Supplementary Material to "D-CCA: A Decomposition-based Canonical Correlation Analysis for High-Dimensional Datasets"

Hai Shu, Xiao Wang, and Hongtu Zhu

Abstract

Two important propositions and all technical proofs are given in Section S.1. Additional simulations are included in Section S.2.

S.1 Propositions and Technical Proofs

Proposition S.1. *Let* $\boldsymbol{x} = (\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top)^\top$*,* $\boldsymbol{e} = (\boldsymbol{e}_1^\top, \boldsymbol{e}_2^\top)^\top$*, and* $r_x = \operatorname{rank}(\operatorname{cov}(\boldsymbol{x}))$*. Assume that ranks* r_1 *and* r_2 *are constants. When* $p := \min(p_1, p_2) \to \infty$*, if* $\min_{k\in\{1,2\}} \lambda_{r_k}(\operatorname{cov}(\boldsymbol{x}_k))$ *diverges and* $\max_{k\in\{1,2\}} \lambda_1(\operatorname{cov}(\boldsymbol{e}_k))$ *is bounded, then* $\lambda_{r_x}(\operatorname{cov}(\boldsymbol{x}))$ *diverges and* $\lambda_1(\operatorname{cov}(\boldsymbol{e}))$ *is bounded.*

Remark S.1. *Hallin and Liška (2011) proposed a decomposition method under a general dynamic factor model that includes our approximate factor model given in* (6) *and* (7) *as a special case. Their decomposition method divides each of two observed vector processes into four components that are called strongly common, weakly common, weakly idiosyncratic, and strongly idiosyncratic, respectively. Consider applying their method to our approximate factor model that has i.i.d. samples. By their Assumption A3 and Proposition 1(a) and (b) as well as Weyl's inequality, ranks* $\{r_k\}_{k=1}^2$ *are constant, and* $\min_{k\in\{1,2\}} \lambda_{r_k}(\operatorname{cov}(\boldsymbol{x}_k))$ *diverges but* $\max_{k\in\{1,2\}} \lambda_1(\operatorname{cov}(\boldsymbol{e}_k))$ *is bounded when* $p \to \infty$*. Then with the additional condition* $\operatorname{cov}(\boldsymbol{x}, \boldsymbol{e}) = \boldsymbol{0}$*, it follows from our Proposition S.1 and their Proposition 1(c) that for each* $\boldsymbol{y}_k$*,* $\boldsymbol{x}_k$ *is the sum of strongly common and weakly common components,* $\boldsymbol{e}_k$ *is the strongly idiosyncratic component, and no weakly idiosyncratic component exists. Furthermore, if* $\operatorname{span}(\boldsymbol{x}_1^\top) \cap \operatorname{span}(\boldsymbol{x}_2^\top) = \{0\}$*, i.e., the first signal canonical correlation* $\rho_1 < 1$*, then there is no strongly common component, and* $\boldsymbol{x}_k$ *is entirely the weakly common component of* $\boldsymbol{y}_k$*.*

Proof of Proposition S.1. Recall that $\mathbf{\Sigma}_k = \operatorname{cov}(\boldsymbol{x}_k) = \mathbf{V}_k\mathbf{\Lambda}_k\mathbf{V}_k^\top$ and $\mathbf{\Sigma}_{12} = \operatorname{cov}(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Using (S.4) that will be shown later, we have

$$\begin{aligned}\operatorname{cov}(\boldsymbol{x}) = \begin{bmatrix}\mathbf{\Sigma}_1 & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{12}^\top & \mathbf{\Sigma}_2\end{bmatrix} &= \begin{bmatrix}\mathbf{V}_1 & \\ & \mathbf{V}_2\end{bmatrix}\begin{bmatrix}\mathbf{\Lambda}_1^{1/2} & \\ & \mathbf{\Lambda}_2^{1/2}\end{bmatrix}\begin{bmatrix}\mathbf{U}_{\theta 1} & \\ & \mathbf{U}_{\theta 2}\end{bmatrix}\begin{bmatrix}\mathbf{I}_{r_1\times r_1} & \mathbf{\Lambda}_\theta\\ \mathbf{\Lambda}_\theta^\top & \mathbf{I}_{r_2\times r_2}\end{bmatrix}\\ &\quad\cdot\begin{bmatrix}\mathbf{U}_{\theta 1}^\top & \\ & \mathbf{U}_{\theta 2}^\top\end{bmatrix}\begin{bmatrix}\mathbf{\Lambda}_1^{1/2} & \\ & \mathbf{\Lambda}_2^{1/2}\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^\top & \\ & \mathbf{V}_2^\top\end{bmatrix}\\ &=: \mathbf{V}_0\mathbf{\Lambda}_0^{1/2}\mathbf{U}_0\mathbf{\Phi}_0\mathbf{U}_0^\top\mathbf{\Lambda}_0^{1/2}\mathbf{V}_0^\top. \end{aligned} \tag{S.1}$$

