\|\mathbf{\Sigma}_0\|_2^4)^{1/3}\}$, *there exists a test function* $\phi_n:\mathbb{R}^{n\times p}\to[0,1]$, *such that*

$$\mathbb{E}_0(\phi_n)\le 3\exp\left\{(2+C_4)(\tau_n\log p+2s_n)-\frac{C_4\sqrt{M}}{\sqrt{2}}n\epsilon_n^2\right\},$$
$$\sup_{\mathbf{\Sigma}\in H_1}\mathbb{E}_{\mathbf{\Sigma}}(1-\phi_n)\le\exp\left\{C_4(\tau_n+2s_n)-\frac{C_4M}{8}n\epsilon_n^2\right\}$$

for some absolute constant $C_4>0$.

Proof of Lemma C.3. To proof Lemma C.3, we need the following oracle testing lemma from Gao and Zhou (2015):

Lemma C.4 (Gao and Zhou, 2015)**.** *Let* $\mathbf{y}_i\sim\mathrm{N}_d(\mathbf{0}_d,\mathbf{\Sigma})$, *where* $\mathbf{\Sigma}\in\mathbb{R}^{d\times d}$. *Then for any* $M>0$, *there exists a test function* ϕ_n *such that*

$$\mathbb{E}_{\mathbf{\Sigma}^{(1)}}(\phi_n)\le\exp\left(C_4d-\frac{C_4M^2}{4\|\mathbf{\Sigma}^{(1)}\|_2^2}n\epsilon^2\right)+2\exp\left(C_4d-C_4\sqrt{M}n\right),$$
$$\sup_{\|\mathbf{\Sigma}^{(2)}-\mathbf{\Sigma}^{(1)}\|_2>M\epsilon}\mathbb{E}_{\mathbf{\Sigma}^{(2)}}(1-\phi_n)\le\exp\left[C_4d-\frac{C_4Mn\epsilon^2}{4}\max\left\{1,\frac{M}{(\sqrt{M}+2)^2\|\mathbf{\Sigma}^{(1)}\|_2^2}\right\}\right].$$

with some absolute constant $C_4>0$.

Let $S_0=\mathrm{supp}(\mathbf{U}_0)$ and $S(\delta)=\mathrm{supp}_\delta(\mathbf{B})$. Then there exists some permutation matrix $\mathbf{P}$ such that

$$\mathbf{B}=\mathbf{P}\begin{bmatrix}\mathbf{B}_\delta\\ \mathbf{A}_\delta\end{bmatrix}\quad\text{and}\quad\mathbf{U}_0=\mathbf{P}\begin{bmatrix}\mathbf{U}_{0\delta}\\ \mathbf{0}\end{bmatrix},$$

where $\mathbf{B}_\delta$ and $\mathbf{U}_{0\delta}$ are $|S(\delta)\cup S_0|\times r$ matrices. Hence for $\mathbf{\Sigma}\in\mathcal{F}(\delta,\tau,t)$, it holds that

$$\begin{aligned}
\|\mathbf{\Sigma}-\mathbf{\Sigma}_0\|_2 &= \left\|\mathbf{P}\begin{bmatrix}\mathbf{B}_\delta\mathbf{B}_\delta^{\mathrm{T}}+\sigma^2\mathbf{I}-\mathbf{U}_{0\delta}\mathbf{\Lambda}_0\mathbf{U}_{0\delta}^{\mathrm{T}}-\sigma_0^2\mathbf{I} & \mathbf{B}_\delta\mathbf{A}_\delta^{\mathrm{T}}\\ \mathbf{A}_\delta\mathbf{B}_\delta^{\mathrm{T}} & \mathbf{A}_\delta\mathbf{A}_\delta^{\mathrm{T}}+(\sigma^2-\sigma_0^2)\mathbf{I}_d\end{bmatrix}\mathbf{P}^{\mathrm{T}}\right\|_2\\
&\le \left\|\begin{bmatrix}\mathbf{B}_\delta\mathbf{B}_\delta^{\mathrm{T}}+\sigma^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma^2\end{bmatrix}-\begin{bmatrix}\mathbf{U}_{0\delta}\mathbf{\Lambda}_0\mathbf{U}_{0\delta}^{\mathrm{T}}+\sigma_0^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_0^2\end{bmatrix}\right\|_2+\left\|\begin{bmatrix}\mathbf{0} & \mathbf{B}_\delta\mathbf{A}_\delta^{\mathrm{T}}\\ \mathbf{A}_\delta\mathbf{B}_\delta^{\mathrm{T}} & \mathbf{A}_\delta\mathbf{A}_\delta^{\mathrm{T}}\end{bmatrix}\right\|_{\mathrm{F}}\\
&\le \left\|\mathbf{\Sigma}_{S(\delta)}-\mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_2+(\|\mathbf{A}_\delta\|_2^2+2\|\mathbf{B}_\delta\|_2^2)^{1/2}\|\mathbf{A}_\delta\|_{\mathrm{F}}\\
&\le \left\|\mathbf{\Sigma}_{S(\delta)}-\mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_2+(\sqrt{p}\delta+2t)\sqrt{p}\delta,
\end{aligned}$$

where

$$\mathbf{\Sigma}_{S(\delta)}=\begin{bmatrix}\mathbf{B}_\delta\mathbf{B}_\delta^{\mathrm{T}}+\sigma^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma^2\end{bmatrix}\quad\text{and}\quad \mathbf{\Sigma}_{S(\delta)}^{(0)}=\begin{bmatrix}\mathbf{U}_{0\delta}\mathbf{\Lambda}_0\mathbf{U}_{0\delta}^{\mathrm{T}}+\sigma_0^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_0^2\end{bmatrix}.$$

By taking $M\ge 2M_1$, we obtain

$$\begin{aligned}
&\left\{\mathbf{\Sigma}=\mathbf{B}\mathbf{B}^{\mathrm{T}}+\sigma^2\mathbf{I}:\|\mathbf{\Sigma}-\mathbf{\Sigma}_0\|>M\epsilon_n,\mathbf{B}\in\mathcal{F}(\delta_n,\tau_n,t_n)\right\}\\
&\quad\subset\left\{\mathbf{\Sigma}:\left\|\mathbf{\Sigma}_{S(\delta)}-\mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_2>\frac{M}{2}\epsilon_n:\mathbf{B}\in\mathcal{F}(\delta_n,\tau_n,t_n)\right\}\\
&\quad\subset\bigcup_{S(\delta_n)\subset[p]:|S(\delta)|\le\tau_n}\left\{\mathbf{\Sigma}:\left\|\mathbf{\Sigma}_{S(\delta)}-\mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_2>\frac{M}{2}\epsilon_n\right\}.
\end{aligned}$$

Since both $\mathbf{\Sigma}_{S(\delta_n)}$ and $\mathbf{\Sigma}_{S(\delta_n)}^{(0)}$ are $(|S(\delta_n)\cup S_0|+1)\times(|S(\delta_n)\cup S_0|+1)$ square matrices, and

$$|S(\delta_n)\cup S_0|+1\le|S(\delta_n)|+S_0+1\le\tau_n+2s_n,$$

then for each $S(\delta_n)\subset[p]$ with $|S(\delta_n)|\le\tau_n$, and for each $M\ge\max\{M_1/2,(128\|\mathbf{\Sigma}_0\|_2^4)^{1/3}\}$, we invoke Lemma C.4 to construct a test $\phi_{S(\delta_n)}$ depending on the index set $S(\delta_n)$, such that the type I error probability satisfies

$$\begin{aligned}
\mathbb{E}_{\mathbf{\Sigma}_{S(\delta_n)}^{(0)}}\left(\phi_{S(\delta_n)}\right)&\le\exp\left\{C_4(\tau_n+2s_n)-\frac{C_4M^2n\epsilon_n^2}{16\|\mathbf{\Sigma}_{S(\delta_n)}^{(0)}\|_2^2}\right\}\\
&\quad+2\exp\left\{C_4(\tau_n+2s_n)-C_4\sqrt{\frac{M}{2}}n\right\}\\
&\le 3\exp\left\{C_4(\tau_n+2s_n)-C_4\min\left(\frac{M^2}{16\|\mathbf{\Sigma}_0\|_2^2},\sqrt{\frac{M}{2}}\right)n\epsilon_n^2\right\}\\
&\le 3\exp\left\{C_4(\tau_n+2s_n)-C_4\sqrt{\frac{M}{2}}n\epsilon_n^2\right\},
\end{aligned}$$

and for all $\mathbf{\Sigma}_{S(\delta_n)}\in\{\|\mathbf{\Sigma}_{S(\delta_n)}-\mathbf{\Sigma}_{S(\delta_n)}^{(0)}\|_2>M\epsilon_n/2\}$, the type II error probability satisfies

$$\mathbb{E}_{\mathbf{\Sigma}_{S(\delta_n)}^{(1)}}\left(1-\phi_{S(\delta_n)}\right)\le\exp\left[C_4(\tau_n+2s_n)-\frac{C_4Mn\epsilon_n^2}{8}\max\left\{1,\frac{M}{(\sqrt{M}+2)^2\|\mathbf{\Sigma}_{S(\delta_n)}^{(0)}\|_2^2}\right\}\right]$$

$$\leq \exp\left\{C_4(\tau_n + 2s_n) - \frac{C_4 M n\epsilon_n^2}{8}\right\}.$$

Notice that for each index set $S(\delta_n)$, the test function $\phi_{S(\delta_n)}$ is only a function of $\mathbf{Y}_n$ through the coordinates $[y_{ij} : i \in [n], j \in S(\delta_n) \cup S_0]$. Hence, $\mathbb{E}_{\mathbf{\Sigma}^{(0)}_{S(\delta_n)}}(\phi_{S(\delta_n)}) = \mathbb{E}_0(\phi_{S(\delta_n)})$, and for any $p \times p$ covariance matrix $\mathbf{\Sigma}$ with $\|\mathbf{\Sigma}_{S(\delta_n)} - \mathbf{\Sigma}^{(0)}_{S(\delta_n)}\|_2 > M\epsilon_n/2$, it holds that $\mathbb{E}_{\mathbf{\Sigma}_{S(\delta_n)}}(1 - \phi_{S(\delta_n)}) = \mathbb{E}_{\mathbf{\Sigma}}(1 - \phi_{S(\delta_n)})$. Therefore, by aggregating the test functions

$$\phi_n = \max_{S(\delta_n) \subset [p] : |S(\delta_n)| \leq \tau_n} \phi_{S(\delta_n)},$$

we obtain

$$\begin{aligned}
\mathbb{E}_0(\phi_n) &\leq \sum_{S(\delta_n) \subset [p] : |S(\delta_n)| \leq \tau_n} \mathbb{E}_{\mathbf{\Sigma}^{(0)}_{S(\delta_n)}}(\phi_{S(\delta_n)}) \\
&\leq 3 \sum_{q=0}^{\lfloor \tau_n \rfloor} \frac{p!}{q!(p-q)!} \exp\left\{C_4(\tau_n + 2s_n) - C_4\sqrt{\frac{M}{2}} n\epsilon_n^2\right\} \\
&\leq 3(\tau_n + 1)\exp(\tau_n \log p)\exp\left\{C_4(\tau_n + 2s_n) - C_4\sqrt{\frac{M}{2}} n\epsilon_n^2\right\} \\
&\leq 3\exp\left\{\tau_n + \tau_n \log p + C_4(\tau_n + 2s_n) - C_4\sqrt{\frac{M}{2}} n\epsilon_n^2\right\} \\
&\leq 3\exp\left\{(2 + C_4)(\tau_n \log p + 2s_n) - C_4\sqrt{\frac{M}{2}} n\epsilon_n^2\right\},
\end{aligned}$$

and

$$\begin{aligned}
\sup_{\mathbf{\Sigma} \in H_1} \mathbb{E}_{\mathbf{\Sigma}}(1 - \phi_n) &\leq \sup_{S(\delta_n) \subset [p] : |S(\delta_n)| \leq \tau_n} \sup_{\left\{\mathbf{\Sigma} : \|\mathbf{\Sigma}_{S(\delta_n)} - \mathbf{\Sigma}^{(0)}_{S(\delta_n)}\|_2 > M\epsilon_n/2\right\}} \mathbb{E}_{\mathbf{\Sigma}_{S(\delta_n)}}\left(1 - \phi_{S(\delta_n)}\right) \\
&\leq \exp\left\{C_4(\tau_n + 2s_n) - \frac{C_4 M}{8} n\epsilon_n^2\right\}.
\end{aligned}$$

The proof is thus completed. □

C.2 Proofs of Theorems 3.1 and 3.2

Recall that $\mathcal{U}_n = \{\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_2 \leq M\epsilon_n\}$ and the posterior probability $\Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n)$ can be written as $\Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n) = N_n(\mathcal{U}_n^c)/D_n$, where

$$N_n(\mathcal{U}_n^c) = \int_{\mathcal{A}} \exp\{\ell_n(\mathbf{\Sigma}) - \ell_n(\mathbf{\Sigma}_0)\}\Pi(\mathrm{d}\mathbf{\Sigma}), \quad D_n = \int \exp\{\ell_n(\mathbf{\Sigma}) - \ell_n(\mathbf{\Sigma}_0)\}\Pi(\mathrm{d}\mathbf{\Sigma}),$$

and $\ell_n(\mathbf{\Sigma}) = \sum_{i=1}^n \log p(\mathbf{y}_i \mid \mathbf{\Sigma})$ is the log-likelihood function of $\mathbf{\Sigma}$.

