Least Favorable Direction Test for Multivariate Analysis of Variance in High Dimension

Beijing Institute of Technology

Supplementary Material

This file contains the proofs of Propositions and Theorems given in the main text.

S1 Technical lemmas

Lemma 1. *Suppose* $\mathbf{A}$ *is a* $p \times r$ *matrix with rank* r *and* $\mathbf{B}$ *is a* $p \times p$ *non-zero positive semi-definite matrix. Denote by* $\mathbf{A} = \mathbf{U_A D_A V_A^\top}$ *the singular value decomposition of* $\mathbf{A}$*, where* $\mathbf{U_A}$ *and* $\mathbf{V_A}$ *are* $p \times r$ *and* $r \times r$ *column orthogonal matrices, respectively, and* $\mathbf{D_A}$ *is a* $r \times r$ *diagonal matrix. Let* $\mathbf{P_A} = \mathbf{U_A U_A^\top}$ *be the projection matrix onto the column space of* $\mathbf{A}$*. Then*

$$\max_{a^\top a = 1, a^\top \mathbf{A}\mathbf{A}^\top a = 0} a^\top \mathbf{B} a = \lambda_1\big(\mathbf{B}(\mathbf{I}_p - \mathbf{P_A})\big).$$

Proof. It can be seen that $a^\top \mathbf{A}\mathbf{A}^\top a = 0$ if and only if $a = (\mathbf{I}_p - \mathbf{P_A})a$.

Then

$$\max_{a^\top a=1, a^\top \mathbf{A}\mathbf{A}^\top a=0} a^\top \mathbf{B} a = \max_{a^\top a=1, \mathbf{P}_{\mathbf{A}} a=0} a^\top (\mathbf{I}_p - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}_p - \mathbf{P}_{\mathbf{A}})a, \tag{S1.1}$$

which is obviously no greater than $\lambda_1\big((\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\big)$. To prove that they are equal, without loss of generality, we can assume $\lambda_1\big((\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\big) > 0$. Let α_1 be one eigenvector corresponding to the largest eigenvalue of $(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}-\mathbf{P}_{\mathbf{A}})$. Since $(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{P}_{\mathbf{A}} = (\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{A}}) = \mathbf{O}_{p\times p}$ and $\mathbf{P}_{\mathbf{A}}$ is symmetric, the rows of $\mathbf{P}_{\mathbf{A}}$ are eigenvetors of $(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}-\mathbf{P}_{\mathbf{A}})$ corresponding to eigenvalue 0. It follows that $\mathbf{P}_{\mathbf{A}}\alpha_1 = 0$. Therefore, α_1 satisfies the constraint of (S1.1) and thus (S1.1) is no less than $\lambda_1\big((\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\big)$. The conclusion now follows by noting that $\lambda_1\big((\mathbf{I}-\mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\big) = \lambda_1\big(\mathbf{B}(\mathbf{I}-\mathbf{P}_{\mathbf{A}})\big)$.

$\square$

Lemma 2. *Let $\xi_{n,i}$, $i=1,\ldots,n$, $n=1,2,\ldots$, be iid s-dimensional random vectors with mean zero, covariance matrix $\mathbf{M}$ and finite fourth moment. For $n=1,2,\ldots$, let $\{a_{n,i}\}_{i=1}^n$ be real random variables which are independent of $\{\xi_{n,i}\}_{i=1}^n$ and satisfy*

$$\frac{\max_{1\le i\le n} a_{n,i}^2}{\sum_{i=1}^n a_{n,i}^2} \xrightarrow{P} 0. \tag{S1.2}$$

Then

$$\Big(\sum_{i=1}^n a_{n,i}^2\Big)^{-1/2} \sum_{i=1}^n a_{n,i}\xi_{n,i} \xrightarrow{\mathcal{L}} \mathcal{N}_s(\mathbf{0}_s, \mathbf{M}).$$

Proof. First we observe that if $\{a_{n,i}\}_{i=1}^n$ are fixed numbers satisfying (S1.2), then Lyapunov central limit theorem and continuity theorem imply that for any $t \in \mathbb{R}^s$,

$$\mathrm{E}\left[\exp\left((\sum_{i=1}^n a_{n,i}^2)^{-1/2}\sum_{i=1}^n a_{n,i} i t^\top \xi_{n,i}\right)\right] \to \exp\left(-\frac{1}{2}t^\top \mathbf{M} t\right).$$

We only need to prove that for every subsequence of $\{n\}$, there is a further subsequence along which the conclusion holds. Let $\{m(n)\}$ be a subsequence of $\{n\}$. We can find a further subsequence of $\{m(n)\}$ along which (S1.2) holds almost surely. Then along this subsequence, our previous argument implies that for any $t \in \mathbb{R}^s$,

$$\mathrm{E}\left[\exp\left((\sum_{i=1}^n a_{n,i}^2)^{-1/2}\sum_{i=1}^n a_{n,i} i t^\top \xi_{n,i}\right)\bigg| a_{n,1},\ldots,a_{n,n}\right] \to \exp\left(-\frac{1}{2}t^\top \mathbf{M} t\right)$$

almost surely. Then by dominated convergence theorem, we have

$$\mathrm{E}\left[\exp\left((\sum_{i=1}^n a_{n,i}^2)^{-1/2}\sum_{i=1}^n a_{n,i} i t^\top \xi_{n,i}\right)\right] \to \exp\left(-\frac{1}{2}t^\top \mathbf{M} t\right)$$

along this further subsequence. This implies the conclusion holds along this further subsequence, which completes the proof.

□

Lemma 3 (Weyl's inequality)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two symmetric* $n \times n$ *ma-*

trices. If $r+s-1 \le i \le j+k-n$, we have

$$\lambda_j(\mathbf{A}) + \lambda_k(\mathbf{B}) \le \lambda_i(\mathbf{A}+\mathbf{B}) \le \lambda_r(\mathbf{A}) + \lambda_s(\mathbf{B}).$$

See, for example, Horn and Johnson (2012) Theorem 4.3.1.

Lemma 4 (von Neumann's trace theorem)**.** *Let $\mathbf{A}$ and $\mathbf{B}$ be two $m \times n$ matrices. Let $\sigma_1(\mathbf{A}) \ge \ldots \ge \sigma_q(\mathbf{A})$ and $\sigma_1(\mathbf{B}) \ge \cdots \ge \sigma_q(\mathbf{B})$ denote the non-increasingly ordered singular values of $\mathbf{A}$ and $\mathbf{B}$, respectively. Then*

$$\operatorname{tr}(\mathbf{A}\mathbf{B}^\top) \le \sum_{i=1}^{\min(m,n)} \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}).$$

See, for example, Horn and Johnson (2012) Theorem 7.4.1.1.

Lemma 5. *Let $\{Z_i\}_{i=1}^n$ be iid m-dimensional random vectors with common distribution $\mathcal{N}_m(\mathbf{0}_m, \mathbf{I}_m)$. Then for any n-dimensional vector $\omega = (\omega_1, \ldots, \omega_n)^\top$, we have*

$$\left\| \sum_{i=1}^n \omega_i (Z_i Z_i^\top - \mathbf{I}_m) \right\| = O_P(|\omega|_2 \sqrt{m} + |\omega|_\infty m),$$

where $|\omega|_2 = \sqrt{\sum_{i=1}^n \omega_i^2}$ and $|\omega|_\infty = \max_{1 \le i \le n} |\omega_i|$.

Remark 1. Our proof implies that the conclusion is still valid if ω is random and is independent of $\{Z_i\}_{i=1}^n$.

Proof. Our proof is adapted from the proof of Theorem 5.39 in Vershynin (2010). By Lemma 5.2 and Lemma 5.4 of Vershynin (2010), there exists a set $\mathcal{C} \subset \{x \in \mathbb{R}^m : |x|_2 = 1\}$ satisfying $\mathrm{Card}(\mathcal{C}) \le 9^m$ such that for any $m \times m$ symmetric matrix $\mathbf{A}$,

$$\|A\| \le 2 \max_{x \in \mathcal{C}} \left| x^\top \mathbf{A} x \right|. \tag{S1.3}$$

Then for $t > 4$,

$$\begin{aligned}
&\Pr\left(\left\| \sum_{i=1}^n \omega_i (Z_i Z_i^\top - \mathbf{I}_m) \right\| > t(|\omega|_2 \sqrt{m} + |\omega|_\infty m) \right) \\
\le &\Pr\left(2 \max_{x \in \mathcal{C}} \left| \sum_{i=1}^n \omega_i (x^\top Z_i Z_i^\top x - 1) \right| > t(|\omega|_2 \sqrt{m} + |\omega|_\infty m) \right) \\
\le &\sum_{x \in \mathcal{C}} \Pr\left(\left| \sum_{i=1}^n \omega_i (x^\top Z_i Z_i^\top x - 1) \right| > 2|\omega|_2 \sqrt{\frac{mt}{4}} + 2|\omega|_\infty \frac{mt}{4} \right) \\
\le &2 \cdot 9^m \exp\left(-\frac{mt}{4} \right) = 2 \exp\left((2 \log 3 - t/4) m \right),
\end{aligned}$$

where the first inequality follows from (S1.3), the second inequality follows from the union bound and the third inequality follows Lemma 1 of Laurent and Massart (2000). The upper bound $2\exp\left((2\log 3 - t/4)m\right)$ can be arbitrarily small as long as t is large enough. This completes the proof. $\square$

S2 Proofs of Propositions 1-4

Proof of Proposition 1. We only need to deal with the matrix $n^{-1}\mathbf{Z}^\top \mathbf{\Lambda} \mathbf{Z}$ since it shares the same non-zero eigenvalues as $\hat{\mathbf{\Sigma}}$. Write

$$\begin{aligned}
n^{-1}\mathbf{Z}^\top \mathbf{\Lambda} \mathbf{Z} =& n^{-1}\mathbf{Z}_1^\top \mathbf{\Lambda}_1 \mathbf{Z}_1 + n^{-1}\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \\
=& n^{-1}\mathbf{Z}_1^\top \mathbf{\Lambda}_1 \mathbf{Z}_1 + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n + n^{-1}\left(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n \right).
\end{aligned}$$

Then Weyl's inequality implies that for $i = 1, \ldots, r$,

$$\left|\lambda_i\left(n^{-1}\mathbf{Z}^\top\boldsymbol{\Lambda}\mathbf{Z}\right) - \lambda_i(n^{-1}\mathbf{Z}_1^\top\boldsymbol{\Lambda}_1\mathbf{Z}_1) - n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right| \leq n^{-1}\left\|\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_n\right\|. \tag{S2.1}$$

Using Weyl's inequality, we can derive the following lower bound for $\lambda_i(\mathbf{Z}_1^\top\boldsymbol{\Lambda}_1\mathbf{Z}_1)$, $i = 1, \ldots, r$.