According to Theorem 3.3.16(d) in Horn and Johnson (1994), we have

$$\begin{aligned}\sigma_{r_x}(\mathbf{U}_0\mathbf{\Phi}_0\mathbf{U}_0^\top) &= \sigma_{r_x}(\mathbf{\Lambda}_0^{-1/2}\mathbf{\Lambda}_0^{1/2}\mathbf{U}_0\mathbf{\Phi}_0\mathbf{U}_0^\top\mathbf{\Lambda}_0^{1/2}\mathbf{\Lambda}_0^{-1/2})\\ &\le \sigma_1(\mathbf{\Lambda}_0^{-1/2})\sigma_{r_x}(\mathbf{\Lambda}_0^{1/2}\mathbf{U}_0\mathbf{\Phi}_0\mathbf{U}_0^\top\mathbf{\Lambda}_0^{1/2})\sigma_1(\mathbf{\Lambda}_0^{-1/2}).\end{aligned}$$

Since r_k and $\sigma_\ell(\mathbf{\Theta}) = \rho_\ell$ are constant for $k \le 2$ and $\ell \le r_{12}$, $\sigma_{r_x}(\mathbf{U}_0\mathbf{\Phi}_0\mathbf{U}_0^\top) = \sigma_{r_x}(\mathbf{\Phi}_0)$ is a positive constant. Thus, when $p \to \infty$, we have

$$\begin{aligned}\lambda_{r_x}(\operatorname{cov}(\boldsymbol{x})) &= \sigma_{r_x}(\operatorname{cov}(\boldsymbol{x})) = \sigma_{r_x}(\mathbf{\Lambda}_0^{1/2}\mathbf{U}_0\mathbf{\Phi}_0\mathbf{U}_0^\top\mathbf{\Lambda}_0^{1/2})\\ &\ge \sigma_{r_x}(\mathbf{U}_0\mathbf{\Phi}_0\mathbf{U}_0^\top)/\sigma_1^2(\mathbf{\Lambda}_0^{-1/2}) = \sigma_{r_x}(\mathbf{U}_0\mathbf{\Phi}_0\mathbf{U}_0^\top)\min_{k\in\{1,2\}}\lambda_{r_k}(\operatorname{cov}(\boldsymbol{x}_k)) \to \infty.\end{aligned}$$

For the noise covariance matrices, we denote their compact SVDs by $\operatorname{cov}(\boldsymbol{e}_k) = \mathbf{V}_{e,k}\mathbf{\Lambda}_{e,k}\mathbf{V}_{e,k}^\top$ for $k = 1, 2$. Similar to (S.1), we have

$$\begin{aligned}\operatorname{cov}(\boldsymbol{e}) &= \begin{bmatrix}\mathbf{V}_{e,1} & \\ & \mathbf{V}_{e,2}\end{bmatrix}\begin{bmatrix}\mathbf{\Lambda}_{e,1}^{1/2} & \\ & \mathbf{\Lambda}_{e,2}^{1/2}\end{bmatrix}\mathbf{U}_e\begin{bmatrix}\mathbf{I}_{r_{e,1}\times r_{e,1}} & \mathbf{\Lambda}_{e12}\\ \mathbf{\Lambda}_{e12}^\top & \mathbf{I}_{r_{e,2}\times r_{e,2}}\end{bmatrix}\mathbf{U}_e^\top\begin{bmatrix}\mathbf{\Lambda}_{e,1}^{1/2} & \\ & \mathbf{\Lambda}_{e,2}^{1/2}\end{bmatrix}\begin{bmatrix}\mathbf{V}_{e,1}^\top & \\ & \mathbf{V}_{e,2}^\top\end{bmatrix}\\ &=: \mathbf{V}_e\mathbf{\Lambda}_e^{1/2}\mathbf{U}_e\mathbf{\Phi}_e\mathbf{U}_e^\top\mathbf{\Lambda}_e^{1/2}\mathbf{V}_e^\top,\end{aligned}$$

where $\mathbf{U}_e$ is an orthogonal matrix, $r_{e,k} = \operatorname{rank}(\operatorname{cov}(\boldsymbol{e}_k))$ for $k = 1, 2$, and $\mathbf{\Lambda}_{e12}$ is a $r_{e,1} \times r_{e,2}$ rectangular diagonal matrix with the canonical correlations between $\boldsymbol{e}_k$'s on its main diagonal. Finally, as $p \to \infty$, we have

$$\lambda_1(\operatorname{cov}(\boldsymbol{e})) = \|\operatorname{cov}(\boldsymbol{e})\|_2 \le \|\mathbf{\Phi}_e\|_1 \cdot \max_{k\in\{1,2\}}\lambda_1(\operatorname{cov}(\boldsymbol{e}_k)) \le 2 \cdot \max_{k\in\{1,2\}}\lambda_1(\operatorname{cov}(\boldsymbol{e}_k)) < \infty.$$