Step 1: Prior concentration. Let $\eta_n = \sqrt{(s \log p)/n}$. Then by Lemma C.1, there exists a sequence of events $(\mathcal{A}_n)_{n=1}^{\infty}$ such that

$$\mathcal{A}_n \subset \{D_n \geq \Pi(\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}} \leq \eta_n) \exp\left(-C_3' s \log p\right)\}$$

for $\eta_n = \sqrt{(s \log p)/n} \leq \sigma_0^2/2$, and

$$\mathbb{P}_0(\mathcal{A}_n^c) \leq 2 \exp\left\{-\tilde{C}_3 \min\left(1, \|\mathbf{\Sigma}_0^{-1}\|_2^{-2}\right) s \log p\right\}, \tag{C.3}$$

where C_3' and $\tilde{C}_3$ are some absolute constants. Denote $\mathbf{B}_0 = \mathbf{U}_0 \mathbf{\Lambda}_0^{1/2}$, where $\mathbf{\Lambda}_0^{1/2} = \operatorname{diag}(\lambda_{01}^{1/2}, \dots, \lambda_{0r}^{1/2})$. Then we analyze the prior concentration using a union bound as follows:

$$\begin{aligned}\Pi(\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}} \leq \eta_n) &\geq \Pi\left(\|\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{B}_0\mathbf{B}_0^{\mathrm{T}}\|_{\mathrm{F}} + \|\sigma^2 \mathbf{I}_p - \sigma_0^2 \mathbf{I}_p\|_{\mathrm{F}} \leq \eta_n\right) \\ &\geq \Pi\left(\|\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{B}_0\mathbf{B}_0^{\mathrm{T}}\|_{\mathrm{F}} \leq \frac{\eta_n}{2}\right) \Pi\left(|\sigma_0^2 - \sigma^2| \leq \frac{\eta_n}{2\sqrt{p}}\right).\end{aligned}$$

On one hand, for $\eta_n = \sqrt{(s \log p)/n} \leq \sigma_0^2/2$, we have

$$\Pi\left(|\sigma_0^2 - \sigma^2| \leq \frac{\eta_n}{2\sqrt{p}}\right) \geq \left\{\min_{\sigma \in [\sigma_0^2/2, 3\sigma_0^2/2]} \pi_\sigma(\sigma^2)\right\} \frac{\eta_n}{\sqrt{p}} \geq C(\sigma_0^2) \mathrm{e}^{-\log p},$$

where the constant $C(\sigma_0^2) = \min_{\sigma_0^2/2 \leq \sigma^2 \leq 3\sigma_0^2/2} \pi_\sigma(\sigma^2) > 0$ depends only on σ_0^2. On the other hand, for $\eta_n = \sqrt{(s \log p)/n} \leq \min(\sigma_0^2/2, 16\|\mathbf{B}_0\|_2^{1/2})$, we proceed by union bound to derive

$$\begin{aligned}\Pi\left(\|\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{B}_0\mathbf{B}_0^{\mathrm{T}}\|_{\mathrm{F}} \leq \frac{\eta_n}{2}\right) &\geq \Pi\left(\|\mathbf{B} - \mathbf{B}_0\|_{\mathrm{F}} \|\mathbf{B} - \mathbf{B}_0 + \mathbf{B}_0\|_2 + \|\mathbf{B}_0\|_2 \|\mathbf{B} - \mathbf{B}_0\|_{\mathrm{F}} \leq \frac{\eta_n}{2}\right) \\ &\geq \Pi\left\{\|\mathbf{B} - \mathbf{B}_0\|_{\mathrm{F}} \left(\|\mathbf{B} - \mathbf{B}_0\|_{\mathrm{F}} + 2\|\mathbf{B}_0\|_2\right) \leq \frac{\eta_n}{2}\right\} \\ &\geq \Pi\left\{\|\mathbf{B} - \mathbf{B}_0\|_{\mathrm{F}} \leq \min\left(\frac{\eta_n}{8\|\mathbf{B}_0\|_2}, 2\|\mathbf{B}_0\|_2\right)\right\} \\ &= \Pi\left(\|\mathbf{B} - \mathbf{B}_0\|_{\mathrm{F}} \leq \frac{\eta_n}{8\|\mathbf{B}_0\|_2}\right).\end{aligned}$$

Invoking Lemma 3.1, we see that there exists some constant $C(\lambda, \mathbf{B}_0)$ depending on λ and $\|\mathbf{B}_0\|_{2\to\infty}$ only, such that

$$\begin{aligned}\Pi\left(\|\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{B}_0\mathbf{B}_0^{\mathrm{T}}\|_{\mathrm{F}} \leq \frac{\eta_n}{2}\right) &\geq \Pi\left(\|\mathbf{B} - \mathbf{B}_0\|_{\mathrm{F}} \leq \frac{\eta_n}{8\|\mathbf{B}_0\|_2}\right) \\ &\geq \exp\left[-C_1 \max\left\{\lambda^2 s \|\mathbf{B}_0\|_{2\to\infty}^2, s \log p, sr \left|\log\left(\lambda \frac{\sqrt{\log p}}{\sqrt{rn}}\right)\right|\right\}\right] \\ &\geq \exp\left\{-C(\lambda, \mathbf{B}_0) s \log p\right\}.\end{aligned}$$

Therefore, for $\eta_n = \sqrt{(s \log p)/n} \leq \min(\sigma_0^2, 16\|\mathbf{B}_0\|_2^{1/2})$ we obtain

$$\Pi(\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_{\mathrm{F}} \leq \eta_n) \geq C(\sigma_0^2) \exp\left[-\{1 + C(\lambda, \mathbf{B}_0)\} s \log p\right],$$

and over $\mathcal{A}_n$, we have

$$D_n \geq C(\sigma_0^2)\exp\left(-C_{0\lambda}s\log p\right) \tag{C.4}$$

for some constant $C_{0\lambda}$ depending only on λ and $\|\mathbf{B}_0\|_{2\to\infty}$.

Step 2: Construct subsets $(\mathcal{F}_n)_{n=1}^{\infty}$. Take $\epsilon_n = \sqrt{(s\log p)/n}$, $\tau_n = \beta s_n$, $t_n = (sr\log p)^2$, and $\delta_n = \epsilon_n/(t_n\sqrt{p})$, where $\beta > 0$ is some constant to be specified later. Clearly, there exists some $\gamma > 0$ such that

$$\delta_n = \frac{\epsilon_n}{t_n\sqrt{p}} = \frac{\sqrt{s\log p}}{\sqrt{np}(sr\log p)^2} = \frac{1}{\sqrt{nps^3r^4(\log p)^3}} \geq \frac{1}{p^\gamma}.$$

Now let $\beta > 4\mathrm{e}\gamma$ and $\mathcal{F}_n = \mathcal{F}(\delta_n, \tau_n, t_n)$ be defined in Lemma C.3. Since

$$\begin{aligned}
\min\left\{\left(\frac{t_n}{\beta sr}\right)^2, \left(\frac{t_n}{r}\right)^2, \frac{t_n}{r}\right\} &= \min\left\{\frac{(sr)^2(\log p)^4}{\beta^2}, s^4r^2(\log p)^4, s^2r(\log p)\right\}\\
&= \min\left\{\frac{sr^2(\log p)^3}{\beta^2}, s^3r^2(\log p)^3, sr\log p\right\}s\log p\\
&\geq \beta s\log p
\end{aligned}$$

for sufficiently large n, and $t_n/(sr) = (sr)\log p \to \infty$, we then can invoke Lemmas 3.2 and 3.3 to obtain

$$\begin{aligned}
\Pi(\mathcal{F}_n^c) &\leq \Pi(|\mathrm{supp}_{\delta_n}(\mathbf{B})| > \beta s_n) + \Pi\left[\sum_{j=1}^{p}\|\mathbf{B}_{j*}\|_2^2\mathbb{1}\{j \in \mathrm{supp}_{\delta_n}(\mathbf{B})\cup\mathrm{supp}(\mathbf{U}_0)\} > t_n^2\right]\\
&\leq 2\exp(-\beta s\log p) + 5\exp\left\{-\min\left(\frac{\beta\kappa}{2}, \frac{\beta}{2\mathrm{e}} - 2\gamma\right)s\log p\right\}\\
&\leq 7\exp\left\{-\min\left(\frac{\beta\kappa}{2}, \frac{\beta}{2\mathrm{e}} - 2\gamma\right)s\log p\right\}
\end{aligned} \tag{C.5}$$

for sufficiently large n (and hence sufficiently small $s\log p/p$).

Step 3: Decompose the integral $\mathbb{E}_0\{\Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n)\}$. Since by construction we have

$$(\sqrt{p}\delta_n + 2t_n)\sqrt{p}\delta_n = \left(\sqrt{p}\frac{\epsilon_n}{t_n\sqrt{p}} + 2t_n\right)\sqrt{p}\delta_n \leq 3t_n\sqrt{p}\delta_n = 3\epsilon_n.$$

Then by Lemma C.3, for each $M \geq \max\{3/2, (128\|\mathbf{\Sigma}_0\|_2^4)^{1/3}\}$, there exists a test function ϕ_n such that

$$\mathbb{E}_0(\phi_n) \leq 3\exp\left[-\left\{\frac{C_4\sqrt{M}}{\sqrt{2}} - (2 + C_4)(\beta + 2)\right\}s\log p\right], \tag{C.6}$$

$$\sup_{\mathbf{\Sigma}\in\mathcal{U}_n^c\cap\mathcal{F}_n}\mathbb{E}_{\mathbf{\Sigma}}(1 - \phi_n) \leq \exp\left[-\left\{\frac{C_4M}{8} - C_4(\beta + 2)\right\}s\log p\right] \tag{C.7}$$

for some absolute constant $C_4 > 0$ for sufficiently large n. Now we decompose the target integral $\mathbb{E}_0\{\Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n)\}$ using (C.3) and (C.6) as follows:

$$\begin{aligned}\mathbb{E}_0\{\Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n)\} &\leq \mathbb{E}_0(\phi_n) + \mathbb{E}_0\left\{(1-\phi_n)\Pi(\mathcal{U}_n \mid \mathbf{Y}_n)\mathbb{1}(\mathcal{A}_n)\right\} + \mathbb{P}_0(\mathcal{A}_n^c)\\ &\leq 3\exp\left[-\left\{\frac{C_4\sqrt{M}}{\sqrt{2}} - (2+C_4)(\beta+2)\right\} s\log p\right]\\ &\quad + 2\exp\left\{-\tilde{C}_3\min\left(1, \|\boldsymbol{\Sigma}_0^{-1}\|_2^{-2}\right) s\log p\right\}\\ &\quad + \mathbb{E}_0\left[(1-\phi_n)\left\{\frac{N_n(\mathcal{U}_n^c)}{D_n}\right\}\mathbb{1}(\mathcal{A}_n)\right].\end{aligned}$$

Now we focus on the third term on the right-hand side of the preceding display. By (C.4), we obtain

$$\begin{aligned}&\mathbb{E}_0\left[(1-\phi_n)\left\{\frac{N_n(\mathcal{U}_n^c)}{D_n}\right\}\mathbb{1}(\mathcal{A}_n)\right]\\ &\quad\leq \frac{\exp\left(C_{0\lambda}s\log p\right)}{C(\sigma_0^2)}\mathbb{E}_0\left\{(1-\phi_n)\int_{\mathcal{U}_n^c}\prod_{i=1}^n\frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)}\Pi(\mathrm{d}\boldsymbol{\Sigma})\right\}.\end{aligned}$$

Observe that by Fubini's theorem,

$$\begin{aligned}&\mathbb{E}_0\left\{(1-\phi_n)\int_{\mathcal{U}_n^c}\prod_{i=1}^n\frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)}\Pi(\mathrm{d}\boldsymbol{\Sigma})\right\}\\ &\leq \mathbb{E}_0\left\{(1-\phi_n)\int_{\mathcal{U}_n^c\cap\mathcal{F}_n}\prod_{i=1}^n\frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)}\Pi(\mathrm{d}\boldsymbol{\Sigma})\right\} + \mathbb{E}_0\left\{\int_{\mathcal{F}_n^c}\prod_{i=1}^n\frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)}\Pi(\mathrm{d}\boldsymbol{\Sigma})\right\}\\ &= \int_{\mathcal{U}_n^c\cap\mathcal{F}_n}\mathbb{E}_0\left\{(1-\phi_n)\prod_{i=1}^n\frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)}\right\}\Pi(\mathrm{d}\boldsymbol{\Sigma}) + \int_{\mathcal{F}_n^c}\left\{\mathbb{E}_0\prod_{i=1}^n\frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)}\right\}\Pi(\mathrm{d}\boldsymbol{\Sigma})\\ &\leq \int_{\mathcal{U}_n^c\cap\mathcal{F}_n}\mathbb{E}_{\boldsymbol{\Sigma}}(1-\phi_n)\Pi(\mathrm{d}\boldsymbol{\Sigma}) + \Pi(\mathcal{F}_n^c)\\ &\leq \exp\left[-\left\{\frac{C_4M}{8} - C_4(\beta+2)\right\}s\log p\right] + 7\exp\left\{-\min\left(\frac{\beta\kappa}{2}, \frac{\beta}{2\mathrm{e}} - 2\gamma\right)s\log p\right\},\end{aligned}$$

where the testing type II error probability bound (C.7) and (C.5) are applied to the last inequality. Then by taking

$$\begin{aligned}\beta &= \max\left\{\frac{4}{\kappa}C_{0\lambda}, 2\mathrm{e}\left(2\gamma + 2C_{0\lambda}\right)\right\},\\ M = M_0 &= \max\left[\frac{8}{C_4}\left\{C_4(\beta+2) + 2C_{0\lambda}\right\}, \frac{2}{C_4^2}\left\{C_{0\lambda} + (2+C_4)(\beta+2)\right\}^2\right],\end{aligned}$$

we obtain the following result:

$$\mathbb{E}_0\left[(1-\phi_n)\left\{\frac{N_n(\mathcal{U}_n^c)}{D_n}\right\}\mathbb{1}(\mathcal{A}_n)\right]$$

$$
\begin{aligned}
&\leq \frac{1}{C(\sigma_0^2)} \exp\left[-\left\{\frac{C_4 M}{8} - C_4(\beta+2) - C_{0\lambda}\right\} s\log p\right] \\
&\quad + \frac{1}{C(\sigma_0^2)} 7\exp\left[-\left\{\min\left(\frac{\beta\kappa}{2}, \frac{\beta}{2\mathrm{e}} - 2\gamma\right) - C_{0\lambda}\right\} s\log p\right] \\
&\leq \frac{8}{C(\sigma_0^2)} \exp\left\{-C_{0\lambda} s\log p\right\}.
\end{aligned}
$$

