SUPPLEMENT TO "MINIMAX ESTIMATION IN SPARSE CANONICAL CORRELATION ANALYSIS"

BY Chao Gao[1], Zongming Ma[2], Zhao Ren[1] and Harrison H. Zhou[1]

[1]Yale University and [2]University of Pennsylvania

APPENDIX A: PROOFS OF TECHNICAL RESULTS

In this appendix, we prove Theorem 4 and Lemmas 7 – 12 in order.

A.1. Proof of Theorem 4. We first need a lemma for perturbation bound of square root matrices.

LEMMA 16. *Let A, B be positive semi-definite matrices, and then for any unitarily invariant norm $\|\cdot\|$,*

$$\|A^{1/2} - B^{1/2}\| \leq \frac{1}{\sigma_{\min}(A^{1/2}) + \sigma_{\min}(B^{1/2})}\|A - B\|.$$

PROOF. The proof essentially follows the idea of [27]. Let $D = B - A$ and $X = B^{1/2} - A^{1/2}$. Then we have for every sufficient large $q > 0$,

$$X = E_2 X E_1 + F,$$

where

$$\begin{aligned} E_1 &= (qI + A^{1/2})^{-1}(qI - A^{1/2}), \\ E_2 &= (qI + B^{1/2})^{-1}(qI - A^{1/2}), \\ F &= 2q(qI + B^{1/2})^{-1} D (qI + A^{1/2})^{-1}. \end{aligned}$$

Take the desired norm on both sides, we have

$$\|X\| \leq \|E_2 X E_1\| + \|F\| \leq \|E_1\|_{\mathrm{op}} \|E_2\|_{\mathrm{op}} \|X\| + \|F\|.$$

Here, the first inequality is due the triangle inequality and the second is due to [6, Prop. IV.2.4]. By the proof of Lemma 2.1 in [27], $\|E_i\|_{\mathrm{op}} < 1$ for $i = 1, 2$ when q is sufficiently large, and hence

$$\|X\| \leq \frac{\|F\|}{1 - \|E_1\|_{\mathrm{op}} \|E_2\|_{\mathrm{op}}}.$$

Sending $q \to \infty$ in the last display leads to the desired bound. □

We prove (43) and (44) respectively.

PROOF OF (43).

$$
\begin{aligned}
& \|A_1 D_1 B_1' - \widehat{A}_1 \widehat{D}_1 \widehat{B}_1'\| \\
= \; & \|A_1 A_1' X B_1 B_1' - \widehat{A}_1 \widehat{A}_1' Y \widehat{B}_1 \widehat{B}_1'\| \\
\leq \; & \|A_1 A_1 X (B_1 B_1' - \widehat{B}_1 \widehat{B}_1')\| + \|A_1 A_1' (X - Y) \widehat{B}_1 \widehat{B}_1'\| \\
& + \|(A_1 A_1' - \widehat{A}_1 \widehat{A}_1) Y \widehat{B}_1 \widehat{B}_1'\| \\
\leq \; & (d_1 + \widehat{d}_1) \frac{\sqrt{2}\epsilon}{\delta} + \epsilon,
\end{aligned}
$$

where the last inequality is by Wedin's sin-theta theorem [35]. □

PROOF OF (44). Without loss of generality, we assume $p \geq m$, and hence the columns of B_1 and B_2 span $\mathbb{R}^m$. We first have the decomposition

$$
\begin{aligned}
\widehat{A}_1 \widehat{B}_1' - A_1 B_1' \; = \; & (I - A_1 A_1') \widehat{A}_1 \widehat{B}_1' && (89) \\
& - A_1 D_1^{-1} B_1' (\widehat{B}_1' \widehat{D}_1 \widehat{A}_1' - B_1 D_1 A_1') \widehat{A}_1 \widehat{B}_1' && (90) \\
& + A_1 D_1^{-1} B_1' (\widehat{B}_1 \widehat{D}_1 \widehat{B}_1' - B_1 D_1 B_1'). && (91)
\end{aligned}
$$