$$\begin{aligned}
\lambda_i(\mathbf{Z}_1^\top\boldsymbol{\Lambda}_1\mathbf{Z}_1) \geq& \lambda_i(\mathbf{Z}_1^\top\operatorname{diag}(\boldsymbol{\lambda}_i\mathbf{I}_i, \mathbf{O}_{(r-i)\times(r-i)})\mathbf{Z}_1) \\
=& \lambda_i\Big(\boldsymbol{\lambda}_i\mathbf{Z}_1^\top\mathbf{Z}_1 - \boldsymbol{\lambda}_i\mathbf{Z}_1^\top\operatorname{diag}(\mathbf{O}_{i\times i}, \mathbf{I}_{r-i})\mathbf{Z}_1\Big) \\
\geq& \lambda_r\Big(\boldsymbol{\lambda}_i\mathbf{Z}_1^\top\mathbf{Z}_1\Big) + \lambda_{n+i-r}\Big(-\boldsymbol{\lambda}_i\mathbf{Z}_1^\top\operatorname{diag}(\mathbf{O}_{i\times i}, \mathbf{I}_{r-i})\mathbf{Z}_1\Big) \\
=& \boldsymbol{\lambda}_i\lambda_r(\mathbf{Z}_1\mathbf{Z}_1^\top).
\end{aligned}$$

Similarly, we can derive the following upper bound for $\lambda_i(\mathbf{Z}_1^\top\boldsymbol{\Lambda}_1\mathbf{Z}_1)$, $i = 1, \ldots, r$.

$$\begin{aligned}
&\lambda_i(\mathbf{Z}_1^\top\boldsymbol{\Lambda}_1\mathbf{Z}_1) \\
=& \lambda_i\Big(\mathbf{Z}_1^\top\big(\operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{i-1}, \mathbf{O}_{(r-i+1)\times(r-i+1)}) + \operatorname{diag}(\mathbf{O}_{(i-1)\times(i-1)}, \boldsymbol{\lambda}_i, \ldots, \boldsymbol{\lambda}_r)\big)\mathbf{Z}_1\Big) \\
\leq& \lambda_i\Big(\mathbf{Z}_1^\top\big(\operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{i-1}, \mathbf{O}_{(r-i+1)\times(r-i+1)})\Big) + \lambda_1\Big(\operatorname{diag}(\mathbf{O}_{(i-1)\times(i-1)}, \boldsymbol{\lambda}_i, \ldots, \boldsymbol{\lambda}_r)\big)\mathbf{Z}_1\Big) \\
\leq& \lambda_1(\mathbf{Z}_1^\top\operatorname{diag}(\mathbf{O}_{(i-1)\times(i-1)}, \boldsymbol{\lambda}_i\mathbf{I}_{r-i+1})\mathbf{Z}_1) \leq \boldsymbol{\lambda}_i\lambda_1(\mathbf{Z}_1\mathbf{Z}_1^\top).
\end{aligned}$$

The above lower bound and upper bound imply

$$\begin{aligned}
\left|\lambda_i(n^{-1}\mathbf{Z}_1^\top\boldsymbol{\Lambda}_1\mathbf{Z}_1) - \boldsymbol{\lambda}_i\right| \leq& \boldsymbol{\lambda}_i\max\left(|\lambda_1(n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top) - 1|, |\lambda_r(n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top) - 1|\right) \\
=& \boldsymbol{\lambda}_i\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\|.
\end{aligned} \tag{S2.2}$$

Combining the bounds (S2.1) and (S2.2) gives that for $i = 1, \ldots, r$,

$$\begin{aligned}&\left|\lambda_i\left(n^{-1}\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}\right) - \boldsymbol{\lambda}_i - n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right| \\ \le & n^{-1}\left\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n\right\| + \boldsymbol{\lambda}_i\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\|.\end{aligned}$$

From Lemma 5, we have

$$\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\| = O_P\left(\sqrt{\frac{r}{n}}\right), \tag{S2.3}$$

$$n^{-1}\left\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n\right\| = O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1}\right). \tag{S2.4}$$

This proves the first statement.

Next we prove the second statement. Note that

$$\begin{aligned}\sum_{i=r+1}^{n}\lambda_i(\hat{\mathbf{\Sigma}}) &= \sum_{i=r+1}^{n}\lambda_i(n^{-1}\ \mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}) \\ &= \operatorname{tr}(n^{-1}\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}) - \sum_{i=1}^{r}\lambda_i(n^{-1}\ \mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}) \\ &= \operatorname{tr}(n^{-1}\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2) - \frac{r}{n}\operatorname{tr}(\mathbf{\Lambda}_2) \\ &\quad - \left(\sum_{i=1}^{r}\lambda_i(n^{-1}\ \mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}) - \operatorname{tr}(n^{-1}\mathbf{Z}_1^\top\mathbf{\Lambda}_1\mathbf{Z}_1) - \frac{r}{n}\operatorname{tr}(\mathbf{\Lambda}_2)\right).\end{aligned}$$

It follows from inequalities (S2.1) and (S2.4) that

$$\begin{aligned}&\left|\sum_{i=1}^{r}\lambda_i(n^{-1}\ \mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}) - \operatorname{tr}(n^{-1}\mathbf{Z}_1^\top\mathbf{\Lambda}_1\mathbf{Z}_1) - \frac{r}{n}\operatorname{tr}(\mathbf{\Lambda}_2)\right| \\ \le & \frac{r}{n}\left\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n\right\| = O_P\left(r\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right).\end{aligned}$$

Thus,

$$\sum_{i=r+1}^{n}\lambda_i(\hat{\mathbf{\Sigma}}) = \operatorname{tr}(n^{-1}\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2) - \frac{r}{n}\operatorname{tr}(\mathbf{\Lambda}_2) + O_P\left(r\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right).$$

It is straightforward to show that

$$\mathrm{E}\,\mathrm{tr}(n^{-1}\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2)=\mathrm{tr}(\mathbf{\Lambda}_2),\quad \mathrm{Var}\left(\mathrm{tr}(n^{-1}\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2)\right)=\frac{2}{n}\,\mathrm{tr}(\mathbf{\Lambda}_2^2).$$

Hence

$$\begin{aligned}
&\sum_{i=r+1}^{n}\lambda_i(\hat{\mathbf{\Sigma}})\\
&=\mathrm{tr}(\mathbf{\Lambda}_2)+O_P\left(\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{n}}\right)-\frac{r}{n}\,\mathrm{tr}(\mathbf{\Lambda}_2)+O_P\left(r\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{n}}+r\boldsymbol{\lambda}_{r+1}\right)\\
&=\mathrm{tr}(\mathbf{\Lambda}_2)-\frac{r}{n}\,\mathrm{tr}(\mathbf{\Lambda}_2)+O_P\left(r\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{n}}+r\boldsymbol{\lambda}_{r+1}\right).
\end{aligned}$$

This completes the proof of the second statement. □

Proof of Proposition 2. The first two statements are direct consequences of Proposition 1 and the condition $r=o(n)$. Next we prove the third statement. We have $\widehat{\mathrm{tr}(\mathbf{\Lambda}_2^2)}=n^{-2}\sum_{i=r+1}^{n}\lambda_i^2(\mathbf{Y}^\top\mathbf{Y}-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)$. Note that Weyl's inequality implies that for $i=r+1,\ldots,n$,

$$\lambda_i(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)\le\lambda_i(\mathbf{Y}^\top\mathbf{Y}-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)\le\lambda_{i-r}(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n).$$

Define

$$\begin{aligned}
\mathcal{C}_1&=\left\{i:1\le i\le n,\,\lambda_i\left(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right)>0\right\},\\
\mathcal{C}_2&=\left\{i:r+1\le i\le n,\,\lambda_{i-r}\left(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right)\le 0\right\}.
\end{aligned}$$

It can be seen that $\mathcal{C}_1\cap\mathcal{C}_2=\emptyset$ and $\mathrm{Card}(\mathcal{C}_1\cup\mathcal{C}_2)\ge n-r$. For $i\ge r+1$ and $i\in\mathcal{C}_1$,

$$\lambda_i^2(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)\le\lambda_i^2(\mathbf{Y}^\top\mathbf{Y}-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)\le\lambda_{i-r}^2(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2-\widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n);$$

for $i \in \mathcal{C}_2$,

$$\lambda_{i-r}^2(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \lambda_i^2(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \lambda_i^2(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n);$$

for $i \geq r+1$ and $i \notin \mathcal{C}_1 \cup \mathcal{C}_2$,

$$\lambda_i^2(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \max\left(\lambda_{i-r}^2(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n), \lambda_i^2(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)\right).$$

Therefore,

$$\begin{aligned}
&\left|\sum_{i=r+1}^n \lambda_i^2\left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right) - \mathrm{tr}(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)^2\right| \\
\leq &\left|\sum_{i>r,\, i\in\mathcal{C}_1} \lambda_i^2\left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right) - \sum_{i\in\mathcal{C}_1} \lambda_i^2\left(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right)\right| \\
&+ \left|\sum_{i>r,\, i\in\mathcal{C}_2} \lambda_i^2\left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right) - \sum_{i\notin\mathcal{C}_1} \lambda_i^2\left(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right)\right| \\
&+ \left|\sum_{i>r,\, i\notin\mathcal{C}_1\cup\mathcal{C}_2} \lambda_i^2\left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right)\right| \\
\leq &3r\|\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\|^2 \\
\leq &3r\left(\|\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \mathrm{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n\| + \left|\mathrm{tr}(\mathbf{\Lambda}_2) - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\right|\right)^2 \\
= &O_P\left(rn\,\mathrm{tr}(\mathbf{\Lambda}_2^2) + rn^2\boldsymbol{\lambda}_{r+1}^2\right).
\end{aligned} \tag{S2.5}$$

where the last equality follows from (S2.4) and the second statement of the proposition.

Now we deal with $\mathrm{tr}(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)^2$. Let $Z_{2,i}$ be the ith column of $\mathbf{Z}_2$, $i = 1, \ldots, n$. Then

$$\operatorname{tr}(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)} \mathbf{I}_n)^2 = \sum_{i=1}^n (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,i} - \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)})^2 + 2 \sum_{1 \le i < j \le n} (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,j})^2.$$

For the first term, we have

$$\sum_{i=1}^n (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,i} - \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)})^2 \le 2 \sum_{i=1}^n (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,i} - \operatorname{tr}(\mathbf{\Lambda}_2))^2 + 2n(\widehat{\operatorname{tr}(\mathbf{\Lambda}_2)} - \operatorname{tr}(\mathbf{\Lambda}_2))^2.$$

Then it follows from the second statement of the proposition and the fact $\operatorname{E} \sum_{i=1}^n (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,i} - \operatorname{tr}(\mathbf{\Lambda}_2))^2 = 2n \operatorname{tr}(\mathbf{\Lambda}_2^2)$ that

$$\sum_{i=1}^n (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,i} - \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)})^2 = O_P\left((n + r^2) \operatorname{tr}(\mathbf{\Lambda}_2^2) + r^2 n \boldsymbol{\lambda}_{r+1}^2\right). \tag{S2.6}$$

For the second term, it is straightforward to show that $\operatorname{E} 2 \sum_{1 \le i < j \le n} (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,j})^2 = n(n-1) \operatorname{tr}(\mathbf{\Lambda}_2^2)$. Furthermore, Chen et al. (2010), Proposition A.2 implies that

$$\begin{aligned}
\operatorname{Var}\left(2 \sum_{1 \le i < j \le n} (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,j})^2\right) &= O\left(n^2 \operatorname{tr}^2(\mathbf{\Lambda}_2^2) + n^3 \operatorname{tr}(\mathbf{\Lambda}_2^4)\right) \\
&= O\left(n^2 \operatorname{tr}^2(\mathbf{\Lambda}_2^2) + n \operatorname{tr}(\mathbf{\Lambda}_2^2) n^2 \boldsymbol{\lambda}_{r+1}^2\right) \\
&= O\left(n^2 \operatorname{tr}^2(\mathbf{\Lambda}_2^2) + n^4 \boldsymbol{\lambda}_{r+1}^4\right).
\end{aligned}$$

Thus,

$$2 \sum_{1 \le i < j \le n} (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,j})^2 = n^2 \operatorname{tr}(\mathbf{\Lambda}_2^2) + O_P\left(n \operatorname{tr}(\mathbf{\Lambda}_2^2) + n^2 \boldsymbol{\lambda}_{r+1}^2\right).$$

Combining the last display and (S2.6) yields

$$\operatorname{tr}(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)} \mathbf{I}_n)^2 = n^2 \operatorname{tr}(\mathbf{\Lambda}_2^2) + O_P\left((n + r^2) \operatorname{tr}(\mathbf{\Lambda}_2^2) + (n + r^2) n \boldsymbol{\lambda}_{r+1}^2\right).$$

Combine the last display and (S2.5), we have

$$\sum_{i=r+1}^{n} \lambda_i^2 \left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right) = O_P\left(rn \operatorname{tr}(\mathbf{\Lambda}_2^2) + rn^2 \boldsymbol{\lambda}_{r+1}^2\right).$$

This completes the proof.