Combining the above results, we finally obtain

$$
\begin{aligned}
\mathbb{E}_0\{\Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n)\} &\leq \left\{3 + \frac{8}{C(\sigma_0^2)}\right\} \exp\left\{-C_{0\lambda} s\log p\right\} \\
&\quad + 2\exp\left\{-\tilde{C}_3 \min\left(1, \|\mathbf{\Sigma}_0^{-1}\|_2^{-2}\right) s\log p\right\} \\
&\leq \left\{5 + \frac{11}{C(\sigma_0^2)}\right\} \exp\left[-\min\left\{C_{0\lambda}, \tilde{C}_3, \tilde{C}_3\|\mathbf{\Sigma}_0^{-1}\|_2^{-2}\right\} s\log p\right] \\
&= R_0 \exp(-C_0 s\log p)
\end{aligned}
$$

by taking $C_0 = \min\left\{C_{0\lambda}, \tilde{C}_3, \tilde{C}_3\|\mathbf{\Sigma}_0^{-1}\|_2^{-2}\right\}$ and $R_0 = \left\{5 + 11/C(\sigma_0^2)\right\}$. Therefore, there exists some constant M_0, such that for all sufficiently large n, we have

$$
\begin{aligned}
\mathbb{E}_0\left\{\Pi\left(\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_2 > M\epsilon_n \,\middle|\, \mathbf{Y}_n\right)\right\} &\leq \mathbb{E}_0\left\{\Pi\left(\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_2 > M_0\epsilon_n \,\middle|\, \mathbf{Y}_n\right)\right\} \\
&\leq R_0 \exp(-C_0 s\log p)
\end{aligned}
$$

for some absolute constants C_0 and R_0 depending on $\mathbf{\Sigma}_0$ and the hyperparameters only.

Step 4: Bounding the projection spectral norm loss using the sine-theta theorem. To prove the posterior contraction for $\mathbf{U}$ with respect to the projection spectral norm loss (3.2), we need the following version of the Davis-Kahan sine-theta theorem, which follows as a recasting of Theorem VII.3.7 in Bhatia (1997) in the language of Yu et al. (2015):

Theorem C.1. *Let $\mathbf{X}$, $\widehat{\mathbf{X}} \in \mathbb{R}^{p\times p}$ be symmetric matrices with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ and $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_p$, respectively. Write $\mathbf{E} = \widehat{\mathbf{X}} - \mathbf{X}$ and fix $1 \leq r \leq s \leq p$. Assume that $\delta_{\mathrm{gap}} := \min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}) > 0$ where $\lambda_0 := \infty$ and $\lambda_{p+1} := -\infty$. Let $d = s - r + 1$ and let $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_s] \in \mathbb{R}^{p\times d}$ and $\widehat{\mathbf{V}} = [\hat{\mathbf{v}}_r, \ldots, \hat{\mathbf{v}}_s] \in \mathbb{R}^{p\times d}$ have orthonormal columns satisfying $\mathbf{X}\mathbf{v}_j = \lambda_j \mathbf{v}_j$ and $\widehat{\mathbf{X}}\hat{\mathbf{v}}_j = \hat{\lambda}_j \hat{\mathbf{v}}_j$ for $j = r, r+1, \ldots, s$. Then*

$$
\|\widehat{\mathbf{V}}\widehat{\mathbf{V}}^{\mathrm{T}} - \mathbf{V}\mathbf{V}^{\mathrm{T}}\|_2 \leq \frac{2}{\delta_{\mathrm{gap}}}\|\mathbf{E}\|_2.
$$

To apply the sine-theta theorem, we let $\mathbf{X} = \mathbf{\Sigma}_0 = \mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}} + \sigma_0^2\mathbf{I}_p$, $\widehat{\mathbf{X}} = \mathbf{B}\mathbf{B}^{\mathrm{T}} + \sigma^2\mathbf{I}_p$, and take "$s$"$= r$ and "r"$= 1$, in which case $\delta_{\mathrm{gap}} = \min\{\infty, \lambda_r(\mathbf{\Sigma}_0) - \lambda_{r+1}(\mathbf{\Sigma}_0)\} = \lambda_{0r}$, $\mathbf{V} = \mathbf{U}_0$, $\widehat{\mathbf{V}} = \mathbf{U}_{\mathbf{B}}$, and $\mathbf{E} = \mathbf{\Sigma} - \mathbf{\Sigma}_0$. Then by the sine-theta theorem and (3.1), we have

$$
\|\mathbf{U}_{\mathbf{B}}\mathbf{U}_{\mathbf{B}}^{\mathrm{T}} - \mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\|_2 \leq \frac{2}{\lambda_{0r}}\|\mathbf{E}\|_2 = \frac{2}{\lambda_{0r}}\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_2
$$

and hence, by the posterior contraction for $\mathbf{\Sigma}$, we have

$$\mathbb{E}_0\left\{\Pi\left(\|\mathbf{U_B}\mathbf{U_B}^{\mathrm{T}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\|_2>\frac{2M\epsilon_n}{\lambda_{0r}}\;\Big|\;\mathbf{Y}_n\right)\right\}\leq\mathbb{E}_0\left\{\Pi\left(\|\mathbf{\Sigma}-\mathbf{\Sigma}_0\|_2>M\epsilon_n\;\Big|\;\mathbf{Y}_n\right)\right\}$$
$$\leq R_0\exp(-C_0 s\log p).$$

C.3 Proof of Theorem 3.3

For any random matrix $\mathbf{X}\in\mathbb{R}^{p\times p}$, we have

$$\|\mathbf{E}(\mathbf{X})\|_2^2=\max_{\|\mathbf{u}\|_2=1}\{\mathbb{E}(\mathbf{Xu})\}^{\mathrm{T}}\{\mathbb{E}(\mathbf{Xu})\}\leq\mathbb{E}\|\mathbf{X}\|_2^2$$

by the Jensen's inequality. Now take $\mathbf{X}=\mathbf{U_B}\mathbf{U_B}^{\mathrm{T}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}$. Denote the event $\mathcal{U}_n=\{\|\mathbf{U_B}\mathbf{U_B}^{\mathrm{T}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\|_2\leq M_0\epsilon_n\}$. Invoking the posterior contraction (3.2), we have

$$\begin{aligned}
\mathbb{E}_0\left(\left\|\widehat{\mathbf{\Omega}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\right\|_2^2\right)&=\mathbb{E}_0\left\{\left\|\int\left(\mathbf{U_B}\mathbf{U_B}^{\mathrm{T}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\right)\Pi(\mathrm{d}\mathbf{B}\mid\mathbf{Y}_n)\right\|_2^2\right\}\\
&\leq\mathbb{E}_0\left\{\int_{\mathcal{U}_n}\left\|\left(\mathbf{U_B}\mathbf{U_B}^{\mathrm{T}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\right)\right\|_2^2\Pi(\mathrm{d}\mathbf{B}\mid\mathbf{Y}_n)\right\}\\
&\quad+\mathbb{E}_0\left\{\int_{\mathcal{U}_n^c}\left\|\left(\mathbf{U_B}\mathbf{U_B}^{\mathrm{T}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\right)\right\|_2^2\Pi(\mathrm{d}\mathbf{B}\mid\mathbf{Y}_n)\right\}\\
&\leq M_0^2\epsilon_n^2+\left(\sup_{\mathbf{U}\in\mathbb{O}(p,r)}\|\mathbf{U}\mathbf{U}^{\mathrm{T}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\|_2^2\right)\mathbb{E}_0\left\{\Pi(\mathcal{U}_n^c\mid\mathbf{Y}_n)\right\}\\
&\leq\frac{4M_0^2}{\lambda_{0r}^2}\epsilon_n^2+4R_0\exp(-C_0 s\log p).
\end{aligned}$$

Since for sufficiently large n, we have

$$\epsilon_n^2=\frac{s\log p}{n}=\exp\left(\log s+\log\log p-\log n\right)\geq\exp(-C_0 s\log p),$$

we obtain

$$\mathbb{E}_0\left(\left\|\widehat{\mathbf{\Omega}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\right\|_2\right)\leq\left\{\mathbb{E}_0\left(\left\|\widehat{\mathbf{\Omega}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\right\|_2^2\right)\right\}^{1/2}\leq\epsilon_n\left(\frac{2M_0}{\lambda_{0r}}+2\sqrt{R_0}\right).$$

Since the columns of $\widehat{\mathbf{U}}$ are the leading r-eigenvectors of $\widehat{\mathbf{\Omega}}$ corresponding to $\lambda_1(\widehat{\mathbf{\Omega}}),\ldots,\lambda_r(\widehat{\mathbf{\Omega}})$, *i.e.*, $\widehat{\mathbf{\Omega}}\widehat{\mathbf{U}}_{*k}=\lambda_k(\widehat{\mathbf{\Omega}})\widehat{\mathbf{U}}_{*k}$, then applying the sine-theta theorem (Theorem C.1) yields

$$\mathbb{E}_0\left(\|\widehat{\mathbf{U}}\widehat{\mathbf{U}}^{\mathrm{T}}-\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\|_2\right)\leq\left(\frac{4M_0}{\lambda_{0r}}+4\sqrt{R_0}\right)\epsilon_n.$$

D Proofs of Lemma 3.4 and Theorem 3.4

D.1 An oracle testing lemma for the matrix infinity norm

We first present an oracle testing lemma for the matrix infinity norm that is useful to derive the posterior contraction rate under the matrix infinity norm loss.

Lemma D.1. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \sim \mathrm{N}_d(\mathbf{0}_d, \boldsymbol{\Sigma})$ independently, where $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. Let $\epsilon \in (0,1)$. Then there exists some absolute constant $C_6 > 0$, such that for each M satisfying $M \geq \max[4, \{(2\log 2)/C_6\}^2]$, and $M\epsilon \leq \min(1, 2\|\boldsymbol{\Sigma}_0\|_2)$, there exists a test function $\phi_n : \mathbb{R}^{n\times d} \to [0,1]$, such that*

$$\mathbb{E}_0(\phi_n) \leq 4\exp\left(4d - \frac{C_6 M^2 n\epsilon^2}{4\|\boldsymbol{\Sigma}_0\|_\infty^2}\right) + 8\exp\left(4d - \frac{C_6\sqrt{M}n}{2}\right)$$

$$\sup_{\{\|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0\|_\infty > M\epsilon\}} \mathbb{E}_{\boldsymbol{\Sigma}}(1-\phi_n) \leq 4\exp\left\{4d - \frac{C_6 M n\epsilon^2}{4}\min\left(1, \frac{1}{4\|\boldsymbol{\Sigma}_0\|_\infty^2}\right)\right\}.$$

The proof relies on the following matrix concentration inequality for the sample covariance matrix with regard to the matrix infinity norm.

Lemma D.2. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \sim \mathrm{N}_d(\mathbf{0}_d, \boldsymbol{\Sigma})$ independently, where $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. Then there exists an absolute constant $C_6 > 0$, such that for any $t > 0$,*

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \boldsymbol{\Sigma}\right\|_\infty > t\|\boldsymbol{\Sigma}\|_\infty\right) \leq 4\exp\{4d - C_6 n\min(t, t^2)\}$$

Proof of Lemma D.2. By definition,

$$\|\mathbf{A}\|_\infty = \sup_{\|\mathbf{v}\|_\infty = 1}\|\mathbf{A}\mathbf{v}\|_\infty = \max_{j\in[p]}\sup_{\|\mathbf{v}\|_\infty=1}\mathbf{e}_j^{\mathrm{T}}\mathbf{A}\mathbf{v},$$

where $\mathbf{e}_j$ is the unit vector along the jth coordinate direction. Now let $S_\infty^{d-1}(1/2)$ be an $1/2$-net of the ℓ_∞-sphere in $\mathbb{R}^d$ ($\{\mathbf{v}\in\mathbb{R}^d : \|\mathbf{v}\|_\infty = 1\}$) with minimum cardinality. Then for each $\mathbf{v}$ with $\|\mathbf{v}\|_\infty = 1$, there exists some $\mathbf{v}' \in S_\infty^{d-1}(1/2)$ such that $\|\mathbf{v}-\mathbf{v}'\|_\infty < 1/2$. Therefore,

$$\begin{aligned}
\|\mathbf{A}\|_\infty &= \max_{j\in[d]}\sup_{\|\mathbf{v}\|_\infty=1}\mathbf{e}_j^{\mathrm{T}}\mathbf{A}\mathbf{v} \leq \max_{j\in[d]}\sup_{\|\mathbf{v}\|_\infty=1}\left\{\mathbf{e}_j^{\mathrm{T}}\mathbf{A}(\mathbf{v}-\mathbf{v}') + \mathbf{e}_j^{\mathrm{T}}\mathbf{A}\mathbf{v}'\right\} \\
&\leq \max_{j\in[d]}\sup_{\|\mathbf{v}\|_\infty=1}\mathbf{e}_j^{\mathrm{T}}\mathbf{A}(\mathbf{v}-\mathbf{v}') + \max_{j\in[d]}\sup_{\mathbf{v}\in S_\infty^{d-1}(1/2)}\mathbf{e}_j^{\mathrm{T}}\mathbf{A}\mathbf{v} \\
&\leq \frac{1}{2}\|\mathbf{A}\|_\infty + \max_{j\in[d]}\sup_{\mathbf{v}\in S_\infty^{d-1}(1/2)}\mathbf{e}_j^{\mathrm{T}}\mathbf{A}\mathbf{v},
\end{aligned}$$

and hence,

$$\|\mathbf{A}\|_\infty \leq 2\max_{j\in[d]}\sup_{\mathbf{v}\in S_\infty^{d-1}(1/2)}\mathbf{e}_j^{\mathrm{T}}\mathbf{A}\mathbf{v}.$$