We bound each of the three terms above. By Wedin's sin-theta theorem [35], the first term (89) can be bounded by

$$
\begin{aligned}
\|(I - A_1 A_1') \widehat{A}_1 \widehat{B}_1'\| \; = \; & \|(\widehat{A}_1 \widehat{A}_1' - A_1 A_1') \widehat{A}_1 \widehat{B}_1'\| \\
\leq \; & \|\widehat{A}_1 \widehat{A}_1' - A_1 A_1'\| \leq \frac{\sqrt{2}\epsilon}{\delta}.
\end{aligned}
$$

Next, we use (43) to bound (90) by

$$
d_r^{-1} \|\widehat{B}_1' \widehat{D}_1 \widehat{A}_1' - B_1 D_1 A_1'\| \leq \frac{d_1 + \widehat{d}_1}{d_r} \frac{\sqrt{2}\epsilon}{\delta} + \frac{\epsilon}{d_r}.
$$

Lastly, (91) is bounded by

$$
\begin{aligned}
& d_r^{-1} \|\widehat{B}_1 \widehat{D}_1 \widehat{B}_1' - B_1 D_1 B_1'\| \\
\leq \; & d_r^{-1} \|\widehat{B}_1 \widehat{D}_1 \widehat{B}_1' + d_1 \widehat{B}_2 \widehat{B}_2' - B_1 D_1 B_1' - d_1 B_2 B_2'\| \\
& + \frac{d_1}{d_r} \|\widehat{B}_2 \widehat{B}_2' - B_2 B_2'\| \\
\leq \; & d_r^{-2} \|\widehat{B}_1 \widehat{D}_1^2 \widehat{B}_1' + d_1^2 \widehat{B}_2 \widehat{B}_2' - B_1 D_1^2 B_1' - d_1^2 B_2 B_2'\| \\
& + \frac{d_1}{d_r} \|\widehat{B}_2 \widehat{B}_2' - B_2 B_2'\| \\
\leq \; & d_r^{-2} \|\widehat{B}_1 \widehat{D}_1^2 \widehat{B}_1' - B_1 D_1^2 B_1'\| + \left(\frac{d_1}{d_r} + \frac{d_1^2}{d_r^2} \right) \|\widehat{B}_2 \widehat{B}_2' - B_2 B_2'\|, && (92)
\end{aligned}
$$

where we have used Lemma 16 in the second inequality above. The second term of (92) is

$$\left(\frac{d_1}{d_r}+\frac{d_1^2}{d_r^2}\right)\|\widehat{B}_2\widehat{B}_2'-B_2B_2'\|=\left(\frac{d_1}{d_r}+\frac{d_1^2}{d_r^2}\right)\|\widehat{B}_1\widehat{B}_1'-B_1B_1'\|\leq\left(\frac{d_1}{d_r}+\frac{d_1^2}{d_r^2}\right)\frac{\sqrt{2}\epsilon}{\delta},$$

by Wedin's sin-theta theorem [35]. The first term of (92) is bounded by

$$\begin{aligned}&d_r^{-2}\|B_1D_1A_1'(A_1D_1B_1'-\widehat{A}_1'\widehat{D}_1\widehat{B}_1')+(B_1D_1A_1'-\widehat{B}_1\widehat{D}_1\widehat{A}_1')\widehat{A}_1\widehat{D}_1\widehat{B}_1'\|\\ \leq\;&\frac{d_1+\widehat{d}_1}{d_r^2}\|A_1D_1B_1'-\widehat{A}_1'\widehat{D}_1\widehat{B}_1'\|\leq\frac{d_1+\widehat{d}_1}{d_r^2}\left((d_1+\widehat{d}_1)\frac{\sqrt{2}\epsilon}{\delta}+\epsilon\right),\end{aligned}$$

by (43). Combining the bounds above, we have

$$\begin{aligned}&\|\widehat{A}_1\widehat{B}_1'-A_1B_1'\|\\ \leq\;&\left(1+\frac{d_1+\widehat{d}_1}{d_r}+\frac{(d_1+\widehat{d}_1)^2}{d_r^2}+\frac{d_1}{d_r}+\frac{d_1^2}{d_r^2}\right)\frac{\sqrt{2}\epsilon}{\delta}+\frac{1+d_r^{-1}(d_1+\widehat{d}_1)}{d_r}\epsilon\\ \leq\;&\frac{C\epsilon}{\delta},\end{aligned}$$