□

Proposition 6. *Suppose that* $r = o(n)$ *and* $r\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. *Then*

$$\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}^*_{\mathbf{Y},1}\| = O_P\left(\frac{\boldsymbol{\lambda}_{r+1} + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\boldsymbol{\lambda}_r + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}\right),$$

where

$$\mathbf{P}^*_{\mathbf{Y},1} = \mathbf{U}\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\left(\mathbf{I}_r + \mathbf{Q}^\top\mathbf{Q}\right)^{-1}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\mathbf{U}^\top.$$

Proof. The following intermediate matrix

$$\begin{aligned}\hat{\mathbf{\Sigma}}_0 =& n^{-1}\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1\mathbf{Z}_1^\top\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^\top + n^{-1}\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top + n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_1^\top\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^\top \\ &+ n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top\end{aligned}$$

plays a key role in the proof. It can be seen that

$$\hat{\mathbf{\Sigma}}_0 = n^{-1}\mathbf{U}\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1\mathbf{Z}_1^\top\mathbf{\Lambda}_1^{1/2}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\mathbf{U}^\top.$$

Consequently, $\hat{\mathbf{\Sigma}}_0$ is a positive semi-definite matrix with rank r, and $\mathbf{P}^*_{\mathbf{Y},1}$ is the projection matrix onto the rank r principal subspace of $\hat{\mathbf{\Sigma}}_0$.

From Cai et al. (2015), Proposition 1, we have

$$\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}^*_{\mathbf{Y},1}\| \le \frac{2\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0\|}{\lambda_r(\hat{\mathbf{\Sigma}}_0)}. \tag{S2.7}$$

We have the following upper bound for $\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0\|$.

$$
\begin{aligned}
\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0\| =& n^{-1} \left\| \mathbf{U}_2 \mathbf{\Lambda}_2^{1/2} \mathbf{Z}_2 \mathbf{Z}_2^\top \mathbf{\Lambda}_2^{1/2} \mathbf{U}_2^\top - \mathbf{U}_2 \mathbf{\Lambda}_2^{1/2} \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} \mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^{1/2} \mathbf{U}_2^\top \right\| \\
=& n^{-1} \left\| \mathbf{\Lambda}_2^{1/2} \mathbf{Z}_2 (\mathbf{I}_n - \mathbf{V}_{\mathbf{Z}_1} \mathbf{V}_{\mathbf{Z}_1}^\top) \mathbf{Z}_2^\top \mathbf{\Lambda}_2^{1/2} \right\| \\
\le& n^{-1} \left\| \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \right\| \\
\le& n^{-1} \left\| \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2) \mathbf{I}_n \right\| + n^{-1} \operatorname{tr}(\mathbf{\Lambda}_2) \\
=& O_P \left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1} + n^{-1} \operatorname{tr}(\mathbf{\Lambda}_2) \right) \\
=& O_P \left(\boldsymbol{\lambda}_{r+1} + n^{-1} \operatorname{tr}(\mathbf{\Lambda}_2) \right),
\end{aligned}
\tag{S2.8}
$$

where the second last equality follows from (S2.4) and the last equality follows from

$$
\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}} \le \sqrt{\frac{\boldsymbol{\lambda}_{r+1} \operatorname{tr}(\mathbf{\Lambda}_2)}{n}} \le \frac{1}{2} \left(\boldsymbol{\lambda}_{r+1} + n^{-1} \operatorname{tr}(\mathbf{\Lambda}_2) \right).
$$

Now we deal with $\lambda_r(\hat{\mathbf{\Sigma}}_0)$. We have

$$
\begin{aligned}
\lambda_r(\hat{\mathbf{\Sigma}}_0) =& \lambda_r \left(n^{-1} (\mathbf{Z}_1 \mathbf{Z}_1^\top)^{1/2} \mathbf{\Lambda}_1^{1/2} (\mathbf{I}_r + \mathbf{Q}^\top \mathbf{Q}) \mathbf{\Lambda}_1^{1/2} (\mathbf{Z}_1 \mathbf{Z}_1^\top)^{1/2} \right) \\
=& \lambda_r \left(n^{-1} (\mathbf{Z}_1 \mathbf{Z}_1^\top)^{1/2} \mathbf{\Lambda}_1 (\mathbf{Z}_1 \mathbf{Z}_1^\top)^{1/2} + n^{-1} \mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} \right).
\end{aligned}
$$

It can be seen that $\mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1}$ is a $(p - r) \times r$ random matrix with iid $\mathcal{N}(0, 1)$

entries. Then Lemma 5 implies that

$$
\begin{aligned}
\left\|n^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} - n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\| =& O_P\left(n^{-1}\sqrt{r\operatorname{tr}\left(\mathbf{\Lambda}_2^2\right)} + rn^{-1}\boldsymbol{\lambda}_{r+1}\right) \\
=& O_P\left(n^{-1}\sqrt{r\boldsymbol{\lambda}_{r+1}\operatorname{tr}\left(\mathbf{\Lambda}_2\right)} + rn^{-1}\boldsymbol{\lambda}_{r+1}\right) \\
=& o_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right),
\end{aligned}
\tag{S2.9}
$$

where the last equality follows from the condition $r\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. Then it follows from Weyl's inequality that

$$
\begin{aligned}
&\left|\lambda_r(\hat{\mathbf{\Sigma}}_0) - \lambda_r\left(n^{-1}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{1/2}\mathbf{\Lambda}_1(\mathbf{Z}_1\mathbf{Z}_1^\top)^{1/2} + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right)\right| \\
\leq& \left\|n^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} - n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\| \\
=& o_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right).
\end{aligned}
$$

On the other hand, (S2.2) and (S2.3) imply that

$$
\begin{aligned}
&\lambda_r\left(n^{-1}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{1/2}\mathbf{\Lambda}_1(\mathbf{Z}_1\mathbf{Z}_1^\top)^{1/2} + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right) \\
=& \lambda_r\left(n^{-1}\mathbf{Z}_1^\top\mathbf{\Lambda}_1\mathbf{Z}_1\right) + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2) \\
=& \boldsymbol{\lambda}_r + o_P(\boldsymbol{\lambda}_r) + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2).
\end{aligned}
$$

Hence we have

$$
\lambda_r(\hat{\mathbf{\Sigma}}_0) = (1 + o_P(1))(\boldsymbol{\lambda}_r + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)). \tag{S2.10}
$$

Then the conclusion follows from (S2.7), (S2.8) and (S2.10). □

Proof of Proposition 3. Note that

$$
\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1}^\dagger\right\| \leq \left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1}^*\right\| + \left\|\mathbf{P}_{\mathbf{Y},1}^* - \mathbf{P}_{\mathbf{Y},1}^\dagger\right\|.
$$

Under the condition $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$, Proposition 6 implies that

$$\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}^*_{\mathbf{Y},1}\right\| = O_P\left(\frac{\boldsymbol{\lambda}_{r+1}}{\boldsymbol{\lambda}_r} + \frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right).$$

So we only need to deal with $\|\mathbf{P}^*_{\mathbf{Y},1} - \mathbf{P}^\dagger_{\mathbf{Y},1}\|$. We have

$$\begin{aligned}
&\left\|\mathbf{P}^*_{\mathbf{Y},1} - \mathbf{P}^\dagger_{\mathbf{Y},1}\right\| \\
\leq&\left\|\mathbf{P}^*_{\mathbf{Y},1} - \mathbf{U}\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\mathbf{U}^\top\right\| + \left\|\mathbf{U}\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\mathbf{U}^\top - \mathbf{P}^\dagger_{\mathbf{Y},1}\right\| \\
=&\left\|\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\left(\left(\mathbf{I}_r + \mathbf{Q}^\top\mathbf{Q}\right)^{-1} - \mathbf{I}_r\right)\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\right\| + \left\|\mathbf{U}_2\mathbf{Q}\mathbf{Q}^\top\mathbf{U}_2^\top\right\| \\
=&\left\|\left(\left(\mathbf{I}_r + \mathbf{Q}^\top\mathbf{Q}\right)^{-1} - \mathbf{I}_r\right)\left(\mathbf{I}_r + \mathbf{Q}^\top\mathbf{Q}\right)\right\| + \left\|\mathbf{U}_2\mathbf{Q}\mathbf{Q}^\top\mathbf{U}_2^\top\right\| \\
=&2\left\|\mathbf{Q}^\top\mathbf{Q}\right\|.
\end{aligned}$$

Note that

$$\begin{aligned}
\|\mathbf{Q}^\top\mathbf{Q}\| =& \left\|\mathbf{\Lambda}_1^{-1/2}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1/2}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1/2}\mathbf{\Lambda}_1^{-1/2}\right\| \\
\leq&\boldsymbol{\lambda}_r^{-1}\left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\right\|\left\|\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \qquad (S2.11)\\
=&O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right),
\end{aligned}$$

where the second last equality follows from the fact $\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\| = \lambda_r(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}$, (S2.3), (S2.9) and Weyl's inequality. Therefore, we have

$$\left\|\mathbf{P}^*_{\mathbf{Y},1} - \mathbf{P}^\dagger_{\mathbf{Y},1}\right\| = O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right).$$

This completes the proof.

□

Proposition 7. *Suppose that* $r = o(n)$ *and* $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$. *Then*

$$\left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}^*_{\mathbf{Y},2}\right\| = O_P\left(\min\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}, 1\right)\right).$$

where $\mathbf{P}^*_{\mathbf{Y},2} = \mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top$.