Denote

$$\mathbf{E} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \boldsymbol{\Sigma}.$$

Now we apply the union bound to derive

$$\begin{aligned}
&\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \boldsymbol{\Sigma}\right\|_\infty > t\|\boldsymbol{\Sigma}\|_\infty\right) \\
&\quad= \mathbb{P}\left[\bigcup_{j\in[d]} \bigcup_{\mathbf{v}\in S_\infty^{d-1}(1/2)} \left\{\mathbf{e}_j^{\mathrm{T}}\mathbf{E}\mathbf{v} > \frac{t}{2}\|\boldsymbol{\Sigma}\|_\infty\right\}\right] \\
&\quad\le \sum_{j=1}^{d} \sum_{\mathbf{v}\in S_\infty^{d-1}(1/2)} \mathbb{P}\left\{\mathbf{e}_j^{\mathrm{T}}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \boldsymbol{\Sigma}\right)\mathbf{v} > \frac{t}{2}\|\boldsymbol{\Sigma}\|_\infty\right\} \\
&\quad= \sum_{j=1}^{d} \sum_{\mathbf{v}\in S_\infty^{d-1}(1/2)} \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}(\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i)(\mathbf{v}^{\mathrm{T}}\mathbf{x}_i) - \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v} > \frac{t}{2}\|\boldsymbol{\Sigma}\|_\infty\right\}.
\end{aligned}$$

Observe that

$$\begin{bmatrix}\mathbf{v}^{\mathrm{T}}\mathbf{x}_i \\ \mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i\end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v} & \mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{e}_j \\ \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v} & \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{e}_j\end{bmatrix}\right),$$

then we can decompose $(\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i)(\mathbf{v}^{\mathrm{T}}\mathbf{x}_i)$ by projecting $\mathbf{v}^{\mathrm{T}}\mathbf{x}_i$ onto the space spanned by $\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i$ as follows:

$$\begin{aligned}
(\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i)(\mathbf{v}^{\mathrm{T}}\mathbf{x}_i) &= \left(\mathbf{v}^{\mathrm{T}}\mathbf{x}_i - \frac{\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v}}{\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{e}_j}\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i\right)(\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i) + \frac{\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v}}{\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{e}_j}(\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i)^2 \\
&\stackrel{d}{=} \sqrt{\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{e}_j\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v} - (\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v})^2}\,\zeta_{i1}\zeta_{i2} + \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v}\zeta_{i2}^2,
\end{aligned}$$

where ζ_{i1} and ζ_{i2} are independent $\mathrm{N}(0,1)$ random variables, $i = 1, \ldots, n$, and $\stackrel{d}{=}$ indicates the equality in distribution. Hence,

$$\begin{aligned}
&\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}(\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i)(\mathbf{v}^{\mathrm{T}}\mathbf{x}_i) - \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v} > \frac{t}{2}\|\boldsymbol{\Sigma}\|_\infty\right\} \\
&\quad\le \mathbb{P}\left\{\sqrt{\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{e}_j\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v} - (\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v})^2}\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_{i1}\zeta_{i2}\right| + |\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v}|\left|\frac{1}{n}\sum_{i=1}^{n}\left(\zeta_{i2}^2 - 1\right)\right| > \frac{t}{2}\|\boldsymbol{\Sigma}\|_\infty\right\} \\
&\quad\le \mathbb{P}\left\{\|\boldsymbol{\Sigma}\|_\infty\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_{i1}\zeta_{i2}\right| + \|\boldsymbol{\Sigma}\|_\infty\left|\frac{1}{n}\sum_{i=1}^{n}\left(\zeta_{i2}^2 - 1\right)\right| > \frac{t}{2}\|\boldsymbol{\Sigma}\|_\infty\right\} \\
&\quad\le \mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_{i1}\zeta_{i2}\right| > \frac{t}{4}\right\} + \mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(\zeta_{i2}^2 - 1\right)\right| > \frac{t}{4}\right\}.
\end{aligned}$$

Since $\zeta_{i1}\zeta_{i2}$ and $\zeta_{i2}^2 - 1$ are mean-zero sub-exponential random variables, it follows from the large-deviation inequality for sub-exponential random variables that

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n \zeta_{i1}\zeta_{i2}\right| > \frac{t}{4}\right\} + \mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n (\zeta_{i2}^2 - 1)\right| > \frac{t}{4}\right\} \leq 4\exp\left\{-C_6 n\min(t, t^2)\right\}$$

for some absolute constant $C_6 > 0$. It suffices to bound $|S_\infty^{d-1}(1/2)|$. For any metric space $(\mathcal{X}, d)$ where $d(\cdot,\cdot) : \mathcal{X}\times\mathcal{X} \to [0, +\infty)$ is the distance function, let $\mathcal{N}(\epsilon, \mathcal{X}, d)$ be the covering number of $\mathcal{X}$ with radii ϵ, namely, the smallest numbers of the balls of the form $B_d(x, \epsilon) := \{y \in \mathcal{X} : d(x, y) < \epsilon\}$ that are needed to cover $\mathcal{X}$. Since

$$\begin{aligned}
|S_\infty^{d-1}(1/2)| &= \mathcal{N}(1/2, \{\mathbf{v}\in\mathbb{R}^d : \|\mathbf{v}\|_\infty = 1\}, \|\cdot\|_\infty)\\
&\leq \mathcal{N}(1/2, \{\mathbf{v}\in\mathbb{R}^d : \|\mathbf{v}\|_\infty \leq 1\}, \|\cdot\|_\infty) \leq 6^d,
\end{aligned}$$

it follows that

$$\begin{aligned}
&\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma}\right\|_\infty > t\|\mathbf{\Sigma}\|_\infty\right)\\
&\quad\leq \sum_{j=1}^d \sum_{\mathbf{v}\in S_\infty^{d-1}(1/2)} \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^n (\mathbf{e}_j^{\mathrm{T}}\mathbf{x}_i)(\mathbf{v}^{\mathrm{T}}\mathbf{x}_i) - \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}\mathbf{v} > \frac{t}{2}\|\mathbf{\Sigma}\|_\infty\right\}\\
&\quad\leq 4d\exp\{d\log 6 - C_6 n\min(t, t^2)\} \leq 4\exp\{4d - C_6 n\min(t, t^2)\},
\end{aligned}$$

and the proof is thus completed. □

Proof of Lemma D.1. Denote the alternative set by $\mathcal{H}_1 = \{\mathbf{\Sigma} : \|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_\infty > M\epsilon\}$ and decompose it as follows: $\mathcal{H}_1 = \bigcup_{j=0}^\infty \mathcal{H}_{1j}$, where

$$\begin{aligned}
\mathcal{H}_{10} &= \left\{\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_\infty > M\epsilon, \|\mathbf{\Sigma}\|_\infty \leq (\sqrt{M} + 2)\|\mathbf{\Sigma}_0\|_\infty\right\}\\
\mathcal{H}_{1j} &= \left\{(\sqrt{M} + 2)(M\epsilon^2)^{-(j-1)/2}\|\mathbf{\Sigma}_0\|_\infty < \|\mathbf{\Sigma}\|_\infty \leq (\sqrt{M} + 2)(M\epsilon^2)^{-j/2}\|\mathbf{\Sigma}_0\|_\infty\right\}.
\end{aligned}$$

For each $\mathcal{H}_{1j}$, we construct test functions ϕ_{nj} as follows:

$$\begin{aligned}
\phi_{n0} &= \mathbb{1}\left\{\left\|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma}_0\right\|_\infty > M\epsilon/2\right\},\\
\phi_{nj} &= \mathbb{1}\left\{\left\|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right\|_\infty > \frac{\sqrt{M} + 2}{2}\|\mathbf{\Sigma}_0\|_\infty(M\epsilon^2)^{-(j-1)/2}\right\}.
\end{aligned}$$

We first control the type I error. By Lemma D.2,

$$\mathbb{E}_0(\phi_{n0}) \leq 4\exp\left\{4d - C_6 n\min\left(\frac{M\epsilon}{2\|\mathbf{\Sigma}_0\|_\infty}, \frac{M^2\epsilon^2}{4\|\mathbf{\Sigma}_0\|_\infty^2}\right)\right\} \leq 4\exp\left(4d - \frac{C_6 M^2 n\epsilon^2}{4\|\mathbf{\Sigma}_0\|_\infty^2}\right)$$

since $M\epsilon < 2\|\mathbf{\Sigma}_0\|_\infty$ by assumption. In addition, $M\epsilon^2 \leq \sqrt{M}M\epsilon^2 \leq (M\epsilon)^2 \leq 1$, and hence, for any $j \geq 1$,

$$\begin{aligned}
\mathbb{E}_0(\phi_{nj}) &\leq \mathbb{P}_0 \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma}_0 \right\|_\infty + \|\mathbf{\Sigma}_0\|_\infty > \frac{\sqrt{M}+2}{2}\|\mathbf{\Sigma}_0\|_\infty (M\epsilon^2)^{-(j-1)/2} \right\} \\
&\leq \mathbb{P}_0 \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma}_0 \right\|_\infty > \frac{\sqrt{M}}{2}\|\mathbf{\Sigma}_0\|_\infty (M\epsilon^2)^{-(j-1)/2} \right\} \\
&\leq 4\exp\left[4d - C_6 n \min\left\{ \frac{M(M\epsilon^2)^{2-j}}{4}, \frac{\sqrt{M}(M\epsilon^2)^{1/2-j/2}}{2} \right\} \right] \\
&\leq 4\exp\left(4d - C_6 \frac{M^{1-j/2}n\epsilon^{-(j-1)}}{2} \right).
\end{aligned}$$

Next we consider the type II error. For any $\mathbf{\Sigma} \in \mathcal{H}_{10}$, the type II error probability can be upper bounded by

$$\begin{aligned}
\mathbb{E}_{\mathbf{\Sigma}}(1-\phi_{n0}) &\leq \mathbb{P}_{\mathbf{\Sigma}} \left\{ \|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_\infty - \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma} \right\|_\infty \leq M\epsilon/2 \right\} \\
&\leq \mathbb{P}_{\mathbf{\Sigma}} \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma} \right\|_\infty > M\epsilon/2 \right\} \\
&\leq \mathbb{P}_{\mathbf{\Sigma}} \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma} \right\|_\infty > \|\mathbf{\Sigma}\|_\infty \frac{M\epsilon}{2(\sqrt{M}+2)\|\mathbf{\Sigma}_0\|_\infty} \right\} \\
&\leq 4\exp\left\{ 4d - \frac{C_6 M^2 n\epsilon^2}{4(\sqrt{M}+2)^2\|\mathbf{\Sigma}_0\|_\infty^2} \right\},
\end{aligned}$$

where the last inequality is due to Lemma D.2 and the assumption $M\epsilon < 2\|\mathbf{\Sigma}_0\|_\infty$. For any $\mathbf{\Sigma} \in \mathcal{H}_{1j}$ with $j \geq 1$, we estimate the type II error as follows:

$$\begin{aligned}
\mathbb{E}_{\mathbf{\Sigma}}(1-\phi_{nj}) &\leq \mathbb{P}_{\mathbf{\Sigma}} \left\{ \|\mathbf{\Sigma}\|_\infty - \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma} \right\|_\infty \leq \frac{\sqrt{M}+2}{2}\|\mathbf{\Sigma}_0\|_\infty (M\epsilon^2)^{-(j-1)/2} \right\} \\
&\leq \mathbb{P}_{\mathbf{\Sigma}} \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma} \right\|_\infty > \frac{\sqrt{M}+2}{2}\|\mathbf{\Sigma}_0\|_\infty (M\epsilon^2)^{-(j-1)/2} \right\} \\
&= \mathbb{P}_{\mathbf{\Sigma}} \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma} \right\|_\infty > \frac{(M\epsilon^2)^{1/2}}{2}(\sqrt{M}+2)(M\epsilon^2)^{-j/2}\|\mathbf{\Sigma}_0\|_\infty \right\} \\
&\leq \mathbb{P}_{\mathbf{\Sigma}} \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \mathbf{\Sigma} \right\|_\infty > \frac{(M\epsilon^2)^{1/2}}{2}\|\mathbf{\Sigma}\|_\infty \right\} \\
&\leq 4\exp\left(4d - \frac{C_6 M n\epsilon^2}{4} \right)
\end{aligned}$$

since $M\epsilon^2 \le M\epsilon \le 1$. Now we aggregate the individual tests by taking $\phi_n = \sup_{j\ge 0}\phi_{nj}$. Then the overall type I error probability can be bounded by

$$\begin{aligned}
\mathbb{E}_0(\phi_n) &\le \sum_{j=0}^{\infty}\mathbb{E}_0(\phi_{nj})\\
&\le 4\exp\left(4d - \frac{C_6M^2n\epsilon^2}{4\|\mathbf{\Sigma}_0\|_\infty^2}\right) + \sum_{j=1}^{\infty}4\exp\left(4d - C_6\frac{M^{1-j/2}n\epsilon^{-(j-1)}}{2}\right)\\
&= 4\exp\left(4d - \frac{C_6M^2n\epsilon^2}{4\|\mathbf{\Sigma}_0\|_\infty^2}\right) + 4\exp(4d)\sum_{j=1}^{\infty}\exp\left\{-\frac{C_6Mn\epsilon}{2}\left(\frac{1}{\sqrt{M}\epsilon}\right)^j\right\}\\
&\le 4\exp\left(4d - \frac{C_6M^2n\epsilon^2}{4\|\mathbf{\Sigma}_0\|_\infty^2}\right) + 4\exp(4d)\sum_{j=1}^{\infty}\exp\left\{-j\frac{C_6Mn\epsilon}{2}\left(\frac{1}{\sqrt{M}}\right)\right\}\\
&= 4\exp\left(4d - \frac{C_6M^2n\epsilon^2}{4\|\mathbf{\Sigma}_0\|_\infty^2}\right) + 8\exp\left(4d - \frac{C_6\sqrt{M}n}{2}\right),
\end{aligned}$$