under the assumption that $d_1\vee\widehat{d}_1\leq\bar{\kappa}d_r$. □

A.2. Proof of Lemma 7. Note that for $i=1,2$,

$$\Sigma_{(i)}=I+\frac{\lambda}{2}\begin{bmatrix}U_{(i)}\\V_{(i)}\end{bmatrix}\begin{bmatrix}U_{(i)}' & V_{(i)}'\end{bmatrix}-\frac{\lambda}{2}\begin{bmatrix}U_{(i)}\\-V_{(i)}\end{bmatrix}\begin{bmatrix}U_{(i)}' & -V_{(i)}'\end{bmatrix}.$$

Thus, $\Sigma_{(i)}$ has two eigenvalues $1\pm\lambda$, both of multiplicity r and the rest are all ones. This, in particular, implies that

$$\det\Sigma_{(1)}=\det\Sigma_{(2)}. \tag{93}$$

Now the KL divergence is

$$\begin{aligned}D(P_{(1)}||P_{(2)})&=\frac{n}{2}\left[\mathsf{Tr}(\Sigma_{(2)}^{-1}\Sigma_{(1)})-(p+m)-\log\det(\Sigma_{(2)}^{-1}\Sigma_{(1)})\right]\\&=\frac{n}{2}\left[\mathsf{Tr}(\Sigma_{(2)}^{-1}\Sigma_{(1)})-(p+m)\right]\\&=\frac{n}{2}\left[\mathsf{Tr}(\Sigma_{(2)}^{-1}(\Sigma_{(1)}-\Sigma_{(2)}))\right]. \qquad (94)\end{aligned}$$

Here, the second equality is due to (93).

Note that $\Sigma_{(1)}-\Sigma_{(2)} = \begin{bmatrix} 0 & U_{(1)}V'_{(1)} - U_{(2)}V'_{(2)} \\ V_{(1)}U'_{(1)} - V_{(2)}U'_{(2)} & 0 \end{bmatrix}$ and that the block inversion formula implies

$$\Sigma_{(2)}^{-1} = \begin{bmatrix} I_p + \frac{\lambda^2}{1-\lambda^2}U_{(2)}U'_{(2)} & -\frac{\lambda}{1-\lambda^2}U_{(2)}V'_{(2)} \\ -\frac{\lambda}{1-\lambda^2}V_{(2)}U'_{(2)} & I_m + \frac{\lambda^2}{1-\lambda^2}V_{(2)}V'_{(2)}. \end{bmatrix}$$

Plugging these expressions into (94), we obtain

$$\begin{aligned} D(P_{(1)}||P_{(2)}) &= \frac{n\lambda^2}{2(1-\lambda^2)}(\mathsf{Tr}(U_{(2)}V'_{(2)}(V_{(2)}U'_{(2)} - V_{(1)}U'_{(1)})) \\ &\qquad + \mathsf{Tr}(V_{(2)}U'_{(2)}(U_{(2)}V'_{(2)} - U_{(1)}V'_{(1)}))) \\ &= \frac{n\lambda^2}{2(1-\lambda^2)}\, 2\,\mathsf{Tr}\left(I_r - V'_{(1)}V_{(2)}U'_{(2)}U_{(1)}\right) \\ &= \frac{n\lambda^2}{2(1-\lambda^2)}\|U_{(1)}V'_{(1)} - U_{(2)}V'_{(2)}\|_{\mathrm{F}}^2. \end{aligned}$$

This completes the proof.

A.3. Proof of Lemma 8. Before stating the proof, we need the following Bernstein's inequality of Gaussian quadratic form.

LEMMA 17. *Let $\{Z_i\}_{1\leq i\leq n}$ be i.i.d. observations from $N(0, I_d)$, and K be a fixed matrix satisfying $\|K\|_{\mathrm{F}} \leq 1$. Then, there exists some $C > 0$, such that*

$$\mathbb{P}\left(\left|\left\langle \frac{1}{n}\sum_{i=1}^n Z_iZ_i' - I_d, K\right\rangle\right| > t\right) \leq \exp(-Cn(t^2\wedge t)),$$

for any $t > 0$.