Proof. We only need to prove that for any subsequence of $\{n\}$, there is a further subsequence along which the conclusion holds. Thus, without loss of generality, we assume $\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to c \in [0, +\infty]$. Since $\mathbf{P}_{\mathbf{Y},2}$ and $\mathbf{P}^*_{\mathbf{Y},2}$ are both projection matrices, we have $\left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}^*_{\mathbf{Y},2}\right\| \le 2$. Therefore, the conclusion holds if $c > 0$. In the rest of the proof, we assume $c = 0$, that is $\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$.

Note that $\mathbf{U}_{\mathbf{Y},2}$ is in fact the leading $n-r$ eigenvectors of $(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})$. Under the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$, Proposition 3 implies that

$$\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}^\dagger_{\mathbf{Y},1}\right\| = O_P\left(\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right).$$

It can be seen that

$$\begin{aligned}&\left\|(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}) - (\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\right\| \\ \le&\left\|(\mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1})\right\| + 2\left\|(\mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\right\|.\end{aligned}$$

Under the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$, Proposition 1 implies that

$$\|\hat{\boldsymbol{\Sigma}}\| = \boldsymbol{\lambda}_1\left(1 + \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_1} + O_P\left(\sqrt{\frac{r}{n}} + \sqrt{\frac{\boldsymbol{\lambda}_{r+1}}{\boldsymbol{\lambda}_1}\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_1}} + \frac{\boldsymbol{\lambda}_{r+1}}{\boldsymbol{\lambda}_1}\right)\right) = \boldsymbol{\lambda}_1(1+o_P(1)).$$

Then

$$\left\|(\mathbf{P}_{\mathbf{Y},1}^{\dagger}-\mathbf{P}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{P}_{\mathbf{Y},1}^{\dagger}-\mathbf{P}_{\mathbf{Y},1})\right\| \leq \|\hat{\mathbf{\Sigma}}\| \left\|\mathbf{P}_{\mathbf{Y},1}^{\dagger}-\mathbf{P}_{\mathbf{Y},1}\right\|^2 = O_P\left(\frac{\operatorname{tr}^2(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}\right). \tag{S2.12}$$

On the other hand, we have

$$\begin{aligned}
&\left\|(\mathbf{P}_{\mathbf{Y},1}^{\dagger}-\mathbf{P}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| \\
\leq&\left\|\mathbf{P}_{\mathbf{Y},1}^{\dagger}-\mathbf{P}_{\mathbf{Y},1}\right\| \left\|n^{-1}\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}\right\| \left\|\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| \\
=&n^{-1/2}\left\|\mathbf{P}_{\mathbf{Y},1}^{\dagger}-\mathbf{P}_{\mathbf{Y},1}\right\| \left\|\hat{\mathbf{\Sigma}}\right\|^{1/2} \left\|\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| \\
=&O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right)\left\|\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\|.
\end{aligned}$$

It is straightforward to show that

$$\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger}) = \tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top - \mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{Z}_1^\top(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\mathbf{\Lambda}_1^{-1/2}\mathbf{U}_1^\top. \tag{S2.13}$$

Then

$$\left\|\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| \leq \left\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\right\|^{1/2} + \boldsymbol{\lambda}_r^{-1/2}\left\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\right\| \left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\right\|^{1/2}.$$

It follows from (S2.4) and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$ that

$$\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\| = (1+o_P(1))\operatorname{tr}(\mathbf{\Lambda}_2). \tag{S2.14}$$

Consequently,

$$\left\|\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| = O_P\left(\operatorname{tr}^{1/2}(\mathbf{\Lambda}_2)\right) + O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{n\boldsymbol{\lambda}_r}}\right) = O_P\left(\operatorname{tr}^{1/2}(\mathbf{\Lambda}_2)\right).$$

Thus,

$$\left\|(\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\right\|=O_P\left(\frac{\mathrm{tr}^{3/2}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right). \tag{S2.15}$$

Combine (S2.12) and (S2.15), we obtain

$$\begin{aligned}&\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})-(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\right\|\\=&O_P\left(\frac{\mathrm{tr}^2(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}+\frac{\mathrm{tr}^{3/2}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right).\end{aligned}$$

Now we deal with $(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})$. In view of (S2.13), we have

$$\begin{aligned}&(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\\=&n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top-n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{Q}\mathbf{U}_1^\top\\&-n^{-1}\mathbf{U}_1\mathbf{Q}^\top\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top+n^{-1}\mathbf{U}_1\mathbf{Q}^\top\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{Q}\mathbf{U}_1^\top.\end{aligned}$$

Then

$$\begin{aligned}&\left\|(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})-n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\|\\\le&n^{-1}\left\|\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{Q}\right\|+n^{-1}\left\|\mathbf{Q}^\top\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{Q}\right\|\\\le&n^{-1}\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\|\|\mathbf{Q}^\top\mathbf{Q}\|^{1/2}+n^{-1}\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\|\|\mathbf{Q}^\top\mathbf{Q}\|\\=&O_P\left(\frac{\mathrm{tr}^{3/2}(\mathbf{\Lambda}_2)}{n^{3/2}\boldsymbol{\lambda}_r^{1/2}}\right),\end{aligned}$$

where the last equality follows from (S2.11) and (S2.14).

Combine the above bounds, we obtain

$$\begin{aligned}&\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})-n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\|\\=&O_P\left(\frac{\mathrm{tr}^2(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}+\frac{\mathrm{tr}^{3/2}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right).\end{aligned} \tag{S2.16}$$

The matrix $n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top$ shares the same non-zero eigenvalues as $n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$. Note that $\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ is a $p \times (n-r)$ random matrix with iid $\mathcal{N}(0,1)$ entries. Then it follows from Lemma 5 and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$ that

$$\begin{aligned}
\left\|n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1} - n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_{n-r}\right\| =& O_P\left(n^{-1/2}\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)} + \boldsymbol{\lambda}_{r+1}\right) \\
=& O_P\left(n^{-1/2}\sqrt{\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\mathbf{\Lambda}_2)} + \boldsymbol{\lambda}_{r+1}\right) \\
=& o_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right).
\end{aligned} \tag{S2.17}$$

This bound, combined with Weyl's inequality, leads to

$$\lambda_{n-r}\left(n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right) = (1+o_P(1))n^{-1}\operatorname{tr}(\mathbf{\Lambda}_1). \tag{S2.18}$$

It can be seen that the matrix $\mathbf{P}_{\mathbf{Y},2}^*$ is the projection matrix onto the rank $n-r$ principal subspace of $n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top$. Therefore, Cai et al. (2015), Proposition 1 implies that

$$\begin{aligned}
&\left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}_{\mathbf{Y},2}^*\right\| \\
\le& \frac{2\left\|(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}) - n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\|}{\lambda_{n-r}\left(n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right)} \\
=& O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2} + \sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}\right) \\
=& O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}\right),
\end{aligned}$$

where the second last equality follows from (S2.16) and (S2.18). This completes the proof. □

Proof of Proposition 4. By some algebra, it can be seen that

$$\begin{aligned}\left\|\mathbf{P}^*_{\mathbf{Y},2} - \mathbf{P}^\dagger_{\mathbf{Y},2}\right\| &= (\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\left\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1} - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_{n-r}\right\| \\ &= O_P\left(\frac{\sqrt{n\operatorname{tr}(\mathbf{\Lambda}_2^2)}}{\operatorname{tr}(\mathbf{\Lambda}_2)} + \frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}\right) \\ &= O_P\left(\sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}\right),\end{aligned}$$

where the second last equality follows from (S2.17) and the last equality follows from the fact $\sqrt{n\operatorname{tr}(\mathbf{\Lambda}_2^2)}/\operatorname{tr}(\mathbf{\Lambda}_2) \le \sqrt{n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)}$ and the condition $\sqrt{n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)} \to 0$. Then the conclusion follows from the last display and Proposition 7.

□

S3 Proofs of Theorems 1 and 2

It can be seen that $\mathbf{XJC}$ is independent of $\mathbf{Y}$. We write $\mathbf{XJC} = \mathbf{\Theta C} + \mathbf{U\Lambda}^{1/2}\mathbf{Z}^\dagger$, where $\mathbf{Z}^\dagger$ is a $p \times (k-1)$ matrix with iid $\mathcal{N}(0,1)$ entries and is independent of $\mathbf{Z}$. Then

$$\begin{aligned}\mathbf{C}^\top\mathbf{J}^\top\mathbf{X}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{XJC} =& \mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U\Lambda}^{1/2}\mathbf{Z}^\dagger + \mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Theta C} \\ &+ \mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U\Lambda}^{1/2}\mathbf{Z}^\dagger + \mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Theta C}.\end{aligned} \tag{S3.1}$$

It can be seen that the first term of (S3.1) can be written as

$$\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger=\sum_{i=1}^p\lambda_i((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top,$$

where $\eta_1,\ldots,\eta_p$ are independent $\mathcal{N}_{k-1}(\mathbf{0}_{k-1},\mathbf{I}_{k-1})$ random vectors and are independent of $\mathbf{P}_\mathbf{Y}$.

Lemma 6. *Suppose that $n\boldsymbol{\lambda}_1/\operatorname{tr}(\mathbf{\Sigma})\to 0$. Then*

$$\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\right)=\operatorname{tr}(\mathbf{\Sigma})-\frac{n\operatorname{tr}(\mathbf{\Sigma}^2)}{\operatorname{tr}(\mathbf{\Sigma})}+O_P\left(n(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p)\sqrt{\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\mathbf{\Sigma})}}\right),$$

$$\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\right)^2=\operatorname{tr}(\mathbf{\Sigma}^2)-\frac{n\operatorname{tr}^2(\mathbf{\Sigma}^2)}{\operatorname{tr}^2(\mathbf{\Sigma})}+O_P(n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p)).$$

Proof. First we approximate $\mathbf{P}_\mathbf{Y}$ by a simple expression. We have

$$\begin{aligned}\left\|\mathbf{P}_\mathbf{Y}-(\operatorname{tr}(\mathbf{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^\top\right\|=&\left\|\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top-(\operatorname{tr}(\mathbf{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^\top\right\|\\=&(\operatorname{tr}(\mathbf{\Sigma}))^{-1}\left\|\mathbf{Y}^\top\mathbf{Y}-\operatorname{tr}(\mathbf{\Sigma})\mathbf{I}_n\right\|.\end{aligned}$$

Then from Lemma 5, we have

$$\begin{aligned}\left\|\mathbf{P}_\mathbf{Y}-(\operatorname{tr}(\mathbf{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^\top\right\|=&(\operatorname{tr}(\mathbf{\Sigma}))^{-1}\left\|\mathbf{Z}^\top\mathbf{\Sigma}\mathbf{Z}-\operatorname{tr}(\mathbf{\Sigma})\mathbf{I}_n\right\|\\=&O_P\left(\frac{\sqrt{n\operatorname{tr}(\mathbf{\Sigma}^2)}}{\operatorname{tr}(\mathbf{\Sigma})}+\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\mathbf{\Sigma})}\right)\\=&O_P\left(\frac{\sqrt{n\boldsymbol{\lambda}_1\operatorname{tr}(\mathbf{\Sigma})}}{\operatorname{tr}(\mathbf{\Sigma})}+\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\mathbf{\Sigma})}\right)\\=&O_P\left(\sqrt{\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\mathbf{\Sigma})}}\right).\end{aligned}\tag{S3.2}$$