since $M \ge \{(2\log 2)/C_6\}^2$. Furthermore, the overall type II error probability can also be bounded:

$$\begin{aligned}
\sup_{\mathbf{\Sigma}\in\mathcal{H}_1}\mathbb{E}_{\mathbf{\Sigma}}(1-\phi_n) &= \sup_{j\ge 0}\sup_{\mathbf{\Sigma}\in\mathcal{H}_{1j}}\mathbb{E}_{\mathbf{\Sigma}}(1-\phi_n) = \sup_{j\ge 0}\sup_{\mathbf{\Sigma}\in\mathcal{H}_{1j}}\mathbb{E}_{\mathbf{\Sigma}}\inf_{j\ge 0}(1-\phi_{jn})\\
&\le \sup_{j\ge 0}\sup_{\mathbf{\Sigma}\in\mathcal{H}_{1j}}\mathbb{E}_{\mathbf{\Sigma}}(1-\phi_{jn})\\
&\le \sup_{j\ge 0}\sup_{\mathbf{\Sigma}\in\mathcal{H}_{1j}}4\exp\left[4d - \frac{C_6Mn\epsilon^2}{4}\min\left\{1, \frac{M}{(\sqrt{M}+2)^2\|\mathbf{\Sigma}_0\|_\infty^2}\right\}\right]\\
&\le 4\exp\left\{4d - \frac{C_6Mn\epsilon^2}{4}\min\left(1, \frac{1}{4\|\mathbf{\Sigma}_0\|_\infty^2}\right)\right\}
\end{aligned}$$

since $M \ge 4$. The proof is thus completed. □

D.2 Existence of tests under the matrix infinity norm

Lemma D.3. *Assume the data $\mathbf{y}_1, \dots, \mathbf{y}_n$ follows $\mathrm{N}_p(\mathbf{0}_p, \mathbf{\Sigma})$, $1 \le r \le p$. Suppose $\mathbf{U}_0 \in \mathbb{O}(p, r)$ satisfy $|\mathrm{supp}(\mathbf{U}_0)| \le s$, and $r \le s \le p$. For any positive δ, t, and τ, define*

$$\mathcal{G}(\delta,\tau,t) = \left\{\mathbf{B}\in\mathbb{R}^{p\times r} : |\mathrm{supp}_\delta(\mathbf{B})| \le \tau, \sum_{j=1}^{p}\|\mathbf{B}_{j*}\|_1\mathbb{1}\{j\in\mathrm{supp}_\delta(\mathbf{B})\cup\mathrm{supp}(\mathbf{U}_0)\} \le t\right\}.$$

Let the positive sequences $(\delta_n, \tau_n, t_n, \epsilon_n)_{n=1}^{\infty}$ satisfy $\max(p\delta_nt_n, \delta_nt_n + p\delta_n^2) \le M_1\epsilon_n$ for some constant $M_1 > 0$, and $\epsilon_n \le 1$. Consider testing $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0 = \mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}} + \sigma_0^2\mathbf{I}_p$ versus

$$H_1 : \mathbf{\Sigma} \in \left\{\mathbf{\Sigma} = \mathbf{B}\mathbf{B}^{\mathrm{T}} + \sigma^2\mathbf{I}_p : \|\mathbf{\Sigma}-\mathbf{\Sigma}_0\|_\infty > M\epsilon_n, \mathbf{B}\in\mathcal{G}(\delta_n,\tau_n,t_n)\right\}.$$

Then there exists some absolute constant $C_6 > 0$, such that for each

$$M \in \left[\max\left\{\frac{M_1}{2}, 8, \frac{8(\log 2)^2}{C_6^2}\right\}, \frac{2\min(1, 2\|\mathbf{\Sigma}_0\|_2)}{\epsilon_n}\right],$$

there exists a test function $\phi_n : \mathbb{R}^{n\times p} \to [0,1]$, such that

$$\mathbb{E}_0(\phi_n) \le 12\exp\left\{6(\tau_n \log p + 2s_n) - C_6 \min\left(\frac{1}{2}, \frac{\|\mathbf{\Sigma}_0\|_\infty^2}{\sqrt{2}}\right)\frac{\sqrt{M}n\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\},$$
$$\sup_{\mathbf{\Sigma}\in H_1} \mathbb{E}_{\mathbf{\Sigma}}(1-\phi_n) \le 4\exp\left\{4(\tau_n + 2s_n) - C_6\min\left(\frac{\|\mathbf{\Sigma}_0\|_\infty^2}{8}, \frac{1}{32}\right)\frac{Mn\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\}.$$

Proof of Lemma D.3. Let $S_0 = \mathrm{supp}(\mathbf{U}_0)$ and $S(\delta) = \mathrm{supp}_\delta(\mathbf{B})$. Then there exists some permutation matrix $\mathbf{P}$ such that

$$\mathbf{B} = \mathbf{P}\begin{bmatrix}\mathbf{B}_\delta\\ \mathbf{A}_\delta\end{bmatrix} \quad \text{and} \quad \mathbf{U}_0 = \mathbf{P}\begin{bmatrix}\mathbf{U}_{0\delta}\\ \mathbf{0}\end{bmatrix},$$

where $\mathbf{B}_\delta$ and $\mathbf{U}_{0\delta}$ are $|S(\delta)\cup S_0| \times r$ matrix. Hence for $\mathbf{\Sigma} \in \mathcal{G}(\delta, \tau, t)$, it holds that

$$\begin{aligned}
\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_\infty &= \left\|\mathbf{P}\begin{bmatrix}\mathbf{B}_\delta\mathbf{B}_\delta^{\mathrm{T}} + \sigma^2\mathbf{I} - \mathbf{U}_{0\delta}\mathbf{\Lambda}_0\mathbf{U}_{0\delta}^{\mathrm{T}} - \sigma_0^2\mathbf{I} & \mathbf{B}_\delta\mathbf{A}_\delta^{\mathrm{T}}\\ \mathbf{A}_\delta\mathbf{B}_\delta^{\mathrm{T}} & \mathbf{A}_\delta\mathbf{A}_\delta^{\mathrm{T}} + (\sigma^2-\sigma_0^2)\mathbf{I}_d\end{bmatrix}\mathbf{P}^{\mathrm{T}}\right\|_\infty\\
&\le \left\|\begin{bmatrix}\mathbf{B}_\delta\mathbf{B}_\delta^{\mathrm{T}} + \sigma^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma^2\end{bmatrix} - \begin{bmatrix}\mathbf{U}_{0\delta}\mathbf{\Lambda}_0\mathbf{U}_{0\delta}^{\mathrm{T}} + \sigma_0^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_0^2\end{bmatrix}\right\|_\infty + \left\|\begin{bmatrix}\mathbf{0} & \mathbf{B}_\delta\mathbf{A}_\delta^{\mathrm{T}}\\ \mathbf{A}_\delta\mathbf{B}_\delta^{\mathrm{T}} & \mathbf{A}_\delta\mathbf{A}_\delta^{\mathrm{T}}\end{bmatrix}\right\|_\infty\\
&\le \left\|\mathbf{\Sigma}_{S(\delta)} - \mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_\infty + \max\left(\|\mathbf{B}_\delta\mathbf{A}_\delta^{\mathrm{T}}\|_\infty, \|\mathbf{A}_\delta\mathbf{B}_\delta^{\mathrm{T}}\|_\infty + \|\mathbf{A}_\delta\mathbf{A}_\delta^{\mathrm{T}}\|_\infty\right),
\end{aligned}$$

where

$$\mathbf{\Sigma}_{S(\delta)} = \begin{bmatrix}\mathbf{B}_\delta\mathbf{B}_\delta^{\mathrm{T}} + \sigma^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma^2\end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma}_{S(\delta)}^{(0)} = \begin{bmatrix}\mathbf{U}_{0\delta}\mathbf{\Lambda}_0\mathbf{U}_{0\delta}^{\mathrm{T}} + \sigma_0^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_0^2\end{bmatrix}.$$

Since

$$\begin{aligned}
&\max\left(\|\mathbf{B}_\delta\mathbf{A}_\delta^{\mathrm{T}}\|_\infty, \|\mathbf{A}_\delta\mathbf{B}_\delta^{\mathrm{T}}\|_\infty + \|\mathbf{A}_\delta\mathbf{A}_\delta^{\mathrm{T}}\|_\infty\right)\\
&\quad \le \max\left(\|\mathbf{B}_\delta\|_\infty\|\mathbf{A}_\delta^{\mathrm{T}}\|_\infty, \|\mathbf{A}_\delta\|_\infty\|\mathbf{B}_\delta^{\mathrm{T}}\|_\infty + \|\mathbf{A}_\delta\|_\infty\|\mathbf{A}_\delta^{\mathrm{T}}\|_\infty\right),
\end{aligned}$$

and

$$\begin{aligned}
\|\mathbf{B}_\delta\|_\infty &= \max_{j\in S_0\cup S(\delta)}\|\mathbf{B}_{j*}\|_1 \le \sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \mathbb{1}\{j \in S(\delta)\cup S_0\} \le t,\\
\|\mathbf{B}_\delta^{\mathrm{T}}\|_\infty &\le \sum_{j=1}^p \|\mathbf{B}_{j*}\|_1\mathbb{1}\{j\in S(\delta)\cup S_0\} \le t,\\
\|\mathbf{A}_\delta\|_\infty &= \max_{j\in S_0^c\cap S(\delta)^c}\|\mathbf{B}_{j*}\|_1 \le \max_{j\in S(\delta)^c}\|\mathbf{B}_{j*}\|_1 \le \delta,
\end{aligned}$$

$$\|\mathbf{A}_\delta^{\mathrm{T}}\|_\infty \le \sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \mathbb{1}\left\{j \in S_0^c \cap S(\delta)^c\right\} \le \sum_{j=1}^p \|\mathbf{B}_{j*}\|_1 \mathbb{1}\left\{j \in S(\delta)^c\right\} \le p\delta,$$

it follows that

$$\|\mathbf{\Sigma}-\mathbf{\Sigma}_0\|_\infty \le \left\|\mathbf{\Sigma}_{S(\delta)} - \mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_\infty + \max(p\delta_n t_n, \delta_n t_n + p\delta_n^2) \le \left\|\mathbf{\Sigma}_{S(\delta)} - \mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_\infty + M_1\epsilon_n.$$

By taking $M \ge 2M_1$, we obtain

$$\begin{aligned}
&\left\{\mathbf{\Sigma} = \mathbf{B}\mathbf{B}^{\mathrm{T}} + \sigma^2\mathbf{I} : \|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_\infty > M\epsilon_n, \mathbf{B} \in \mathcal{G}(\delta_n, \tau_n, t_n)\right\} \\
&\quad \subset \left\{\mathbf{\Sigma} : \left\|\mathbf{\Sigma}_{S(\delta)} - \mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_\infty > \frac{M}{2}\epsilon_n : \mathbf{B} \in \mathcal{G}(\delta_n, \tau_n, t_n)\right\} \\
&\quad \subset \bigcup_{S(\delta_n)\subset[p]:|S(\delta)|\le\tau_n} \left\{\mathbf{\Sigma} : \left\|\mathbf{\Sigma}_{S(\delta)} - \mathbf{\Sigma}_{S(\delta)}^{(0)}\right\|_\infty > \frac{M}{2}\epsilon_n\right\}.
\end{aligned}$$

Since both $\mathbf{\Sigma}_{S(\delta_n)}$ and $\mathbf{\Sigma}_{S(\delta_n)}^{(0)}$ are $(|S(\delta_n)\cup S_0|+1)\times(|S(\delta_n)\cup S_0|+1)$ square matrices, and

$$|S(\delta_n) \cup S_0| + 1 \le |S(\delta_n)| + S_0 + 1 \le \tau_n + 2s_n,$$

then for each $S(\delta_n) \subset [p]$ with $|S(\delta_n)| \le \tau_n$, and for each

$$M \in \left[\max\left\{\frac{M_1}{2}, 8, \frac{8(\log 2)^2}{C_6^2}\right\}, \frac{2\min(1, 2\|\mathbf{\Sigma}_0\|_2)}{\epsilon_n}\right],$$

(implying $M/2 \ge \max\{4, (2\log 2)^2/C_6^2\}$, $(M/2)\epsilon_n \le \min(1, \|\mathbf{\Sigma}_{S(\delta_n)}^{(0)}\|_2) = \min(1, \|\mathbf{\Sigma}_0\|_2)$), we invoke Lemma D.1 to construct a test $\phi_{S(\delta_n)}$ depending on the index set $S(\delta_n)$, such that the type I error probability satisfies

$$\begin{aligned}
\mathbb{E}_{\mathbf{\Sigma}_{S(\delta_n)}^{(0)}}\left(\phi_{S(\delta_n)}\right) &\le 4\exp\left\{4(\tau_n + 2s_n) - \frac{C_6 M^2 n\epsilon_n^2}{16\|\mathbf{\Sigma}_{S(\delta_n)}^{(0)}\|_\infty^2}\right\} \\
&\quad + 8\exp\left\{4(\tau_n + 2s_n) - C_6\sqrt{\frac{M}{2}}n\right\} \\
&\le 12\exp\left\{4(\tau_n + 2s_n) - C_6\min\left(\frac{M^2}{16\|\mathbf{\Sigma}_0\|_\infty^2}, \sqrt{\frac{M}{2}}\right)n\epsilon_n^2\right\} \\
&\le 12\exp\left\{4(\tau_n + 2s_n) - C_6\min\left(\frac{1}{2}, \frac{\|\mathbf{\Sigma}_0\|_\infty^2}{\sqrt{2}}\right)\frac{\sqrt{M}n\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\}.
\end{aligned}$$