PROOF. It is sufficient to consider symmetric K because

$$\left\langle \frac{1}{n}\sum_{i=1}^n Z_iZ_i' - I_d, K\right\rangle = \left\langle \frac{1}{n}\sum_{i=1}^n Z_iZ_i' - I_d, \frac{1}{2}(K+K')\right\rangle.$$

Let K has spectral decomposition $K = \sum_{l=1}^d \eta_l q_l q_l'$. Since K has unit Frobenius norm, we have $\sum_{l=1}^d \eta_l^2 = 1$. Then we have

$$\begin{aligned} \left\langle \frac{1}{n}\sum_{i=1}^n Z_iZ_i' - I_d, K\right\rangle &= \left\langle \frac{1}{n}\sum_{i=1}^n Z_iZ_i' - I_d, \sum_{l=1}^d \eta_l q_l q_l'\right\rangle \\ &= \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^d \eta_l(|q_l'Z_i|^2 - 1). \end{aligned}$$

Notice $|q_l' Z_i|^2 - 1$ is centered sub-exponential random variable for each i and l. Moreover, they are independent across i and l because $\{q_l\}_{l=1}^d$ is an orthonormal basis. Applying Bernstein's inequality for sub-exponential variables [31, Prop. 5.16], the proof is complete. $\square$

Now we are ready to state the main proof.

Proof of Lemma 8. Define the class

$$\mathcal{K}(r) = \left\{ K \in \mathbb{R}^{d\times d} : ||K||_F \leq 1, \text{rank}(K) \leq r \right\}.$$

The strategy is to find an accurate covering number for $\mathcal{K}(r)$ such that we can apply an ϵ-net argument. Suppose we can find a subset $\mathcal{K}_\epsilon(r) = \{K_1, K_2, ..., K_N\} \subset \mathcal{K}(r)$ with finite cardinality $N = N(\epsilon)$ such that for any $K \in \mathcal{K}(r)$, there exists $K_j \in \mathcal{K}_\epsilon(r)$ such that $||K_j - K||_F \leq \epsilon$. Define $S = \frac{1}{n}\sum_{i=1}^n Z_i Z_i^T - I$, and then for any fixed matrix $K \subset \mathcal{K}(r)$, we have that

$$\begin{aligned} &|\langle S, K\rangle| \leq |\langle S, K_j\rangle| + \|K - K_j\|_{\mathrm{F}} \left| \left\langle S, \frac{K - K_j}{\|K - K_j\|_{\mathrm{F}}} \right\rangle \right| \\ \leq\ & |\langle S, K_j\rangle| + \epsilon \sup_{H\in\mathcal{K}(2r)} |\langle S, H\rangle| \\ \leq\ & \max_j |\langle S, K_j\rangle| + 2\epsilon \sup_{H\in\mathcal{K}(r)} |\langle S, H\rangle|, \end{aligned}$$

where we have used the fact that the rank of $\frac{K-K_j}{\|K-K_j\|_{\mathrm{F}}}$ is not more than $2r$ and for any such $H \in \mathcal{K}(2r)$, it can be written as the sum of two matrices with rank not more than r. Therefore, taking sup on both sides, we have that $\sup_{K\in\mathcal{K}(r)} |\langle S, K\rangle| \leq (1-2\epsilon)^{-1} \max_j |\langle S, K_j\rangle|$. Picking $\epsilon = 1/4$, by union bound and Lemma 17, we have

$$\begin{aligned} &\mathbb{P}\left\{ \sup_{K\in\mathcal{K}(r)} \left| \left\langle \frac{1}{n}\sum_{i=1}^n Z_i Z_i^T - I, K \right\rangle \right| > t \right\} \\ \leq\ & \sum_{j=1}^{N(1/4)} \mathbb{P}\left\{ \left| \left\langle \frac{1}{n}\sum_{i=1}^n Z_i Z_i^T - I, K_j \right\rangle \right| > \frac{t}{2} \right\} \\ \leq\ & \exp(\log N(1/4) - Cn(t \wedge t^2)). \end{aligned}$$