Now we deal with $\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)$. It can be seen that

$$\begin{aligned}\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) &= \operatorname{tr}(\boldsymbol{\Sigma}) - \operatorname{tr}(\boldsymbol{\Sigma}\mathbf{P}_{\mathbf{Y}}) \\ &= \operatorname{tr}(\boldsymbol{\Sigma}) - \operatorname{tr}\left(\left(\boldsymbol{\Sigma} - \frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right) - \frac{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}.\end{aligned} \tag{S3.3}$$

For the second term, we have

$$\begin{aligned}&\left|\operatorname{tr}\left(\left(\boldsymbol{\Sigma} - \frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right) - (\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\operatorname{tr}\left(\left(\boldsymbol{\Sigma} - \frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{Y}\mathbf{Y}^\top\right)\right| \\ =&\left|\operatorname{tr}\left(\left(\boldsymbol{\Sigma} - \frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\left(\mathbf{P}_{\mathbf{Y}} - (\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^\top\right)\right)\right| \\ \le& 2n\left\|\boldsymbol{\Sigma} - \frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right\|\left\|\mathbf{P}_{\mathbf{Y}} - (\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^\top\right\| \\ =& O_P\left(n(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p)\sqrt{\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}}\right),\end{aligned}$$

where the last inequality follows from von Neumann's trace theorem and the fact $\operatorname{Rank}\left(\mathbf{P}_{\mathbf{Y}} - (\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^\top\right) \le 2n$, and the last equality follows from (S3.2) and the fact $\operatorname{tr}(\boldsymbol{\Sigma}^2)/\operatorname{tr}(\boldsymbol{\Sigma}) \in [\boldsymbol{\lambda}_p, \boldsymbol{\lambda}_1]$. On the other hand, it is straightforward to show that

$$\operatorname{E}\left((\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\operatorname{tr}\left(\left(\boldsymbol{\Sigma} - \frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{Y}\mathbf{Y}^\top\right)\right) = 0,$$

and

$$\begin{aligned}
&\operatorname{Var}\left((\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\operatorname{tr}\left(\left(\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{Y}\mathbf{Y}^\top\right)\right)\\
=&\frac{2n}{\operatorname{tr}^2(\boldsymbol{\Sigma})}\operatorname{tr}\left(\boldsymbol{\Sigma}^2-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\boldsymbol{\Sigma}\right)^2\\
=&\frac{2n}{\operatorname{tr}^2(\boldsymbol{\Sigma})}\sum_{i=1}^p\boldsymbol{\lambda}_i^2\left(\boldsymbol{\lambda}_i-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\right)^2\\
\leq&\frac{2n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p)^2}{\operatorname{tr}(\boldsymbol{\Sigma})}.
\end{aligned}$$

Thus,

$$\operatorname{tr}\left(\left(\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)=O_P\left(n(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p)\sqrt{\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}}\right).$$

Then the first statement follows from the last display and (S3.3).

Next we deal with $\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)^2$. We have

$$\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)^2=\operatorname{tr}(\boldsymbol{\Sigma}^2)-2\operatorname{tr}(\boldsymbol{\Sigma}^2\mathbf{P}_{\mathbf{Y}})+\operatorname{tr}((\boldsymbol{\Sigma}\mathbf{P}_{\mathbf{Y}})^2).$$

From von Neumann's trace theorem, the second term satisfies

$$\left|\operatorname{tr}(\boldsymbol{\Sigma}^2\mathbf{P}_{\mathbf{Y}})-\frac{n\operatorname{tr}^2(\boldsymbol{\Sigma}^2)}{\operatorname{tr}^2(\boldsymbol{\Sigma})}\right|=\left|\operatorname{tr}\left(\left(\boldsymbol{\Sigma}^2-\frac{\operatorname{tr}^2(\boldsymbol{\Sigma}^2)}{\operatorname{tr}^2(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)\right|\leq n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p),$$

and the third term satisfies

$$\begin{aligned}
\left|\operatorname{tr}((\boldsymbol{\Sigma}\mathbf{P}_{\mathbf{Y}})^2)-\frac{n\operatorname{tr}^2(\boldsymbol{\Sigma}^2)}{\operatorname{tr}^2(\boldsymbol{\Sigma})}\right|=&\left|\operatorname{tr}\left(\left(\boldsymbol{\Sigma}+\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\left(\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)\right|\\
\leq&2n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p).
\end{aligned}$$

This completes the proof of the second statement. □

Proof of Theorem 1. In the current context, Lemma 6 implies that

$$\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) = \operatorname{tr}(\boldsymbol{\Sigma}) - \frac{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})} + o_P(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}), \tag{S3.4}$$

$$\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)^2 = (1 + o_P(1))\operatorname{tr}(\boldsymbol{\Sigma}^2). \tag{S3.5}$$

The fact $\lambda_1\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \leq \boldsymbol{\lambda}_1$ and (S3.5) imply that the first term of (S3.1) satisfies the Lyapunov condition

$$\frac{\lambda_1\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)}{\sqrt{\operatorname{tr}\left(((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}))^2\right)}} \leq \frac{\boldsymbol{\lambda}_1}{\sqrt{(1 + o_P(1))\operatorname{tr}^2(\boldsymbol{\Sigma})}} \xrightarrow{P} 0.$$

From Lemma 2, we have

$$\frac{\mathbf{Z}^{\dagger\top}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}^\dagger - \operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)^2}} \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}.$$

Then it follows from (S3.4), (S3.5) and Slutsky's theorem that

$$\frac{\mathbf{Z}^{\dagger\top}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}^\dagger - \left(\operatorname{tr}(\boldsymbol{\Sigma}) - n\operatorname{tr}(\boldsymbol{\Sigma}^2)/\operatorname{tr}(\boldsymbol{\Sigma})\right)\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}} \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}. \tag{S3.6}$$

Next we consider the second term of (S3.1). Note that

$$\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Theta}\mathbf{C} - \mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}\right\| = \left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\|.$$

We have

$$\begin{aligned}
&\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C} - \operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\| \\
\leq &\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\|\left\|(\mathbf{Y}^\top\mathbf{Y})^{-1} - \operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{I}_n\right\| \\
\leq &\left\|\operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\|\left\|(\mathbf{Y}^\top\mathbf{Y})^{-1}\right\|\left\|\mathbf{Y}^\top\mathbf{Y} - \operatorname{tr}(\boldsymbol{\Sigma})\mathbf{I}_n\right\|.
\end{aligned}$$

From Lemma 5, we have

$$\begin{aligned}\left\|\mathbf{Y}^\top\mathbf{Y}-\operatorname{tr}(\boldsymbol{\Sigma})\mathbf{I}_n\right\| =& \left\|\mathbf{Z}^\top\boldsymbol{\Lambda}\mathbf{Z}-\operatorname{tr}(\boldsymbol{\Sigma})\mathbf{I}_n\right\|\\ =&O_P(\sqrt{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}+n\boldsymbol{\lambda}_1)\\ =&o_P(\operatorname{tr}(\boldsymbol{\Sigma})).\end{aligned}$$

Then $\left\|(\mathbf{Y}^\top\mathbf{Y})^{-1}\right\| = \lambda_n^{-1}(\mathbf{Z}^\top\boldsymbol{\Lambda}\mathbf{Z}) = (1+o_P(1))\operatorname{tr}(\boldsymbol{\Sigma})$. Therefore,

$$\begin{aligned}&\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}-\operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\|\\ =&o_P\left(\left\|\operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\|\right).\end{aligned}$$

Note that the columns of $\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}=\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}$ are iid $\mathcal{N}_{k-1}(\mathbf{0}_{k-1},\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}\mathbf{C})$ random vectors. Hence we can write $\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}=(\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}\mathbf{C})^{1/2}\mathbf{Z}^*$, where $\mathbf{Z}^*$ is a $(k-1)\times n$ random matrix with iid $\mathcal{N}(0,1)$ entries. Then

$$\begin{aligned}&\left\|\operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}-\frac{n}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}\mathbf{C}\right\|\\ \leq&\frac{n}{\operatorname{tr}(\boldsymbol{\Sigma})}\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}\mathbf{C}\right\|\left\|n^{-1}\mathbf{Z}^*\mathbf{Z}^{*\top}-\mathbf{I}_{k-1}\right\|\\ =&o_P\left(\frac{n}{\operatorname{tr}(\boldsymbol{\Sigma})}\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}\mathbf{C}\right\|\right),\end{aligned}$$

where the last equality follows from the law of large numbers. Combine the above arguments, we have

$$\begin{aligned}\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Theta}\mathbf{C}-\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}\right\| =&(1+o_P(1))\frac{n}{\operatorname{tr}(\boldsymbol{\Sigma})}\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}\mathbf{C}\right\|\\ \leq&(1+o_P(1))\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}\right\|\\ =&o_P\left(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}\right).\end{aligned}\tag{S3.7}$$

Now we deal with the cross term of (S3.1). Note that

$$\begin{aligned}\mathrm{E}[\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\|_F^2|\mathbf{Y}] =&(k-1)\operatorname{tr}\left(\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}\right)\\ \leq&(k-1)\boldsymbol{\lambda}_1\operatorname{tr}\left(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\| =&O_P\left(\sqrt{\boldsymbol{\lambda}_1\operatorname{tr}(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C})}\right)\\ =&o_P\left(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\right),\end{aligned}\tag{S3.8}$$

where the last equality follows from the conditions $\boldsymbol{\lambda}_1/\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\to 0$ and $\operatorname{tr}\left(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right)\leq(k-1)\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\| = O(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)})$.

It follows from (S3.7), (S3.8) and Weyl's inequality that

$$\begin{aligned}&\left|T(\mathbf{X})-\left(\lambda_1\left(\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger+\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right)\right)\right|\\ \leq&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}-\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}+\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger+\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}\right\|\\ \leq&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}-\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\|+2\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\right\|\\ =&o_P\left(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\right).\end{aligned}$$

But (S3.6) implies that

$$\begin{aligned}&\frac{\lambda_1\left(\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger+\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right)-(\operatorname{tr}(\mathbf{\Sigma})-n\operatorname{tr}(\mathbf{\Sigma}^2)/\operatorname{tr}(\mathbf{\Sigma}))}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}\\ =&\lambda_1\left(\frac{\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger-(\operatorname{tr}(\mathbf{\Sigma})-n\operatorname{tr}(\mathbf{\Sigma}^2)/\operatorname{tr}(\mathbf{\Sigma}))\,\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}+\frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}\right)\\ \sim&\lambda_1\left(\mathbf{W}_{k-1}+\frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}\right)+o_P(1).\end{aligned}$$

This completes the proof.