In addition, for all $\mathbf{\Sigma}_{S(\delta_n)} \in \{\|\mathbf{\Sigma}_{S(\delta_n)} - \mathbf{\Sigma}_{S(\delta_n)}^{(0)}\|_2 > M\epsilon_n/2\}$, the type II error probability satisfies

$$\mathbb{E}_{\mathbf{\Sigma}_{S(\delta_n)}^{(1)}}\left(1 - \phi_{S(\delta_n)}\right) \le 4\exp\left\{4(\tau_n + 2s_n) - \frac{C_6 M n\epsilon_n^2}{8}\min\left(1, \frac{1}{4\|\mathbf{\Sigma}_0\|_\infty^2}\right)\right\}$$

$$\leq 4\exp\left\{4(\tau_n+2s_n)-C_6\min\left(\frac{\|\mathbf{\Sigma}_0\|_\infty^2}{8},\frac{1}{32}\right)\frac{Mn\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\}.$$

Notice that for each index set $S(\delta_n)$, the test function $\phi_{S(\delta_n)}$ is only a function of $\mathbf{Y}_n$ through the coordinates $[y_{ij}: i\in[n], j\in S(\delta_n)\cup S_0]$. Hence, $\mathbb{E}_{\mathbf{\Sigma}^{(0)}_{S(\delta_n)}}(\phi_{S(\delta_n)}) = \mathbb{E}_0(\phi_{S(\delta_n)})$, and for any $p\times p$ covariance matrix $\mathbf{\Sigma}$ with $\|\mathbf{\Sigma}_{S(\delta_n)}-\mathbf{\Sigma}^{(0)}_{S(\delta_n)}\|_\infty > M\epsilon_n/2$, it holds that

$$\mathbb{E}_{\mathbf{\Sigma}_{S(\delta_n)}}(1-\phi_{S(\delta_n)}) = \mathbb{E}_{\mathbf{\Sigma}}(1-\phi_{S(\delta_n)}).$$

Therefore, by aggregating the test functions

$$\phi_n = \max_{S(\delta_n)\subset[p]:|S(\delta_n)|\leq\tau_n}\phi_{S(\delta_n)},$$

we obtain

$$\begin{aligned}
\mathbb{E}_0(\phi_n) &\leq \sum_{S(\delta_n)\subset[p]:|S(\delta_n)|\leq\tau_n}\mathbb{E}_{\mathbf{\Sigma}^{(0)}_{S(\delta_n)}}(\phi_{S(\delta_n)})\\
&\leq 12\sum_{q=0}^{\lfloor\tau_n\rfloor}\frac{p!}{q!(p-q)!}\exp\left\{4(\tau_n+2s_n)-C_6\min\left(\frac{1}{2},\frac{\|\mathbf{\Sigma}_0\|_\infty^2}{\sqrt{2}}\right)\frac{\sqrt{M}n\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\}\\
&\leq 12(\tau_n+1)\exp(\tau_n\log p)\exp\left\{4(\tau_n+2s_n)-C_6\min\left(\frac{1}{2},\frac{\|\mathbf{\Sigma}_0\|_\infty^2}{\sqrt{2}}\right)\frac{\sqrt{M}n\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\}\\
&\leq 12\exp\left\{\tau_n+\tau_n\log p+4(\tau_n+2s_n)-C_6\min\left(\frac{1}{2},\frac{\|\mathbf{\Sigma}_0\|_\infty^2}{\sqrt{2}}\right)\frac{\sqrt{M}n\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\}\\
&\leq 12\exp\left\{6(\tau_n\log p+2s_n)-C_6\min\left(\frac{1}{2},\frac{\|\mathbf{\Sigma}_0\|_\infty^2}{\sqrt{2}}\right)\frac{\sqrt{M}n\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\},
\end{aligned}$$

and

$$\begin{aligned}
\sup_{\mathbf{\Sigma}\in H_1}\mathbb{E}_{\mathbf{\Sigma}}(1-\phi_n) &\leq \sup_{S(\delta_n)\subset[p]:|S(\delta_n)|\leq\tau_n}\ \sup_{\left\{\mathbf{\Sigma}:\|\mathbf{\Sigma}_{S(\delta_n)}-\mathbf{\Sigma}^{(0)}_{S(\delta_n)}\|_2>M\epsilon_n/2\right\}}\mathbb{E}_{\mathbf{\Sigma}_{S(\delta_n)}}\left(1-\phi_{S(\delta_n)}\right)\\
&\leq 4\exp\left\{4(\tau_n+2s_n)-C_6\min\left(\frac{\|\mathbf{\Sigma}_0\|_\infty^2}{8},\frac{1}{32}\right)\frac{Mn\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}\right\}.
\end{aligned}$$

The proof is thus completed. □

D.3 Proofs of Lemma 3.4 and Theorem 3.4

Before we proceed to the proofs, observe that the bounded coherence assumption on $\mathbf{U}_0$ (*i.e.*, $\|\mathbf{U}_0\|_{2\to\infty}\leq C_\mu\sqrt{r/s}$ for some $C_\mu\geq 1$) implies the following bound for the infinity norm on $\mathbf{\Sigma}_0$:

$$\|\mathbf{\Sigma}_0\|_\infty \leq \|\mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}}\|_\infty+\sigma_0^2 \leq \lambda_{01}\|\mathbf{U}_0\|_\infty\|\mathbf{U}_0^{\mathrm{T}}\|_\infty+\sigma_0^2$$

$$\le \lambda_{01}\left(\sqrt{r}\|\mathbf{U}_0\|_{2\to\infty}\right)\left(\sqrt{s}\|\mathbf{U}_0^{\mathrm{T}}\|_{2\to\infty}\right)+\sigma_0^2 \le C_\mu r\|\mathbf{\Sigma}_0\|_2.$$

Hence,

$$\frac{n\epsilon_n^2}{\|\mathbf{\Sigma}_0\|_\infty^2}=\frac{r^2 s\log p}{C_\mu^2 r^2\|\mathbf{\Sigma}_0\|_2^2}=\frac{s\log p}{C_\mu^2\|\mathbf{\Sigma}_0\|_2^2}.$$

The rest of the proofs breaks down into several steps as follows.

Step 1 remains the same as that in the proof of Theorem 3.1. In what follows we will make use of inequalities (C.3) and (C.4).

Step 2: Construct subsets $(\mathcal{G}_n)_{n=1}^\infty$. This step is also similar to that in the proof of Theorem 3.1. Take $\epsilon_n = r\sqrt{(s\log p)/n}$, $\tau_n=\beta s_n$, $t_n=(sr\log p)^2$, and $\delta_n=\epsilon_n/(pt_n)$, where $\beta>0$ is some constant to be specified later. Clearly, there exists some $\gamma>0$ such that

$$\delta_n=\frac{\epsilon_n}{pt_n}=\frac{r\sqrt{s\log p}}{p\sqrt{n}(sr\log p)^2}=\frac{1}{\sqrt{np^2s^3r^2(\log p)^3}}\ge\frac{1}{p^\gamma}.$$

Now let $\beta>4\mathrm{e}\gamma$ and $\mathcal{G}_n=\mathcal{G}(\delta_n,\tau_n,t_n)$ be defined in Lemma D.3. Since

$$\min\left\{\left(\frac{t_n}{\beta sr}\right)^2,\left(\frac{t_n}{r}\right)^2,\frac{t_n}{r}\right\}\ge\beta s\log p$$

for sufficiently large n, and $t_n/(sr)=(sr)\log p\to\infty$, we then can invoke Lemmas 3.2 and 3.3 to obtain

$$\begin{aligned}\Pi(\mathcal{G}_n^c)&\le\Pi(|\mathrm{supp}_{\delta_n}(\mathbf{B})|>\beta s_n)+\Pi\left[\sum_{j=1}^p\|\mathbf{B}_{j*}\|_1\mathbb{1}\{j\in\mathrm{supp}_{\delta_n}(\mathbf{B})\cup\mathrm{supp}(\mathbf{U}_0)\}>t_n\right]\\&\le 2\exp(-\beta s\log p)+5\exp\left\{-\min\left(\frac{\beta\kappa}{2},\frac{\beta}{2\mathrm{e}}-2\gamma\right)s\log p\right\}\\&\le 7\exp\left\{-\min\left(\frac{\beta\kappa}{2},\frac{\beta}{2\mathrm{e}}-2\gamma\right)s\log p\right\}\end{aligned}\tag{D.1}$$

for sufficiently large n (and hence sufficiently small $(s\log p)/p$).

Step 3: Decompose the integral. Since by construction we have

$$\max(p\delta_nt_n,\delta_nt_n+p\delta_n^2)\le p\delta_nt_n+p\delta_n^2\le 2p\delta_nt_n\le 2\epsilon_n,$$

then by Lemma D.3, there exists some absolute constant $C_6>0$, such that for sufficiently large n, and for each

$$M\in\left[\max\left\{8,\frac{8(\log 2)^2}{C_6^2}\right\},\frac{2\min(1,2\|\mathbf{\Sigma}_0\|_2)}{\epsilon_n}\right],$$

there exists a test function ϕ_n such that

$$\mathbb{E}_0(\phi_n) \le 12 \exp\left[- \left\{ \frac{C_6\sqrt{M}}{C_\mu^2 \|\boldsymbol{\Sigma}_0\|_2^2} \min\left(\frac{1}{2}, \frac{\|\boldsymbol{\Sigma}_0\|_2^2}{\sqrt{2}} \right) - 6(\beta+2) \right\} s \log p \right], \tag{D.2}$$

$$\mathbb{E}_{\boldsymbol{\Sigma}}(1-\phi_n) \le 4 \exp\left[- \left\{ \frac{C_6 M}{C_\mu^2 \|\boldsymbol{\Sigma}_0\|_2^2} \min\left(\frac{\|\boldsymbol{\Sigma}_0\|_2^2}{8}, \frac{1}{32} \right) - 4(\beta+2) \right\} s \log p \right] \tag{D.3}$$

for all $\boldsymbol{\Sigma} \in \{\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\|_\infty > M\epsilon_n\} \cap \mathcal{G}_n$. Denote $\mathcal{V}_n = \{\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\|_\infty \le M\epsilon_n\}$. Now we decompose the target integral $\mathbb{E}_0\{\Pi(\mathcal{V}_n^c \mid \mathbf{Y}_n)\}$ using (C.3) and (D.2) as follows:

$$\begin{aligned}
\mathbb{E}_0\{\Pi(\mathcal{V}_n^c \mid \mathbf{Y}_n)\} &\le \mathbb{E}_0(\phi_n) + \mathbb{E}_0\left\{(1-\phi_n)\Pi(\mathcal{V}_n \mid \mathbf{Y}_n)\mathbb{1}(\mathcal{A}_n)\right\} + \mathbb{P}_0(\mathcal{A}_n^c) \\
&\le 12 \exp\left[- \left\{ \frac{C_6\sqrt{M}}{C_\mu^2 \|\boldsymbol{\Sigma}_0\|_2^2} \min\left(\frac{1}{2}, \frac{\|\boldsymbol{\Sigma}_0\|_2^2}{\sqrt{2}} \right) - 6(\beta+2) \right\} s \log p \right] \\
&\quad + 2\exp\left\{ -\tilde{C}_3 \min\left(1, \|\boldsymbol{\Sigma}_0^{-1}\|_2^{-2}\right) s\log p \right\} \\
&\quad + \mathbb{E}_0\left[(1-\phi_n)\left\{ \frac{N_n(\mathcal{V}_n^c)}{D_n} \right\} \mathbb{1}(\mathcal{A}_n) \right].
\end{aligned}$$

Now we focus on the third term on the right-hand side of the preceding display. By (C.4), we obtain

$$\begin{aligned}
&\mathbb{E}_0\left[(1-\phi_n)\left\{ \frac{N_n(\mathcal{V}_n^c)}{D_n} \right\} \mathbb{1}(\mathcal{A}_n) \right] \\
&\quad \le \frac{\exp\{C_{0\lambda}\} s\log p\}}{C(\sigma_0^2)} \mathbb{E}_0\left\{ (1-\phi_n) \int_{\mathcal{V}_n^c} \prod_{i=1}^n \frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)} \Pi(\mathrm{d}\boldsymbol{\Sigma}) \right\}.
\end{aligned}$$