We finish the proof of Proposition S.1. $\square$

Proposition S.2. *Equation* (15) *is the unique solution to the problem in* (11) *subject to* (12)-(14).

Proof. Let $\theta(\cdot,\cdot)$ denote the angle between two elements in space $(\mathcal{L}_0^2, \mathrm{cov})$. Then, $\cos\theta(\cdot,\cdot) = \mathrm{corr}(\cdot,\cdot)$. Hereafter, we use these two operators exchangeably. Note that $\sum_{k=1}^2 \cos^2\theta(z_{k\ell}, z_{1\ell}) \geq 1$. If $w \perp \mathrm{span}(\{z_{1\ell}, z_{2\ell}\})$, then $\sum_{k=1}^2 \cos^2\theta(z_{k\ell}, w) = 0$, and thus such a w is not an optimal solution to the right-hand side of (11). When $w \not\perp \mathrm{span}(\{z_{1\ell}, z_{2\ell}\})$, since $\cos\theta(z_{k\ell}, w) = \cos\theta(z_{k\ell}, w_0)\cos\theta(w_0, w)$, where w_0 denotes the projection of w onto $\mathrm{span}(\{z_{1\ell}, z_{2\ell}\})$, we only need to consider $w \in \mathrm{span}(\{z_{1\ell}, z_{2\ell}\})$. Let $w = az_{1\ell} + bz_{2\ell}$ with $\mathrm{var}(w) = a^2 + b^2 + 2ab\rho_\ell = 1$. Then, we have

$$\begin{aligned}\sum_{k=1}^2 \mathrm{corr}^2(z_{k\ell}, w) &= (a + b\rho_\ell)^2 + (a\rho_\ell + b)^2 \\ &= a^2 + b^2 + 4ab\rho_\ell + \rho_\ell^2(a^2+b^2) = 1 + 2ab\rho_\ell + \rho_\ell^2(1 - 2ab\rho_\ell) \\ &= 1 + \rho_\ell^2 + 2ab\rho_\ell(1-\rho_\ell^2). \end{aligned} \tag{S.2}$$

We first consider $\rho_\ell \in (0,1)$. Equation (S.2) is maximized only when $ab \geq 0$. Without loss of generality, we assume a and b are nonnegative. Since $2ab \leq a^2 + b^2 = 1 - 2ab\rho_\ell$, the maximizer of ab satisfies $a = b = (2 + 2\rho_\ell)^{-1/2}$. Thus, $c_\ell \propto z_{1\ell} + z_{2\ell}$. Let $c_\ell = \alpha(z_{1\ell} + z_{2\ell})$. From (13), we have

$$\begin{aligned}0 = \mathrm{corr}(d_{1\ell}, d_{2\ell}) &= \mathrm{corr}\big((1-\alpha)z_{1\ell} - \alpha z_{2\ell}, (1-\alpha)z_{2\ell} - \alpha z_{1\ell}\big) \\ &= 2(\rho_\ell + 1)\alpha^2 - 2(\rho_\ell+1)\alpha + \rho_\ell.\end{aligned}$$

Hence, we obtain

$$\alpha = \frac{1}{2}\left(1 \pm \sqrt{\frac{1-\rho_\ell}{1+\rho_\ell}}\right).$$

It follows that

$$\mathrm{var}(c_\ell) = \frac{1}{2}\left(1 \pm \sqrt{1 - \frac{2}{1 + 1/\rho_\ell}}\right)^2 (1 + \rho_\ell). \tag{S.3}$$

To satisfy (14), c_ℓ must be the one in (15) when $\rho_\ell \in (0,1)$.