□

Proof of Corollary 1. It is straightforward to show that $\mathrm{E}\,\widehat{\operatorname{tr}(\mathbf{\Sigma})} = \operatorname{tr}(\mathbf{\Sigma})$ and $\operatorname{Var}\left(\widehat{\operatorname{tr}(\mathbf{\Sigma})}\right) = 2n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)$. Then $\widehat{\operatorname{tr}(\mathbf{\Sigma})} = \operatorname{tr}(\mathbf{\Sigma}) + O_P(\sqrt{n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)})$. Let $Z_1, \ldots, Z_n$ be the columns of $\mathbf{Z}$. Then we have

$$
\begin{aligned}
\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)} =& n^{-2}\operatorname{tr}(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z} - n^{-1}\operatorname{tr}(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z})\mathbf{I}_n)^2 \\
=& n^{-2}\sum_{i=1}^n (Z_i^\top\mathbf{\Lambda}Z_i - n^{-1}\sum_{i=1}^n Z_i^\top\mathbf{\Lambda}Z_i)^2 + 2n^{-2}\sum_{1\le i<j\le n}(Z_i^\top\mathbf{\Lambda}Z_i)^2.
\end{aligned}
$$

It can be seen that $n^{-2}\sum_{i=1}^n (Z_i^\top\mathbf{\Lambda}Z_i - n^{-1}\sum_{i=1}^n Z_i^\top\mathbf{\Lambda}Z_i)^2 = O_P(n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2))$. On the other hand, we have $\mathrm{E}\,2\sum_{1\le i<j\le n}(Z_i^\top\mathbf{\Lambda}Z_i)^2 = n(n-1)\operatorname{tr}(\mathbf{\Sigma}^2)$. Furthermore, Chen et al. (2010), Proposition A.2 implies that

$$
\operatorname{Var}\left(2\sum_{1\le i<j\le n}(Z_i^\top\mathbf{\Lambda}Z_j)^2\right) = O\left(n^2\operatorname{tr}^2(\mathbf{\Sigma}^2) + n^3\operatorname{tr}(\mathbf{\Sigma}^4)\right) = O\left(n^3\operatorname{tr}^2(\mathbf{\Sigma}^2)\right).
$$

Hence $\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)} = (1 + O_P(n^{-1/2}))\operatorname{tr}(\mathbf{\Sigma}^2)$.

Thus, we have

$$
\begin{aligned}
&\widehat{\operatorname{tr}(\mathbf{\Sigma})} - n\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)}/\widehat{\operatorname{tr}(\mathbf{\Sigma})} \\
=& \operatorname{tr}(\mathbf{\Sigma}) + O_P(\sqrt{n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)}) - \frac{n\operatorname{tr}(\mathbf{\Sigma}^2)(1 + O_P(n^{-1/2}))}{\operatorname{tr}(\mathbf{\Sigma})(1 + O_P(\sqrt{n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)/\operatorname{tr}^2(\mathbf{\Sigma})}))} \\
=& \operatorname{tr}(\mathbf{\Sigma}) + O_P(\sqrt{n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)}) - \frac{n\operatorname{tr}(\mathbf{\Sigma}^2)}{\operatorname{tr}(\mathbf{\Sigma})}\left(1 + O_P\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\operatorname{tr}(\mathbf{\Sigma}^2)}{n\operatorname{tr}^2(\mathbf{\Sigma})}}\right)\right) \\
=& \operatorname{tr}(\mathbf{\Sigma}) - \frac{n\operatorname{tr}(\mathbf{\Sigma}^2)}{\operatorname{tr}(\mathbf{\Sigma})} + o_P(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}).
\end{aligned}
$$

Therefore,

$$
Q_1 = \frac{T(\mathbf{X}) - (\operatorname{tr}(\mathbf{\Sigma}) - n\operatorname{tr}(\mathbf{\Sigma}^2)/\operatorname{tr}(\mathbf{\Sigma}))}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}} + o_P(1).
$$

Then the conclusion follows from Theorem 1. $\square$

Lemma 7. *Suppose that $r = o(n)$, $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$, $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. Then uniformly for $i = 1, \ldots, r$,*

$$\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) = n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\left(1 + O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}} + \sqrt{\frac{r}{n}}\right)\right).$$

Proof. Note that

$$(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}) = (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2})(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2}). \quad \text{(S3.9)}$$

We first deal with $(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})$. Under the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$, Proposition 3 implies that

$$\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^\top - \mathbf{P}_{\mathbf{Y},1}^\dagger\| = O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right).$$

From the decomposition

$$\begin{aligned}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}) =& (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger) + (\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\\ &+ (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\mathbf{\Sigma}(\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},1}) + (\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},1}),\end{aligned}$$

we have

$$\begin{aligned}&\left\|(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}) - (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\right\|\\ \le& 2\left\|\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},1}\right\|\left\|\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\right\| + \boldsymbol{\lambda}_1\|\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},1}\|^2.\\ =& O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right)\left\|\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\right\| + O_P\left(\frac{\operatorname{tr}^2(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}\right).\end{aligned}$$

Note that

$$\begin{aligned}\left\|\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| =& \left\|\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top-\mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top-\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{Q}\mathbf{U}_1^\top\right\|\\
\leq&\boldsymbol{\lambda}_{r+1}+\left\|\boldsymbol{\Lambda}_1\mathbf{Q}^\top\right\|+\boldsymbol{\lambda}_{r+1}\left\|\mathbf{Q}\right\|\\
=&\boldsymbol{\lambda}_{r+1}+\left\|\boldsymbol{\Lambda}_1^{1/2}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1/2}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\right\|+\boldsymbol{\lambda}_{r+1}\|\mathbf{Q}^\top\mathbf{Q}\|^{1/2}\\
\leq&\boldsymbol{\lambda}_{r+1}+\boldsymbol{\lambda}_1^{1/2}\left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1/2}\right\|\left\|\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\|^{1/2}+\boldsymbol{\lambda}_{r+1}\|\mathbf{Q}^\top\mathbf{Q}\|^{1/2}\\
=&O_P\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n}}\right),\end{aligned}$$

where the last equality follows from (S2.9), (S2.11) and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2)\to 0$. Thus,

$$\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})-(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\|=O_P\left(\frac{\operatorname{tr}^{3/2}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right).\tag{S3.10}$$

From the decomposition

$$\begin{aligned}&(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\\
=&\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top+\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top-\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{Q}\mathbf{U}_1^\top-\mathbf{U}_1\mathbf{Q}^\top\boldsymbol{\Lambda}_2\mathbf{U}_2^\top+\mathbf{U}_1\mathbf{Q}^\top\boldsymbol{\Lambda}_2\mathbf{Q}\mathbf{U}_1^\top,\end{aligned}$$

we have

$$\begin{aligned}\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})-\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top\right\|\leq&\boldsymbol{\lambda}_{r+1}(1+2\|\mathbf{Q}^\top\mathbf{Q}\|^{1/2}+\|\mathbf{Q}^\top\mathbf{Q}\|)\\
=&O_P\left(\boldsymbol{\lambda}_{r+1}\right),\end{aligned}\tag{S3.11}$$

where the last equality follows from (S2.11). Note that $\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top=$

$\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top$. We have

$$\begin{aligned}
&\left\|\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top - n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\| \\
\leq&\left\|\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\| \left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1} - n^{-1}\mathbf{I}_r\right\| \\
\leq&\left\|\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\right\| \left\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\right\| \\
=&O_P\left(\frac{r^{1/2}\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n^{3/2}}\right),
\end{aligned} \tag{S3.12}$$

where the last equality follows from (S2.3) and (S2.9). From (S3.9), (S3.10), (S3.11) and (S3.12), we obtain that

$$\begin{aligned}
&\left\|(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}) - n^{-1}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2})\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2})\right\| \\
=&O_P\left(\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)} + \sqrt{\frac{r}{n}}\right)\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n}\right).
\end{aligned}$$

Thus, the last display, together with Weyl's inequality, implies that uniformly for $i = 1, \ldots, r$,

$$\begin{aligned}
\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) =& n^{-1}\lambda_i\left(\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top(\mathbf{I} - \mathbf{P}_{\mathbf{Y},2})\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right) \\
&+ O_P\left(\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)} + \sqrt{\frac{r}{n}}\right)\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n}\right).
\end{aligned}$$

Note that

$$
\begin{aligned}
&\Big\| n^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top(\mathbf{I}-\mathbf{P}_{\mathbf{Y},2})\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\\
&-\Big(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r-(n\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\Big)\Big\|\\
\le&\left\|n^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}-n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\|\\
&+n^{-1}\left\|\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\|\left\|\mathbf{P}_{\mathbf{Y},2}-(\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\right\|\\
=&O_P\left(\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}+\sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}\right)\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n}\right),
\end{aligned}
$$

where the last equality follows from (S2.9) and Proposition 4. Then it follows from Weyl's inequality that uniformly for $i=1,\ldots,r$,

$$
\begin{aligned}
&\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\\
=&n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)-(n\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\lambda_{r+1-i}\left(\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right)\\
&+O_P\left(\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}+\sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}+\sqrt{\frac{r}{n}}\right)\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n}\right).
\end{aligned}
\tag{S3.13}
$$

Now we deal with the matrix $\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}$. Note that $\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}$ and $\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ both have iid $\mathcal{N}(0,1)$ entries and they are mutually independent. Then Lemma 5 implies that

$$
\begin{aligned}
&\left\|\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}-\operatorname{tr}(\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2)\mathbf{I}_r\right\|\\
=&O_P\left(\sqrt{r\operatorname{tr}(\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2)^2}+r\left\|\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\right\|\right).
\end{aligned}
$$

By some algebra, we have

$$\begin{aligned}&\left\|\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1} \tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} - \operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}) \mathbf{I}_r\right\| \\ =&O_P\left(\sqrt{r\left\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right\| \operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1})} + r\left\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right\|\right).\end{aligned}$$

Since $\operatorname{E}\operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}) = (n-r)\operatorname{tr}(\mathbf{\Lambda}_2^2)$, we have

$$\operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}) = O_P\left(n\operatorname{tr}(\mathbf{\Lambda}_2^2)\right) = O_P\left(n\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\mathbf{\Lambda}_2)\right).$$

On the other hand, Lemma 5 implies that

$$\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\| = O_P\left(\operatorname{tr}(\mathbf{\Lambda}_2^2) + n\boldsymbol{\lambda}_{r+1}^2\right) = O_P\left(\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\mathbf{\Lambda}_2)\right).$$

Combine these bounds, we have

$$\left\|\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1} \tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} - n\operatorname{tr}(\mathbf{\Lambda}_2^2)\mathbf{I}_r\right\| = O_P\left(\sqrt{rn}\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\mathbf{\Lambda}_2)\right).$$

The last display, combined with Weyl's inequality, implies that uniformly for $i = 1, \ldots, r$,

$$(n\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\lambda_i\left(\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1} \tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1}\right) = O_P(\boldsymbol{\lambda}_{r+1}).$$

Then (S3.13) and the last display implies that uniformly for $i = 1, \ldots, r$,

$$\begin{aligned}&\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \\ =&n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2) + O_P\left(\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}} + \sqrt{\frac{r}{n}}\right)\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n}\right).\end{aligned}$$

This completes the proof.