Now it is sufficient for us to find the covering number, i.e. to show that $\log N(1/4)$ is bounded by $C'rd$ to complete our proof. We write the SVD of any $K \subset \mathcal{K}(r)$ as $K = P\Lambda Q'$. Note that both $P\Lambda$ and $Q\Lambda$ belong to the follwoing class

$$\mathcal{B} = \left\{ B \in \mathbb{R}^{d\times r} : \exists U \in \mathbb{R}^{d\times r}\text{and } D \text{ diagonal s.t. } U'U = I, B = UD, ||D||_F \leq 1 \right\}.$$

It is obvious that $\mathcal{B} \subset \left\{ B \in \mathbb{R}^{d\times r} : ||B||_F \leq 1 \right\}$, the $d \times r$ dimensional unit ball. Hence the well-known covering number of unit ball implies that for small $\epsilon/2 > 0$, we can find a subset $\mathcal{B}_{\epsilon/2} = \{B_1, B_2, ..., B_L\} \subset \mathcal{B}$ with cardinality $L(\epsilon/2) \leq (C\epsilon)^{-C_0 rd}$ such that $\inf_j ||B - B_j||_F \leq \epsilon/2$. We denote each $B_j = U_j D_j$, then we claim the subset $\mathcal{K}_\epsilon(r)$ can be defined as follows

$$\mathcal{K}_\epsilon(r) = \left\{ K_{ij} = U_i D_i U_j' : i, j \leq L(\epsilon/2) \right\}.$$

As a consequence, we obtain that $N(\epsilon) \leq L^2(\epsilon/2) \leq (C\epsilon)^{-2C_0 rd}$ and hence $\log N(1/4) \leq C' rd$. We prove our claim now. First, it is clear that any $K_{ij} \in \mathcal{K}_\epsilon(r)$ and we have $||K_{ij}||_F \leq 1$ and $\text{rank}(K_{ij}) \leq r$. Second, for any $K = P\Lambda Q' \subset \mathcal{K}(r)$, we can find $B_j = U_j D_j$ such that $\|Q\Lambda - B_j\|_{\mathrm{F}} \leq \epsilon/2$ and further can find $B_i = U_i D_i$ such that $\|PD_j - B_i\|_{\mathrm{F}} \leq \epsilon/2$. Hence we have

$$\begin{aligned} \|K - K_{ij}\|_{\mathrm{F}} &= \|P\Lambda Q' - U_i D_i U_j'\|_{\mathrm{F}} \\ &\leq \|P\Lambda Q' - PB_j'\|_{\mathrm{F}} + \|PD_j U_j' - U_i D_i U_j'\|_{\mathrm{F}} \\ &= \|Q\Lambda - B_j\|_{\mathrm{F}} + \|PD_j - B_i\|_{\mathrm{F}} \leq \epsilon. \end{aligned}$$

Therefore the proof is complete. We remark that a similar covering argument is also obtained by Candes and Plan [11, Lemma 3.1]. The proof we provide above is different from theirs, because we avoid the concepts of Grassmann manifold through very elementary calculation. □

A.4. Proof of Lemma 9. Expanding the Frobenius norm, we have

$$\left\| AB' - EF' \right\|_{\mathrm{F}}^2 = 2\,\mathsf{Tr}(I - A'EF'B).$$

On the other hand, we have

$$\left\langle ADB', AB' - EF' \right\rangle = \mathsf{Tr}(D - DA'EF'B) = \sum_{l=1}^{r} d_l (I - A'EF'B)_{ll}.$$

It is clear that $(I - A'EF'B)_{ll} \geq 0$, and thus the result follows.

A.5. Proof of Lemma 10. The proof depends on two facts. The first one is deterministic

$$\|\widehat{\Sigma}_x^{1/2} A \widehat{\Sigma}_y^{1/2}\|_{\mathrm{F}}^2 = \|\widehat{\Sigma}_{xT_uT_u}^{1/2} A_{T_uT_v} \widehat{\Sigma}_{yT_vT_v}^{1/2}\|_{\mathrm{F}}^2. \tag{95}$$

The second one is that with probability at least $1-\exp(-C' k_q^u \log(ep/k_q^u)) - \exp(-C' k_q^v \log(em/k_q^v))$, we have

$$\|\widehat{\Sigma}_{xT_uT_u}^{1/2} - \Sigma_{xT_uT_u}^{1/2}\|_{\mathrm{op}} \vee \|\widehat{\Sigma}_{yT_vT_v}^{1/2} - \Sigma_{yT_vT_v}^{1/2}\|_{\mathrm{op}} \leq \frac{C}{n}(k_q^u \log(ep/k_q^u) + k_q^v \log(em/k_q^v)). \tag{96}$$

The two facts will be derived at the end of the proof.