Now we consider the solution of c_ℓ when $\rho_\ell \in \{0,1\}$. By (S.3), for c_ℓ that is defined in (15) when $\rho_\ell \in (0,1)$, we have $\lim_{\rho_\ell \to 0^+} \mathrm{var}(c_\ell) = 0$ and $\lim_{\rho_\ell \to 1^-} \mathrm{var}(c_\ell) = 1$. Then by (14), when $\rho_\ell = 0$, then $\mathrm{var}(c_\ell) = 0$, and thus $c_\ell = 0$ which satisfies (15) as well as (11) and (13). We also obtain $\mathrm{var}(c_\ell) \geq 1$ when $\rho_\ell = 1$. Now consider $\rho_\ell = 1$. By (S.2), we have $\max_{w \in \mathcal{L}_0^2} \sum_{k=1}^2 \mathrm{corr}^2(z_{k\ell}, w) = 2$, and thus $c_\ell \propto z_{1\ell} = z_{2\ell}$. If $\mathrm{var}(c_\ell) > 1$ or $c_\ell = -z_{1\ell} = -z_{2\ell}$, then $d_{1\ell} = d_{2\ell} \neq 0$ and $\mathrm{corr}(d_{1\ell}, d_{2\ell}) = 1$. Thus, to satisfy (13), $c_\ell = z_{1\ell} = z_{2\ell}$, which is equivalent to (15).

From the above, we have that c_ℓ must be the one in (15) when $\rho_\ell \in [0, 1]$. □

Lemma S.1. *When $n \to \infty$, if $a_n = O_P(b_n)$ holds on a given event $\mathcal{A}_n$ that has $\mathbb{P}(\mathcal{A}_n) \to 1$, then we have $a_n = O_P(b_n)$.*

Proof. By the given assumptions, for any $\varepsilon > 0$, there exist constants M_ε and N_ε such that $\mathbb{P}\left(|a_n| \le M_\varepsilon b_n \middle| \mathcal{A}_n\right) > 1 - \varepsilon$ and $\mathbb{P}(\mathcal{A}_n) > 1 - \varepsilon$ for all $n \ge N_\varepsilon$. Then, $\mathbb{P}(|a_n| \le M_\varepsilon b_n) \ge \mathbb{P}\left(|a_n| \le M_\varepsilon b_n \middle| \mathcal{A}_n\right) \mathbb{P}(\mathcal{A}_n) > (1-\varepsilon)^2 > 1 - 2\varepsilon$. Hence, we obtain $a_n = O_P(b_n)$. □

Proof of Theorem 1. First, we show $\mathrm{rank}(\mathbf{\Theta}) = r_{12}$. Since

$$\boldsymbol{x}_k = \mathrm{cov}(\boldsymbol{x}_k, \boldsymbol{z}_k^*)\boldsymbol{z}_k^* = \mathrm{cov}(\boldsymbol{x}_k, \mathbf{\Lambda}_k^{-1/2}\mathbf{V}_k^\top \boldsymbol{x}_k)\boldsymbol{z}_k^* = \mathbf{V}_k \mathbf{\Lambda}_k^{1/2} \boldsymbol{z}_k^*,$$

we obtain

$$\mathbf{\Sigma}_{12} = \mathrm{cov}(\mathbf{V}_1 \mathbf{\Lambda}_1^{1/2} \boldsymbol{z}_1^*, \mathbf{V}_2 \mathbf{\Lambda}_2^{1/2} \boldsymbol{z}_2^*) = \mathbf{V}_1 \mathbf{\Lambda}_1^{1/2} \mathbf{\Theta} \mathbf{\Lambda}_2^{1/2} \mathbf{V}_2^\top. \tag{S.4}$$

Hence by $\mathrm{rank}(\mathbf{M}_1\mathbf{M}_2) \le \min(\mathrm{rank}(\mathbf{M}_1), \mathrm{rank}(\mathbf{M}_2))$ for real matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, we have $\mathrm{rank}(\mathbf{\Theta}) \ge r_{12}$. Again using the above inequality of the rank of matrix product, by $\mathbf{\Theta} = \mathbf{\Lambda}_1^{-1/2}\mathbf{V}_1^\top \mathbf{\Sigma}_{12} \mathbf{V}_2 \mathbf{\Lambda}_2^{-1/2}$, we have $\mathrm{rank}(\mathbf{\Theta}) \le r_{12}$. Thus, $\mathrm{rank}(\mathbf{\Theta}) = r_{12}$.