$\square$

Lemma 8. *Suppose that* $r = o(n)$, $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$, $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. *Then*

$$\begin{aligned}&\sum_{i=r+1}^{p} \lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\\ =&\operatorname{tr}(\mathbf{\Lambda}_2) - \frac{n\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)} + O_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}\right) + r\boldsymbol{\lambda}_{r+1}\right).\end{aligned}$$

Proof. Write $\mathbf{\Sigma} = \mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^\top + \mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top$. Note that $\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^\top$ is of rank r. Then Weyl's inequality implies that for $i = r+1, \ldots, p$,

$$\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \geq \lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right), \quad \text{(S3.14)}$$

$$\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \leq \lambda_{i-r}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right). \quad \text{(S3.15)}$$

Hence we have

$$\begin{aligned}&\left|\sum_{i=r+1}^{p} \lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) - \operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\right|\\ \leq& r\lambda_1\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\\ \leq& r\boldsymbol{\lambda}_{r+1}.\end{aligned} \quad \text{(S3.16)}$$

Write

$$\begin{aligned}&\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\\ =&\operatorname{tr}\left(\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\right)\\ =&\operatorname{tr}(\mathbf{\Lambda}_2) - \operatorname{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right) - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\operatorname{tr}\left(\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right).\end{aligned} \quad \text{(S3.17)}$$

For the third term, note that $\operatorname{tr}\left(\mathbf{U}_2^\top \mathbf{P}_{\mathbf{Y}} \mathbf{U}_2\right) = \operatorname{tr}(\mathbf{P}_{\mathbf{Y}}) - \operatorname{tr}\left(\mathbf{P}_{\mathbf{Y}} \mathbf{U}_1 \mathbf{U}_1^\top\right)$. Since $\mathbf{P}_{\mathbf{Y}}$ is of rank n and $\mathbf{U}_1$ is of rank r, we have

$$|\operatorname{tr}\left(\mathbf{U}_2^\top \mathbf{P}_{\mathbf{Y}} \mathbf{U}_2\right) - n| \le r. \tag{S3.18}$$

Next we deal with the second term. We have

$$\begin{aligned}
&\left|\operatorname{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top \mathbf{P}_{\mathbf{Y}} \mathbf{U}_2\right) - \operatorname{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top \left(\mathbf{P}_{\mathbf{Y},1}^\dagger + \mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right)\right| \\
=&\left|\operatorname{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top \left(\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right)\right|.
\end{aligned}$$

Since $\operatorname{tr}(\mathbf{\Lambda}_2^2)/\operatorname{tr}(\mathbf{\Lambda}_2) \in [\boldsymbol{\lambda}_p, \boldsymbol{\lambda}_{r+1}]$, we have $\|\mathbf{\Lambda}_2 - (\operatorname{tr}(\mathbf{\Lambda}_2^2)/\operatorname{tr}(\mathbf{\Lambda}_2))\mathbf{I}_{p-r}\| \le \boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p$. Also note that the rank of the matrix $\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},2}^\dagger$ is at most $2n$. Therefore, von Neumann's trace theorem implies that

$$\begin{aligned}
&\left|\operatorname{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top \left(\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right)\right| \\
\le& 2n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left\|\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},2}^\dagger\right\| \\
\le& 2n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1}^\dagger\right\| + \left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}_{\mathbf{Y},2}^\dagger\right\|\right) \\
=& O_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}\right)\right),
\end{aligned} \tag{S3.19}$$

where the last equality follows from Proposition 3 and Proposition 4. Note that

$$\begin{aligned}
&\operatorname{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top \left(\mathbf{P}_{\mathbf{Y},1}^\dagger + \mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right) \\
=&\operatorname{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top \mathbf{P}_{\mathbf{Y},2}^\dagger \mathbf{U}_2\right) \\
=&\frac{1}{\operatorname{tr}(\mathbf{\Lambda}_2)}\operatorname{tr}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \left(\mathbf{\Lambda}_2^2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{\Lambda}_2\right)\mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)
\end{aligned}$$

It is straightforward to show that

$$\mathrm{E}\,\mathrm{tr}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \left(\mathbf{\Lambda}_2^2 - \frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{\Lambda}_2\right) \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right) = 0,$$

and

$$\begin{aligned}
\mathrm{Var}\left(\mathrm{tr}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \left(\mathbf{\Lambda}_2^2 - \frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{\Lambda}_2\right) \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)\right) =& 2(n-r)\,\mathrm{tr}\left(\mathbf{\Lambda}_2^2 - \frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{\Lambda}_2\right)^2 \\
\leq& 2n\,\mathrm{tr}(\mathbf{\Lambda}_2^2)(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)^2 \\
\leq& 2n\boldsymbol{\lambda}_{r+1}\,\mathrm{tr}(\mathbf{\Lambda}_2)(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)^2.
\end{aligned}$$

Thus,

$$\mathrm{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right) \mathbf{U}_2^\top \left(\mathbf{P}_{\mathbf{Y},1}^\dagger + \mathbf{P}_{\mathbf{Y},2}^\dagger\right) \mathbf{U}_2\right) = O_P\left((\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\mathrm{tr}(\mathbf{\Lambda}_2)}}\right).$$

The last display, combined with (S3.19), leads to

$$\mathrm{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right) \mathbf{U}_2^\top \mathbf{P}_{\mathbf{Y}} \mathbf{U}_2\right) = O_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\mathrm{tr}(\mathbf{\Lambda}_2)}}\right)\right).$$

It then follows from (S3.17), (S3.18) and the last display that

$$\begin{aligned}
&\mathrm{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \\
=& \mathrm{tr}(\mathbf{\Lambda}_2) - \frac{n\,\mathrm{tr}(\mathbf{\Lambda}_2^2)}{\mathrm{tr}(\mathbf{\Lambda}_2)} + O_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\mathrm{tr}(\mathbf{\Lambda}_2)}}\right) + r\boldsymbol{\lambda}_{r+1}\right).
\end{aligned}$$

Then the conclusion follows from (S3.16) and the last display. □

Lemma 9. *Suppose $p > n$, we have*

$$\sum_{i=r+1}^{p} \lambda_i^2\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) = \mathrm{tr}(\mathbf{\Lambda}_2^2) - \frac{n\,\mathrm{tr}^2(\mathbf{\Lambda}_2^2)}{\mathrm{tr}^2(\mathbf{\Lambda}_2)} + O_P\left(n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p) + r\boldsymbol{\lambda}_{r+1}^2\right).$$

Proof. From (S3.14) and (S3.15), we have

$$\left|\sum_{i=r+1}^{p} \lambda_i^2\left((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\right)-\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\right)^2\right|$$
$$\le r\lambda_1^2\left((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\right)$$
$$\le r\boldsymbol{\lambda}_{r+1}^2. \tag{S3.20}$$

It is straightforward to show that

$$\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\right)^2=\operatorname{tr}(\boldsymbol{\Lambda}_2^2)-2\operatorname{tr}(\boldsymbol{\Lambda}_2^2\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2)+\operatorname{tr}(\boldsymbol{\Lambda}_2\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2)^2.$$

For the second term, we have

$$\left|\operatorname{tr}(\boldsymbol{\Lambda}_2^2\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2)-\frac{\operatorname{tr}^2(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)}\operatorname{tr}(\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2)\right|=\left|\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2^2-\frac{\operatorname{tr}^2(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2\right)\right|$$
$$\le n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1}-\boldsymbol{\lambda}_p),$$

where the last equality follows from von Neumann's trace theorem. The last display, combined with (S3.18), implies that

$$\operatorname{tr}(\boldsymbol{\Lambda}_2^2\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2)=\frac{n\operatorname{tr}^2(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)}+O_P\left(n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1}-\boldsymbol{\lambda}_p)+r\boldsymbol{\lambda}_{r+1}^2\right).$$

For the third term, von Neumann's trace theorem implies that

$$\left|\operatorname{tr}(\boldsymbol{\Lambda}_2\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2)^2-\frac{\operatorname{tr}^2(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)}\operatorname{tr}(\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2)^2\right|$$
$$=\left|\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2-\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2\left(\boldsymbol{\Lambda}_2+\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\mathbf{P}_\mathbf{Y}\mathbf{U}_2\right)\right|$$
$$\le 2n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1}-\boldsymbol{\lambda}_p).$$

Note that

$$\begin{aligned}\operatorname{tr}(\mathbf{U}_2^\top \mathbf{P}_\mathbf{Y} \mathbf{U}_2)^2 =& \operatorname{tr}\left(\mathbf{P}_\mathbf{Y} - \mathbf{P}_\mathbf{Y}\mathbf{U}_1\mathbf{U}_1^\top\right)^2\\ =& n - 2\operatorname{tr}(\mathbf{P}_\mathbf{Y}\mathbf{U}_1\mathbf{U}_1^\top) + \operatorname{tr}(\mathbf{P}_\mathbf{Y}\mathbf{U}_1\mathbf{U}_1^\top)^2\\ =& n + O_P(r).\end{aligned}$$

Therefore, the third term satisfies

$$\operatorname{tr}(\boldsymbol{\Lambda}_2\mathbf{U}_2^\top \mathbf{P}_\mathbf{Y} \mathbf{U}_2)^2 = \frac{n\operatorname{tr}^2(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)} + O_P\left(n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p) + r\boldsymbol{\lambda}_{r+1}^2\right).$$

Thus,

$$\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right)^2 = \operatorname{tr}(\boldsymbol{\Lambda}_2^2) - \frac{n\operatorname{tr}^2(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)} + O_P\left(n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p) + r\boldsymbol{\lambda}_{r+1}^2\right).$$

Then the conclusion follows from the last display and (S3.20). □

Proof of Theorem 2. We have

$$\begin{aligned}&\mathbf{Z}^{\dagger\top}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}^\dagger\\ =& \sum_{i=1}^r \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top + \sum_{i=r+1}^p \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top.\end{aligned}$$

From Lemma 7, the first term satisfies

$$\sum_{i=1}^r \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top = (1 + o_P(r^{-1/2}))n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\sum_{i=1}^r \eta_i\eta_i^\top.$$

Then

$$\frac{\sum_{i=1}^r \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top - rn^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_{k-1}}{\sqrt{r}n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)} = \frac{\sum_{i=1}^r \eta_i\eta_i^\top - r\mathbf{I}_{k-1}}{\sqrt{r}} + o_P(1). \tag{S3.21}$$

Next we deal with the term $\sum_{i=r+1}^{p} \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top$. In the current context, Lemma 8 and Lemma 9 imply that

$$\sum_{i=r+1}^{p} \lambda_i\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right) = \operatorname{tr}(\boldsymbol{\Lambda}_2) - \frac{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)} + o_P\left(\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right), \tag{S3.22}$$

$$\sum_{i=r+1}^{p} \lambda_i^2\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right) = (1 + o_P(1))\operatorname{tr}(\boldsymbol{\Lambda}_2^2). \tag{S3.23}$$

By Weyl's inequality, we have

$$\begin{aligned}
&\lambda_{r+1}((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})) \\
=&\lambda_{r+1}\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{U}_1^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}) + (\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right) \\
\le&\lambda_1\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right) \\
\le&\boldsymbol{\lambda}_{r+1}.
\end{aligned}$$

The last display and (S3.22) imply that

$$\frac{\lambda_{r+1}^2\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right)}{\sum_{i=r+1}^{p}\lambda_i^2\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right)} \le \frac{\boldsymbol{\lambda}_{r+1}^2}{(1 + o_P(1))\operatorname{tr}(\boldsymbol{\Lambda}_2^2)} \xrightarrow{P} 0.$$

Then Lemma 2 implies that

$$\frac{\sum_{i=r+1}^{p} \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top - \sum_{i=r+1}^{p} \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\mathbf{I}_{k-1}}{\sqrt{\sum_{i=r+1}^{p}\lambda_i^2((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))}} \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}.$$

The last display, combined with (S3.22) and (S3.23), leads to

$$\frac{\sum_{i=r+1}^{p} \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top - \left(\operatorname{tr}(\boldsymbol{\Lambda}_2) - n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2)\right)\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}} \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}. \tag{S3.24}$$

Note that $\sum_{i=1}^{r} \eta_i\eta_i^\top$ is independent of $\sum_{i=r+1}^{p} \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top$.