Observe that by Fubini's theorem,

$$\begin{aligned}
&\mathbb{E}_0\left\{ (1-\phi_n) \int_{\mathcal{V}_n^c} \prod_{i=1}^n \frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)} \Pi(\mathrm{d}\boldsymbol{\Sigma}) \right\} \\
&\quad \le \mathbb{E}_0\left\{ (1-\phi_n) \int_{\mathcal{V}_n^c \cap \mathcal{G}_n} \prod_{i=1}^n \frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)} \Pi(\mathrm{d}\boldsymbol{\Sigma}) \right\} + \mathbb{E}_0\left\{ \int_{\mathcal{G}_n^c} \prod_{i=1}^n \frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)} \Pi(\mathrm{d}\boldsymbol{\Sigma}) \right\} \\
&\quad = \int_{\mathcal{V}_n^c \cap \mathcal{G}_n} \mathbb{E}_0\left\{ (1-\phi_n) \prod_{i=1}^n \frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)} \right\} \Pi(\mathrm{d}\boldsymbol{\Sigma}) + \int_{\mathcal{G}_n^c} \left\{ \mathbb{E}_0 \prod_{i=1}^n \frac{p(\mathbf{y}_i \mid \boldsymbol{\Sigma})}{p(\mathbf{y}_i \mid \boldsymbol{\Sigma}_0)} \right\} \Pi(\mathrm{d}\boldsymbol{\Sigma}) \\
&\quad \le \int_{\mathcal{V}_n^c \cap \mathcal{G}_n} \mathbb{E}_{\boldsymbol{\Sigma}}(1-\phi_n)\Pi(\mathrm{d}\boldsymbol{\Sigma}) + \Pi(\mathcal{G}_n^c) \\
&\quad \le 4 \exp\left[- \left\{ \frac{C_6 M}{C_\mu^2 \|\boldsymbol{\Sigma}_0\|_2^2} \min\left(\frac{\|\boldsymbol{\Sigma}_0\|_2^2}{8}, \frac{1}{32} \right) - 4(\beta+2) \right\} s \log p \right] \\
&\qquad + 7\exp\left\{ -\min\left(\frac{\beta\kappa}{2}, \frac{\beta}{2\mathrm{e}} - 2\gamma \right) s\log p \right\},
\end{aligned}$$

where the testing type II error probability bound (D.3) and (D.1) are applied to the last inequality. Then by taking $M = M_\infty = \max(M_{\infty 1}, M_{\infty 2})$, where

$$\begin{aligned}
\beta &= \max\left[\frac{4}{\kappa}C_{0\lambda}, 2\mathrm{e}\left\{2\gamma + 2C_{0\lambda}\right\}\right],\\
M_{\infty 1} &= \max\left(\frac{32C_\mu^2\|\boldsymbol{\Sigma}_0\|_2^2}{C_6}, \frac{8C_\mu^2}{C_6}\right)\left\{4(\beta+2) + 2C_{0\lambda}\right\},\\
M_{\infty 2} &= \max\left(\frac{4C_\mu^4\|\boldsymbol{\Sigma}_0\|_2^4}{C_6^2}, \frac{2C_\mu^4}{C_6^2}\right)\left\{C_{0\lambda} + 6(\beta+2)\right\}^2,,
\end{aligned}$$

we obtain the following result:

$$\begin{aligned}
&\mathbb{E}_0\left[(1-\phi_n)\left\{\frac{N_n(\mathcal{V}_n^c)}{D_n}\right\}\mathbb{1}(\mathcal{A}_n)\right]\\
&\quad\le \frac{4}{C(\sigma_0^2)}\exp\left[-\left\{\frac{C_6 M}{C_\mu^2\|\boldsymbol{\Sigma}_0\|_2^2}\min\left(\frac{\|\boldsymbol{\Sigma}_0\|_2^2}{8}, \frac{1}{32}\right) - 4(\beta+2) - C_{0\lambda}\right\}s\log p\right]\\
&\qquad + \frac{7}{C(\sigma_0^2)}\exp\left[-\left\{\min\left(\frac{\beta\kappa}{2}, \frac{\beta}{2\mathrm{e}} - 2\gamma\right) - C_{0\lambda}\right\}s\log p\right]\\
&\quad\le \frac{11}{C(\sigma_0^2)}\exp\left(-C_{0\lambda}s\log p\right).
\end{aligned}$$

Combining the above results, we finally obtain

$$\begin{aligned}
\mathbb{E}_0\{\Pi(\mathcal{V}_n^c \mid \mathbf{Y}_n)\} &\le \left\{3 + \frac{11}{C(\sigma_0^2)}\right\}\exp\left(-C_{0\lambda}s\log p\right)\\
&\quad + 2\exp\left\{-\tilde{C}_3\min\left(1, \left\|\boldsymbol{\Sigma}_0^{-1}\right\|_2^{-2}\right)s\log p\right\}\\
&\le \left\{5 + \frac{11}{C(\sigma_0^2)}\right\}\exp\left\{-\min\left(C_{0\lambda}, \tilde{C}_3, \tilde{C}_3\|\boldsymbol{\Sigma}_0^{-1}\|_2^{-2}\right)s\log p\right\}\\
&= R_0\exp(-C_0 s\log p)
\end{aligned}$$

by taking $C_0 = \min\left\{C_{0\lambda}, \tilde{C}_3, \tilde{C}_3\|\boldsymbol{\Sigma}_0^{-1}\|_2^{-2}\right\}$ and $R_0 = \left\{5 + 11/C(\sigma_0^2)\right\}$. Therefore, there exists some constant M_∞, such that for all sufficiently large n, we have

$$\begin{aligned}
\mathbb{E}_0\left\{\Pi\left(\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\|_\infty > M\epsilon_n \,\middle|\, \mathbf{Y}_n\right)\right\} &\le \mathbb{E}_0\left\{\Pi\left(\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\|_\infty > M_0\epsilon_n \,\middle|\, \mathbf{Y}_n\right)\right\}\\
&\le R_0\mathrm{e}^{-C_0 s\log p}
\end{aligned}$$

for some absolute constants C_0 and R_0 depending on $\boldsymbol{\Lambda}_0$ and the hyperparameters only whenever $M \ge M_\infty$.

Step 4: Bounding the two-to-infinity norm loss using the Neumann trick. Let $\mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{U}_{\mathbf{B}}\boldsymbol{\Lambda}\mathbf{U}_{\mathbf{B}}^{\mathrm{T}}$ be the compact spectral decomposition of $\mathbf{B}\mathbf{B}^{\mathrm{T}}$. Denote $\mathbf{E} = \mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{U}_0\boldsymbol{\Lambda}_0\mathbf{U}_0^{\mathrm{T}}$

to be the "error" matrix. Clearly, $(\mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}} + \mathbf{E})\mathbf{U}_{\mathbf{B}} = (\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}\mathbf{U}_{\mathbf{B}}^{\mathrm{T}})\mathbf{U}_{\mathbf{B}} = \mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}$ by definition, yielding the matrix Sylvester equation

$$\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda} - \mathbf{E}\mathbf{U}_{\mathbf{B}} = (\mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}})\mathbf{U}_{\mathbf{B}}.$$

Now consider the events

$$\mathcal{U}_n = \left\{\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_2 \le M_0\sqrt{\frac{s\log p}{n}}\right\}, \quad \mathcal{V}_n = \left\{\|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_\infty \le M_\infty r\sqrt{\frac{s\log p}{n}}\right\}.$$

Suppose $\mathbf{\Sigma} \in \mathcal{U}_n \cap \mathcal{V}_n$. By the Weyl's inequality, for sufficiently large n, we have

$$\begin{aligned}
|\sigma^2 - \sigma_0^2| &= |\lambda_{r+1}(\mathbf{\Sigma}) - \lambda_{r+1}(\mathbf{\Sigma}_0)| \le \max_{k\in[p]}|\lambda_k(\mathbf{\Sigma}) - \lambda_k(\mathbf{\Sigma}_0)| \le \|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_2 \\
&\le M_0\sqrt{\frac{s\log p}{n}}, \\
\lambda_r(\mathbf{\Lambda}) &\ge \lambda_{0r} - |\lambda_{0r} - \lambda_r(\mathbf{\Lambda})| \ge \lambda_{0r} - |(\lambda_{0r} + \sigma_0^2) - \{\lambda_r(\mathbf{\Lambda}) + \sigma^2\}| - |\sigma_0^2 - \sigma^2| \\
&\ge \lambda_{0r} - \max_{k\in[p]}|\lambda_k(\mathbf{\Sigma}) - \lambda_k(\mathbf{\Sigma}_0)| - M_0\sqrt{\frac{s\log p}{n}} \\
&\ge \lambda_{0r} - 2M_0\sqrt{\frac{s\log p}{n}} > \max\left\{\frac{\lambda_{0r}}{2}, 2M_0\sqrt{\frac{s\log p}{n}}\right\}, \\
\|\mathbf{E}\|_2 &\le \|\mathbf{\Sigma} - \mathbf{\Sigma}_0\|_2 + \|(\sigma^2 - \sigma_0^2)\mathbf{I}_p\|_2 \le 2M_0\sqrt{\frac{s\log p}{n}}.
\end{aligned}$$

Therefore, the spectra of $\mathbf{\Lambda}$ and $\mathbf{E}$ are disjoint, and we can apply the Neumann's trick (see Theorem VII.2.2 in Bhatia, 1997) to expand $\mathbf{U}_{\mathbf{B}}$ in terms of a matrix series:

$$\mathbf{U}_{\mathbf{B}} = \sum_{m=0}^{\infty}\mathbf{E}^m(\mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}})\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-(m+1)} \tag{D.4}$$

Now we proceed to bound $\|\mathbf{U}_{\mathbf{B}} - \mathbf{U}_0\mathbf{W}_{\mathbf{U}}\|_{2\to\infty}$ using the techniques developed in Cape et al. (2019). Write

$$\begin{aligned}
\mathbf{U}_{\mathbf{B}} - \mathbf{U}_0\mathbf{W}_{\mathbf{U}} &= (\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}\mathbf{U}_{\mathbf{B}}^{\mathrm{T}} - \mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}})\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-1} + \mathbf{U}_0\mathbf{\Lambda}_0(\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-1} - \mathbf{\Lambda}_0^{-1}\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}}) \\
&\quad + \mathbf{U}_0(\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}} - \mathbf{W}_{\mathbf{U}}) \\
&= \mathbf{E}\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-1} + \mathbf{U}_0\mathbf{\Lambda}_0(\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-1} - \mathbf{\Lambda}_0^{-1}\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}}) + \mathbf{U}_0(\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}} - \mathbf{W}_{\mathbf{U}}).
\end{aligned}$$

By the CS decomposition and the sine-theta theorem, we see that the third term can be bounded:

$$\|\mathbf{U}_0(\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}} - \mathbf{W}_{\mathbf{U}})\|_{2\to\infty} \le \|\mathbf{U}_0\|_{2\to\infty}\|\mathbf{U}_{\mathbf{B}}\mathbf{U}_{\mathbf{B}}^{\mathrm{T}} - \mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}\|_2^2 \le \frac{4M_0^2C_\mu}{\lambda_{0r}^2}\left(\frac{\sqrt{rs}\log p}{n}\right).$$

Now we consider the second term. Denote $R = \mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-1} - \mathbf{\Lambda}_0^{-1}\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}}$. Then the (i,j)-th element of R can be represented as

$$r_{k\ell} = (\mathbf{U}_0)_{*k}^{\mathrm{T}}(\mathbf{U}_{\mathbf{B}})_{*\ell}\left\{\frac{1}{\lambda_\ell(\mathbf{\Lambda})} - \frac{1}{\lambda_{0k}}\right\} = \frac{1}{\lambda_\ell(\mathbf{\Lambda})\lambda_{0k}}\{\lambda_{0k} - \lambda_\ell(\mathbf{\Lambda})\}(\mathbf{U}_0)_{*k}^{\mathrm{T}}(\mathbf{U}_{\mathbf{B}})_{*\ell}.$$

Therefore, by defining $H_1 \in \mathbb{R}^{r\times r}$ by $(h_1)_{k\ell} = 1/\{\lambda_\ell(\mathbf{\Lambda})\lambda_{0k}\}$, we have

$$\begin{aligned}\|R\|_2 &= \|H_1 \circ (\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda} - \mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}})\|_2 \le r\|H_1\|_{\max}\|\mathbf{U}_0^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{B}}\|_2\\ &\le r\|H_1\|_{\max}2M_0\sqrt{\frac{s\log p}{n}},\end{aligned}$$

where $\circ$ represents the Hadamard matrix product (element-wise product), and $\|\cdot\|_{\max}$ is the maximum of the absolute values of the entries of a matrix. Furthermore, using the Weyl's inequality, we have

$$\|H_1\|_{\max} \le \frac{1}{\lambda_r(\mathbf{\Lambda})\lambda_{0r}} \le \frac{2}{\lambda_{0r}^2}$$

for sufficiently large n, since $\|\mathbf{\Lambda}^{-1}\|_2 = 1/\lambda_r(\mathbf{\Lambda}) \le 2/\lambda_{0r}$ for sufficiently large n. Hence, the second term can be bounded:

$$\begin{aligned}\|\mathbf{U}_0\mathbf{\Lambda}_0(\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-1} - \mathbf{\Lambda}_0^{-1}\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_{\mathbf{B}})\|_{2\to\infty} &= \|\mathbf{U}_0\|_{2\to\infty}\|\mathbf{\Lambda}_0\|_2\|R\|_2\\ &\le \frac{4M_0C_\mu\lambda_{01}}{\lambda_{0r}^2}\sqrt{\frac{r^3\log p}{n}}.\end{aligned}$$