The assumption that $\frac{1}{n}(k_q^u \log(ep/k_q^u) + k_q^v \log(em/k_q^v))$ is sufficiently small and the fact (96) immediately imply that there exists some constant $C > 0$ such that

$$\|\widehat{\Sigma}_{xT_uT_u}^{1/2}\|_{\mathrm{op}}, \|\widehat{\Sigma}_{xT_uT_u}^{-1/2}\|_{\mathrm{op}}, \|\widehat{\Sigma}_{yT_vT_v}^{1/2}\|_{\mathrm{op}}, \|\widehat{\Sigma}_{yT_vT_v}^{-1/2}\|_{\mathrm{op}} \in [1/C, C],$$

since the spectra of $\Sigma_{yT_vT_v}^{1/2}$ and $\Sigma_{xT_uT_u}^{1/2}$ are bounded below and above by universal constants. This consequence together with the fact (95) further shows the desired result. Namely,

$$\|A\|_{\mathrm{F}}^2 \leq \|\widehat{\Sigma}_{xT_uT_u}^{-1/2}\|_{\mathrm{op}}^2 \|\widehat{\Sigma}_{yT_vT_v}^{-1/2}\|_{\mathrm{op}}^2 \|\widehat{\Sigma}_{xT_uT_u}^{1/2} A_{T_uT_v} \widehat{\Sigma}_{yT_vT_v}^{1/2}\|_{\mathrm{F}}^2 \leq C^4 \|\widehat{\Sigma}_x^{1/2} A \widehat{\Sigma}_y^{1/2}\|_{\mathrm{F}}^2,$$

$$\|\widehat{\Sigma}_x^{1/2} A \widehat{\Sigma}_y^{1/2}\|_{\mathrm{F}}^2 \leq \|\widehat{\Sigma}_{xT_uT_u}^{1/2}\|_{\mathrm{op}}^2 \|\widehat{\Sigma}_{yT_vT_v}^{1/2}\|_{\mathrm{op}}^2 \|A_{T_uT_v}\|_{\mathrm{F}}^2 \leq C^4 \|A\|_{\mathrm{F}}^2.$$

Now we only need to prove the two facts (95) and (96). The fact (96) is a simple consequence of Lemma 13 and Lemma 16. To see (95), we expand the Frobenius norm by trace product,

$$\begin{aligned} \|\widehat{\Sigma}_x^{1/2} A \widehat{\Sigma}_y^{1/2}\|_{\mathrm{F}}^2 &= \mathsf{Tr}\left(\widehat{\Sigma}_y^{1/2} A' \widehat{\Sigma}_x A \widehat{\Sigma}_y^{1/2}\right) \\ &= \mathsf{Tr}\left((A_{T_uT_v})' \widehat{\Sigma}_{xT_uT_u} A_{T_uT_v} \widehat{\Sigma}_{yT_vT_v}\right) = \|\widehat{\Sigma}_{xT_uT_u}^{1/2} A_{T_uT_v} \widehat{\Sigma}_{yT_vT_v}^{1/2}\|_{\mathrm{F}}^2. \end{aligned}$$

A.6. Proofs of Lemma 11 and Lemma 12.

PROOF OF LEMMA 11. The last claim is proved by the following observation.

$$\widetilde{U}_1' \Sigma_x \widetilde{U}_1 = (\widetilde{U}_{1S_u*})' \Sigma_{xS_uS_u} \widetilde{U}_{1S_u*} = I_r, \;\; \widetilde{V}_1' \Sigma_y \widetilde{V}_1 = (\widetilde{V}_{1S_v*})' \Sigma_{yS_vS_v} \widetilde{V}_{1S_v*} = I_r.$$