Then(S3.21) and (S3.24) implies that

$$\frac{\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger - ((1 + r/n)\operatorname{tr}(\mathbf{\Lambda}_2) - n\operatorname{tr}(\mathbf{\Lambda}_2^2)/\operatorname{tr}(\mathbf{\Lambda}_2))\,\mathbf{I}_{k-1}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}}$$
$$\sim\frac{n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}) + \frac{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}}\mathbf{W}_{k-1} + o_P(1). \tag{S3.25}$$

This completes the proof of the first statement.

Now we prove the second statement. For the second term of (S3.1), we have $\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C} = \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C} - \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_\mathbf{Y}\mathbf{\Theta}\mathbf{C}$. We need to deal with $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_\mathbf{Y}\mathbf{\Theta}\mathbf{C}$. Note that Proposition 3 implies that

$$\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{U}_1\mathbf{U}_1^\top\right\| \le \left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1}^\dagger\right\| + 2\left\|\mathbf{Q}\right\| = o_P(1).$$

It follows from the last display and Proposition 4 that

$$\begin{aligned}
&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_\mathbf{Y}\mathbf{\Theta}\mathbf{C} - \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_1\mathbf{U}_1^\top\mathbf{\Theta}\mathbf{C} - \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\mathbf{\Theta}\mathbf{C}\right\|\\
\le&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},1}\mathbf{\Theta}\mathbf{C} - \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_1\mathbf{U}_1^\top\mathbf{\Theta}\mathbf{C}\right\| + \left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}\mathbf{\Theta}\mathbf{C} - \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\mathbf{\Theta}\mathbf{C}\right\|\\
\le&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\|\left(\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{U}_1\mathbf{U}_1^\top\right\| + \left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}_{\mathbf{Y},2}^\dagger\right\|\right)\\
=&o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right).
\end{aligned}$$

We have

$$\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\mathbf{\Theta}\mathbf{C} = (\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\,\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}.$$

Note that $\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ is a $(p-r)\times(n-r)$ matrix with iid $\mathcal{N}(0,1)$ entries. Then the columns of $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ are iid $\mathcal{N}_{k-1}(\mathbf{0}_{k-1}, \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C})$

random vectors. Write $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1} = (\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C})^{1/2}\mathbf{Z}^*$, where $\mathbf{Z}^*$ is a $(k-1)\times(n-r)$ random matrix with iid $\mathcal{N}(0,1)$ entries. Then

$$\begin{aligned}&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\mathbf{\Theta}\mathbf{C} - \frac{n}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right\| \\ \leq&\frac{n}{\operatorname{tr}(\mathbf{\Lambda}_2)}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right\|\left\|n^{-1}\mathbf{Z}^*\mathbf{Z}^{*\top} - \mathbf{I}_{k-1}\right\| \\ =&o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right),\end{aligned}$$

where the last equality follows from the law of large numbers, the local alternative condition and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$. But

$$\frac{n}{\operatorname{tr}(\mathbf{\Lambda}_2)}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right\| \leq \frac{n\boldsymbol{\lambda}_2}{\operatorname{tr}(\mathbf{\Lambda}_2)}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\| = o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right).$$

Hence $\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\mathbf{\Theta}\mathbf{C}\right\| = o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right)$. Consequently, $\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y}}\mathbf{\Theta}\mathbf{C} - \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_1\mathbf{U}_1^\top\mathbf{\Theta}\mathbf{C}\right\| = o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right)$. Thus, the second term of (S3.1) satisfies

$$\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C} - \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right\| = o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right). \tag{S3.26}$$

Next we consider the cross term of (S3.1). Note that

$$\begin{aligned}&\mathrm{E}[\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\|_F^2|\mathbf{Y}] \\ =&(k-1)\operatorname{tr}(\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C}) \\ \leq&(k-1)\lambda_1\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\operatorname{tr}(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}) \\ =&O_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\|\right),\end{aligned}$$

where the last equality follows from Lemma 7. Under the condition $r \to \infty$ or $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}) \to 0$, we have $n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2) = o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right)$. Therefore,

$$\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\| = o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right).$$

It follows from the last display, (S3.26) and Weyl's inequality that

$$\begin{aligned}
&\left|T(\mathbf{X}) - \lambda_1\left(\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\dagger(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger + \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right)\right| \\
=&o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}\right).
\end{aligned}$$

Then the second statement follows from the last display and (S3.25).

$\square$

Proof of Corollary 2. From Proposition 2, we have

$$rn^{-2}(\widehat{\operatorname{tr}(\mathbf{\Lambda}_2)})^2 + \widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)} = (1 + o_P(1))(rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)),$$

and

$$
\begin{aligned}
&(1+r/n)\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} - n\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}/\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} \\
=&(1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) + O_P\left(r\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right) \\
&- \frac{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)\left(1+O_P\left(r/n + r\boldsymbol{\lambda}_{r+1}^2/\operatorname{tr}(\boldsymbol{\Lambda}_2^2)\right)\right)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)\left(1+O_P\left(r\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/n\operatorname{tr}^2(\boldsymbol{\Lambda}_2)} + r\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2)\right)\right)} \\
=&(1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) + O_P\left(r\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right) \\
&- \frac{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\left(1+O_P\left(\frac{r}{n} + \frac{r\boldsymbol{\lambda}_{r+1}^2}{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)} + r\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n\operatorname{tr}^2(\boldsymbol{\Lambda}_2)}} + \frac{r\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\right)\right) \\
=&(1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) - \frac{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)} + o_P\left(\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right).
\end{aligned}
$$

Therefore,

$$
Q_2 = \frac{T(\mathbf{X}) - ((1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) - n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2))}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}} + o_P(1).
$$

On the other hand, it is not hard to see that the ratio consistency of $\widehat{\operatorname{tr}(\boldsymbol{\lambda}_2)}$ and $\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}$ imply $F_2^{-1}(1-\alpha;\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)},\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}) = F_2^{-1}(1-\alpha;\operatorname{tr}(\boldsymbol{\Lambda}_2),\operatorname{tr}(\boldsymbol{\Lambda}_2^2)) + o_P(1)$. Then the conclusion follows from Theorem 2 and Slutsky's theorem.

□

Proof of Proposition 5. Under the conditions of Theorem 1, we have $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$. From Lemma 5 and Weyl's inequality, we have

$$
\lambda_1(\hat{\boldsymbol{\Sigma}}) = n^{-1}\lambda_1(\mathbf{Z}^\top\boldsymbol{\Lambda}\mathbf{Z}) = n^{-1}\operatorname{tr}(\boldsymbol{\Sigma}) + O_P\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{n}} + \boldsymbol{\lambda}_1\right) = (1+o_P(1))n^{-1}\operatorname{tr}(\boldsymbol{\Sigma}).
$$

From the proof of Corollary 1, we have $\operatorname{tr}(\hat{\mathbf{\Sigma}}) = (1+o_P(1))\operatorname{tr}(\mathbf{\Sigma})$. Therefore,

$$\frac{n\lambda_1(\hat{\mathbf{\Sigma}})}{\operatorname{tr}(\hat{\mathbf{\Sigma}})} \xrightarrow{P} 1.$$

This completes the proof of (i).

Now we prove (ii). Under the conditions of Theorem 2, Proposition 1 implies that

$$\begin{aligned}\frac{n\lambda_1(\hat{\mathbf{\Sigma}})}{\operatorname{tr}(\hat{\mathbf{\Sigma}})} =& \frac{n\lambda_1(\hat{\mathbf{\Sigma}})}{\sum_{i=1}^r \lambda_i(\hat{\mathbf{\Sigma}}) + \sum_{i=r+1}^n \lambda_i(\hat{\mathbf{\Sigma}})}\\ =&(1+o_P(1))\frac{n\boldsymbol{\lambda}_1 + \operatorname{tr}(\mathbf{\Lambda}_2)}{\sum_{i=1}^r \boldsymbol{\lambda}_i + \operatorname{tr}(\mathbf{\Lambda}_2)}\\ \geq&(1+o_P(1))\frac{n\boldsymbol{\lambda}_1}{r\boldsymbol{\lambda}_1 + \operatorname{tr}(\mathbf{\Lambda}_2)} \xrightarrow{P} \infty.\end{aligned}$$

It follows that

$$\Pr\left(\frac{n\lambda_1(\hat{\mathbf{\Sigma}})}{\operatorname{tr}(\hat{\mathbf{\Sigma}})} < \tau\right) \to 0.$$

Next we consider the consistency of $\hat{r}$. Note that

$$\{\hat{r} = r\} = \left\{\frac{n\lambda_{i+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=i+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} \geq \tau, i = 1, \dots, r-1\right\} \cap \left\{\frac{n\lambda_{r+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=r+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} < \tau\right\}.$$

But Proposition 1 implies that uniformly for $i = 1, \dots, r-1$,

$$\begin{aligned}\frac{n\lambda_{i+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=i+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} \geq& \frac{n\lambda_{i+1}(\hat{\mathbf{\Sigma}})}{(r-i)\lambda_{i+1}(\hat{\mathbf{\Sigma}}) + \sum_{j=r+1}^n \lambda_j(\hat{\mathbf{\Sigma}})}\\ =&(1+o_P(1))\frac{n\boldsymbol{\lambda}_{i+1} + \operatorname{tr}(\mathbf{\Lambda}_2)}{(r-i)\boldsymbol{\lambda}_{i+1} + (1-i/n)\operatorname{tr}(\mathbf{\Lambda}_2)} \xrightarrow{P} \infty.\end{aligned}$$

Thus, we only need to prove that

$$\Pr\left(\frac{n\lambda_{r+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=r+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} < \tau\right) \to 1.$$

Weyl' inequality implies that $n\lambda_{r+1}(\hat{\mathbf{\Sigma}}) = \lambda_{r+1}(\mathbf{Z}_1^\top \mathbf{\Lambda}_1 \mathbf{Z}_1 + \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2) \leq \lambda_1(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2)$. Then using Lemma 5, we have $n\lambda_{r+1}(\hat{\mathbf{\Sigma}}) \leq (1+o_P(1))\,\mathrm{tr}(\mathbf{\Lambda}_2)$. Thus,

$$\frac{n\lambda_{r+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=r+1}^{n} \lambda_j(\hat{\mathbf{\Sigma}})} \leq (1 + o_P(1)).$$

This completes the proof.

$\square$

References

Cai, T., Z. Ma, and Y. Wu (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability Theory & Related Fields 161*(3-4), 781–815.

Chen, S. X., L. Zhang, and P. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis* (2nd ed.). New York: Cambridge University Press.

Laurent, B. and P. Massart (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics 28*(5), 1302–1338.

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