Now we focus on the first term. By the Neumann matrix series (D.4), we have

$$\begin{aligned}\|\mathbf{E}\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-1}\|_{2\to\infty} &= \left\|\sum_{m=1}^{\infty}\mathbf{E}^m(\mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}})\mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}^{-(m+1)}\right\|_{2\to\infty}\\ &\le \|\mathbf{E}\mathbf{U}_0\|_{2\to\infty}\|\mathbf{\Lambda}_0\|_2\|\|\mathbf{\Lambda}^{-1}\|_2^2 + \sum_{m=2}^{\infty}\|\mathbf{E}\|_2^m\|\mathbf{\Lambda}_0\|_2\|\mathbf{\Lambda}^{-1}\|_2^{(m+1)}\\ &\le \|\mathbf{E}\mathbf{U}_0\|_{2\to\infty}\left\{\frac{\lambda_{01}}{\lambda_r(\mathbf{\Lambda})^2}\right\} + \left\{\frac{\lambda_{01}}{\lambda_r(\mathbf{\Lambda})}\right\}\frac{\|\mathbf{E}\|_2^2\|\mathbf{\Lambda}^{-1}\|_2^2}{1-\|\mathbf{E}\|_2\|\mathbf{\Lambda}^{-1}\|_2}\\ &\le 4\|\mathbf{E}\|_\infty\|\mathbf{U}_0\|_{2\to\infty}\frac{\lambda_{01}}{\lambda_{0r}^2} + \frac{8\lambda_{01}}{\lambda_{0r}^3}\|\mathbf{E}\|_2^2\\ &\le \frac{4M_\infty C_\mu\lambda_{01}}{\lambda_{0r}^2}\sqrt{\frac{r^3\log p}{n}} + \frac{8M_0^2\lambda_{01}}{\lambda_{0r}^3}\frac{s\log p}{n}\end{aligned}$$

for sufficiently large n. In other words, there exists some constant $M_{2\to\infty}$ depending on M_0, M_∞, $\mathbf{\Lambda}_0$, and hyperparameters, such that

$$\|\mathbf{U}_{\mathbf{B}} - \mathbf{U}_0\mathbf{W}_{\mathbf{U}}\|_{2\to\infty} \le M_{2\to\infty}\max\left(\sqrt{\frac{r^3\log p}{n}}, \frac{s\log p}{n}\right)$$

for sufficiently large n whenever $\mathbf{\Sigma} \in \mathcal{U}_n \cap \mathcal{V}_n$. Therefore,

$$\mathbb{E}_0\left[\Pi\left\{\|\mathbf{U}_{\mathbf{B}} - \mathbf{U}_0\mathbf{W}_{\mathbf{U}}\|_{2\to\infty} > M\max\left(\sqrt{\frac{r^3\log p}{n}}, \frac{s\log p}{n}\right)\right\}\right]$$

$$\leq \mathbb{E}_0 \left[\Pi \left\{ \|\mathbf{U}_{\mathbf{B}} - \mathbf{U}_0 \mathbf{W}_{\mathbf{U}}\|_{2\to\infty} > M_{2\to\infty} \max \left(\sqrt{\frac{r^3 \log p}{n}}, \frac{s \log p}{n} \right) \right\} \right]$$

$$\leq \mathbb{E}_0 \left\{ \Pi(\mathcal{U}_n^c \mid \mathbf{Y}_n) + \Pi(\mathcal{V}_n^c \mid \mathbf{Y}_n) \right\} \leq 2R_0 \mathrm{e}^{-C_0 s \log p},$$

for sufficiently large n when $M \geq M_{2\to\infty}$, completing the proof.

E Additional Numerical Results

E.1 MCMC convergence diagnostics

In this subsection, we present the convergence diagnostics of the MCMC sampler implemented in Section 4. For each dataset, the Metropolis-within-Gibbs sampler is implemented with 1000 burn-in iterations and 4000 post-burn-in samples that are collected after the burn-in period, so that the number of total MCMC iterations is 5000. We take the spectral norm $\|\boldsymbol{\Sigma}\|_2$ of the MCMC samples of $\boldsymbol{\Sigma}$ as an indicator for convergence check. The trace plot and the Gelman-Rubin convergence diagnostics (Gelman and Rubin, 1992) implemented in the `coda` package (Plummer et al., 2015) are adopted to assess the convergence of the MCMC chains.

Below, Figure 1 and Figure 2 present the trace plots of $\|\boldsymbol{\Sigma}\|_2$ after burn-in for Section 4.1 and Section 4.2, respectively, with four different colors highlighting four different MCMC chains with diverse initial values. These trace plots suggest that the MCMC chains mix well after the burn-in iterations. The Gelman-Rubin convergence diagnostics test results are summarized in Table 1 below. The point estimates of the potential scale reduction factors are close to 1 and the upper limits of the 95% confidence intervals are all no greater than 1.1. These numerical results suggest that there is no sign of non-convergence of the MCMC chains after the first 1000 burn-in iterations.

Table 1: Gelman-Rubin convergence diagnostics for Section 4 with the point estimates and the upper 95% confidence limits of the potential scale reduction factor given by the Gelman-Rubin convergence diagnostics implemented in the `coda` package.

s	8		12		20		40		Face data
r	1	4	1	4	1	4	1	4	
Point est.	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.02
Upper CI	1.03	1.01	1.01	1.01	1.02	1.00	1.01	1.02	1.07

E.2 Hyperparameter selection and sensitivity

We now discuss the practical aspects of the hyperparameter selection λ, a_σ, b_σ, κ and the sensitivity analysis through a simulation study. The setup for the synthetic dataset is the same as that in Section 4.1 of the manuscript but we only consider the case where the number of spikes is $r = 1$ and the number of non-zero rows in $\mathbf{U}_0$ is $s = 20$ for simplicity. We perform the sensitivity analysis of the hyperparameters by considering a grid values of them: $\kappa \in \{0.1, 1, 10\}$, $\lambda \in \{0.01, 0.1, 1, 10, 100\}$, and

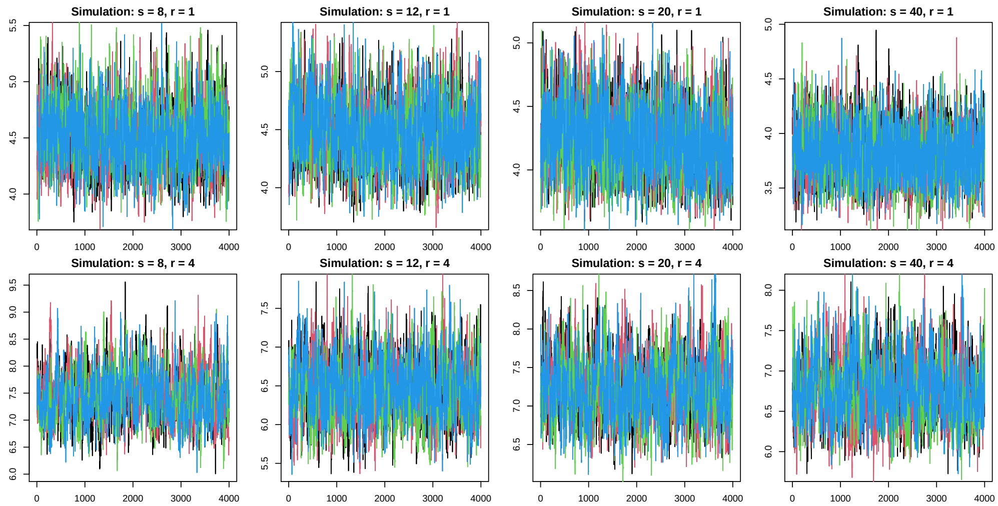

Figure 1: Trace plots of the spectral norm $\|\mathbf{\Sigma}\|_2$ across the post-burn-in MCMC samples for Section 4.1. Four different colors are used to highlight the trace plots of four different MCMC chains with different initializations.

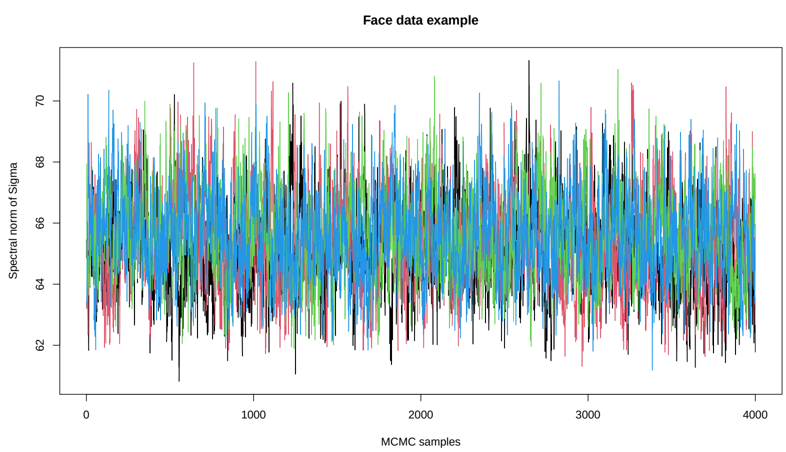

Figure 2: Trace plots of the spectral norm $\|\mathbf{\Sigma}\|_2$ across the post-burn-in MCMC samples after thinning for Section 4.2. Four different colors are used to highlight the trace plots of four different MCMC chains with different initializations.

Table 2: The medians of the spectral norm loss $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_2$ for Section E.2 across 50 Monte Carlo replicates of the synthetic datasets for Section E.2.

(a) $\kappa = 0.1$

λ \ (a_σ, b_σ)	$(0.01, 0.01)$	$(0.1, 0.1)$	$(1, 1)$	$(10, 10)$	$(100, 100)$
0.01	5.9421	5.8327	5.7288	5.8800	5.7559
0.1	5.5094	5.5101	5.1313	5.7221	5.5680
1	4.0154	3.8562	3.8312	3.7106	4.0427
10	5.8991	5.7350	5.8030	5.9359	5.9489
100	9.9511	9.9481	9.9500	9.9517	9.9503

(b) $\kappa = 1$

λ \ (a_σ, b_σ)	$(0.01, 0.01)$	$(0.1, 0.1)$	$(1, 1)$	$(10, 10)$	$(100, 100)$
0.01	5.6795	5.1464	5.3600	5.3734	5.1711
0.1	5.5170	5.1807	5.4149	5.4169	5.1275
1	3.7250	3.7818	3.8528	3.8862	3.8538
10	6.1104	6.1533	5.9961	5.9835	5.7849
100	9.9511	9.9482	9.9473	9.9530	9.9521

(c) $\kappa = 10$

λ \ (a_σ, b_σ)	$(0.01, 0.01)$	$(0.1, 0.1)$	$(1, 1)$	$(10, 10)$	$(100, 100)$
0.01	5.2294	4.8870	5.0485	5.2234	5.0931
0.1	4.4619	4.4141	4.5278	5.0051	4.6650
1	3.6260	3.9294	3.4814	3.6217	3.4913
10	6.2010	5.9664	6.4260	6.4270	5.9808
100	9.9526	9.9508	9.9468	9.9517	9.9519

$(a_\sigma, b_\sigma) \in \{(0.01, 0.01), (0.1, 0.1), (1, 1), (10, 10), (100, 100)\}$. For each hyperparameter setting $(\kappa, a_\sigma, b_\sigma, \lambda)$, we generate 50 independent Monte carlo replicates, compute the posterior mean $\widehat{\mathbf{\Sigma}}$ using the Metropolis-within-Gibbs sampler for each replicate, take the median of the spectral norm loss $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_2$ over the 50 Monte Carlo replicates, and tabulate the results in Table 2 below.

We can see that the spectral norm loss $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_2$ is robust to the hyperparameters (a_σ, b_σ) for $\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma)$ and κ for $\theta \sim \mathrm{Beta}\left(1, p^{1+\kappa}\right)$. In contrast, the posterior is comparatively more sensitive to λ for the "slab" component $\prod_{k=1}^{r} \psi_1(b_{jk} \mid \lambda)$ in MSSL prior and $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_2$ is minimized when $\lambda = 1$ in our grid search. In practice, we recommend setting $\lambda = 1$ and the users are welcome to set the other hyperparameters $a_\sigma, b_\sigma, \kappa$ based on their expertise together with the following interpretation: Small values of (a_σ, b_σ) correspond to weakly informative prior for σ^2 and large values of (a_σ, b_σ)

warrant strongly informative prior centered around $b_\sigma/(a_\sigma - 1)$. The hyperparameter κ for the prior $\theta \sim \text{Beta}(1, p^{1+\kappa})$ controls the prior sparsity on the rows of $\mathbf{B}$, with large values suggesting that the prior favors a large amount of the rows of $\mathbf{B}$ being close to zero.

References

Alzer, H. (1997). "On some inequalities for the incomplete gamma function." *Mathematics of Computation of the American Mathematical Society*, 66(218): 771–778. 5, 7

Bhatia, R. (1997). *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. New York: Springer.
URL http://catdir.loc.gov/catdir/enhancements/fy0818/96032217-d.html; http://www.gbv.de/dms/goettingen/213798093.pdf 22, 35

Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2015). "Dirichlet-Laplace Priors for Optimal Shrinkage." *Journal of the American Statistical Association*, 110(512): 1479–1490. PMID: 27019543.
URL https://doi.org/10.1080/01621459.2014.960967 7

Cape, J., Tang, M., and Priebe, C. E. (2019). "Signal-plus-noise matrix models: Eigenvector deviations and fluctuations." *Biometrika*, 106(1): 243–250.
URL https://doi.org/10.1093/biomet/asy070 35

Gao, C. and Zhou, H. H. (2015). "Rate-optimal posterior contraction for sparse PCA." *The Annals of Statistics*, 43(2): 785–818. 16

Gelman, A. and Rubin, D. B. (1992). "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science*, 7(4): 457 – 472.
URL https://doi.org/10.1214/ss/1177011136 37

Ghosal, S., Ghosh, J. K., and Van Der Vaart, A. W. (2000). "Convergence rates of posterior distributions." *Annals of Statistics*, 28(2): 500–531. 12

Hagerup, T. and Rüb, C. (1990). "A guided tour of Chernoff bounds." *Information processing letters*, 33(6): 305–308. 7

Pati, D., Bhattacharya, A., Pillai, N. S., and Dunson, D. (2014). "Posterior contraction in sparse Bayesian factor models for massive covariance matrices." *The Annals of Statistics*, 42(3): 1102–1130. 4, 13

Plummer, M., Best, N., Cowles, K., and Vines, K. (2015). "Package 'coda'." *URL http://cran. r-project. org/web/packages/coda/coda. pdf, accessed January*, 25: 2015. 37

Stewart, G. W. and Sun, J.-g. (1990). *Matrix Perturbation Theory*. Computer Science and Scientific Computing. Boston, MA: Academic Press.
URL https://www.worldcat.org/title/matrix-perturbation-theory/oclc/908946968 2

Vershynin, R. (2010). "Introduction to the non-asymptotic analysis of random matrices." *arXiv: 1011.3027*. 10

Yu, Y., Wang, T., and Samworth, R. J. (2015). "A useful variant of the Davis–Kahan theorem for statisticians." *Biometrika*, 102(2): 315–323.
URL http://dx.doi.org/10.1093/biomet/asv008 22