To show the first two claims, we need to prove that all the singular values of $(\Sigma_{xS_uS_u})^{1/2} U_{1S_u*}$ and $(\Sigma_{yS_vS_v})^{1/2} V_{1S_u*}$ are close to 1. Indeed, if all singular values are between 0.9 and 1.1, then the range of spectrum of $P\widetilde{\Lambda}_1 Q'$ will be in the interval $[0.9\lambda_r, 1.1\lambda_1]$ according to (51). Therefore our assumptions on λ_1 and λ_r imply $1.1\kappa\lambda \geq \widetilde{\lambda}_1 \geq \widetilde{\lambda}_r \geq 0.9\lambda$. The second term of Ξ in (53) is orthogonal to the first term $P\widetilde{\Lambda}_1 Q'$ and clearly its largest singular value can be bounded by $C\lambda_{r+1}$, which is less than $c\lambda$ by our assumption on λ_{r+1}. Therefore we finish the proof of the first two claims.

Now we bound the singular values of $(\Sigma_{xS_uS_u})^{1/2}U_{1S_u*}$ and $(\Sigma_{yS_vS_v})^{1/2}V_{1S_u*}$. Note

$$
\begin{aligned}
I_r &= U_1'\Sigma_x U_1 = (U_{1S_u*})'\Sigma_{xS_uS_u}U_{1S_u*} + (U_{1S_u*})'\Sigma_{xS_uS_u^c}U_{1S_u^c*} \\
&\quad + (U_{1S_u^c*})'\Sigma_{xS_u^cS_u}U_{1S_u*} + (U_{1S_u^c*})'\Sigma_{xS_u^cS_u^c}U_{1S_u^c*}.
\end{aligned}
$$

Therefore we have

$$
\|(U_{1S_u*})'\Sigma_{xS_uS_u}U_{1S_u*} - I_r\|_{\mathrm{F}}^2 \leq C\|U_{1S_u^c*}\|_{\mathrm{F}}^2 \leq \frac{Cq}{2-q}k_q^u(s_u/k_q^u)^{2/q} \leq 0.01,
$$

where the last two inequalities follow from (56) and (16). Hence we have shown that all singular values of $(\Sigma_{xS_uS_u})^{1/2}U_{1S_u*}$ are bewteen 0.9 and 1.1. Similar analysis implies that the same result holds for $(\Sigma_{yS_vS_v})^{1/2}V_{1S_u*}$. □

PROOF OF LEMMA 12. First of all, note that $\|(U_{1S_u*})'\Sigma_{xS_u*}U_2\|_{\mathrm{F}}^2 \leq C\|U_{1S_u^c*}\|_{\mathrm{F}}^2$ by the following equality,

$$
0 = U_1'\Sigma_x U_2 = (U_{1S_u*})'\Sigma_{xS_u*}U_2 + (U_{1S_u^c*})'\Sigma_{xS_u^c*}U_2.
$$

Moreover, the fact that all singular values of $(\Sigma_{xS_uS_u})^{1/2}U_{1S_u*}$ are between 0.9 and 1.1, which is shown in Lemma 11, implies that there exists $W \in \mathbb{R}^{r\times r}$ with $\|W\|_{\mathrm{op}} \leq 1.2$, such that $P = (\Sigma_{xS_uS_u})^{1/2}U_{1S_u*}W$. Therefore,

$$
\begin{aligned}
\|P'(\Sigma_{xS_uS_u})^{-1/2}\Sigma_{xS_u*}U_2\|_{\mathrm{F}}^2 &= \|W'(U_{1S_u*})'\Sigma_{xS_u*}U_2\|_{\mathrm{F}}^2 \\
&\leq \|(U_{1S_u*})'\Sigma_{xS_u*}U_2\|_{\mathrm{F}}^2\|W\|_{\mathrm{op}}^2 \leq C\|U_{1S_u^c*}\|_{\mathrm{F}}^2.
\end{aligned}
$$

Similar analysis shows the result for $Q'(\Sigma_{yS_vS_v})^{-1/2}\Sigma_{yS_v*}V_2$. Hence the proof is complete. □

DEPARTMENT OF STATISTICS
YALE UNIVERSITY
NEW HAVEN, CT 06511.
E-MAIL: chao.gao@yale.edu
E-MAIL: zhao.ren@yale.edu
E-MAIL: huibin.zhou@yale.edu
URL: http://www.stat.yale.edu/~hz68

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PA 19104.
E-MAIL: zongming@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/~zongming