Now temporarily replace the constraint $\ell \le r_{12}$ by $\ell \le r_{\min}$ for (10) . Let $\{\widetilde{z}_{1\ell}, \widetilde{z}_{2\ell}\}_{\ell=1}^{r_{\min}}$ be an arbitrary solution of (10). We will later see that $\mathrm{corr}(\widetilde{z}_{1\ell}, \widetilde{z}_{2\ell}) = 0$ for all $\ell > r_{12}$. Augment $(\widetilde{z}_{k1}, \ldots, \widetilde{z}_{k,r_{\min}})^\top$ with any $(r_k - r_{\min})$ standardized variables to be $\widetilde{\boldsymbol{z}}_k = (\widetilde{z}_{k1}, \ldots, \widetilde{z}_{k,r_k})^\top$ such that $\widetilde{\boldsymbol{z}}_k^\top$ is an orthonormal basis of $\mathrm{span}(\boldsymbol{x}_k^\top)$. Denote $\widetilde{\mathbf{\Theta}} = \mathrm{cov}(\widetilde{\boldsymbol{z}}_1, \widetilde{\boldsymbol{z}}_2)$. When $\ell = 1$ in (10), z_{11} must be proportional to the projection of z_{21} onto $\mathrm{span}(\boldsymbol{x}_1^\top)$. Hence, $z_{21} \perp \mathrm{span}(\boldsymbol{x}_1^\top) \backslash \mathrm{span}(z_{11})$, and $\widetilde{\mathbf{\Theta}}^{[2:r_1,1]}$ is a zero vector. Similarly, $\widetilde{\mathbf{\Theta}}^{[1,2:r_2]}$ is a zero vector. Using the same argument for $\ell = 2, \ldots, r_{\min}$ yields that the only nonzero entries of $\widetilde{\mathbf{\Theta}}$ are located on the diagonal of $\widetilde{\mathbf{\Theta}}^{[1:r_{\min},1:r_{\min}]}$. Note that there exists an orthogonal matrix $\mathbf{Q}_k$ such that $\widetilde{\boldsymbol{z}}_k = \mathbf{Q}_k \boldsymbol{z}_k^*$. Then, $\widetilde{\mathbf{\Theta}} = \mathbf{Q}_1 \mathbf{\Theta} \mathbf{Q}_2^\top$ has rank r_{12}. Hence, the only nonzero entries of $\widetilde{\mathbf{\Theta}}$ are the first r_{12} elements of its main diagonal. We thus only need $\ell \le r_{12}$ in (10).

The proof is complete. □

Proof of Theorem 2. We only need to show the uniqueness of $\boldsymbol{c}_1$.

Let $\{\widetilde{\boldsymbol{z}}_k\}_{k=1,2}$ be another set of augmented standardized canonical variables. Then, there exists an orthogonal matrix $\mathbf{Q}_k$ such that $\widetilde{\boldsymbol{z}}_k = \mathbf{Q}_k \boldsymbol{z}_k$ with $\boldsymbol{z}_k = \boldsymbol{\Gamma}_k^\top \boldsymbol{x}_k$ defined in (18). By Theorem 1 and the fact that $\mathrm{cov}(\widetilde{\boldsymbol{z}}_1, \widetilde{\boldsymbol{z}}_2) = \mathbf{Q}_1 \boldsymbol{\Lambda}_\theta \mathbf{Q}_2^\top$ has the same singular values of $\boldsymbol{\Lambda}_\theta$, we have $\mathrm{cov}(\widetilde{\boldsymbol{z}}_1, \widetilde{\boldsymbol{z}}_2) = \mathbf{Q}_1 \boldsymbol{\Lambda}_\theta \mathbf{Q}_2^\top = \boldsymbol{\Lambda}_\theta$. Let m be the number of distinct nonzero singular values of $\boldsymbol{\Lambda}_\theta$. Then for $k = 1, 2$, we have $\mathbf{Q}_k = \mathrm{diag}(\mathbf{M}_{k1}, \ldots, \mathbf{M}_{km}, \mathbf{M}_{k,m+1})$, where $\mathbf{M}_{k\ell}, \ell \le m$ is an orthogonal matrix with column dimension equal to the repetition number of the ℓ-th largest distinct nonzero singular value of $\boldsymbol{\Lambda}_\theta$, and $\mathbf{M}_{k,m+1}$ might be an empty matrix. By $\boldsymbol{\Lambda}_\theta \mathbf{Q}_2^\top = \mathbf{Q}_1^\top \boldsymbol{\Lambda}_\theta$, we obtain $\mathbf{M}_{1\ell} = \mathbf{M}_{2\ell}$ for all $\ell \le m$.

By the expression of $\boldsymbol{c}_1$ in (19), we only need to show

$$\mathrm{cov}(\boldsymbol{x}_1, \widetilde{\boldsymbol{z}}_1^{[1:r_{12}]}) \mathbf{A}_C \sum_{k=1}^{2} \widetilde{\boldsymbol{z}}_k^{[1:r_{12}]} = \mathrm{cov}(\boldsymbol{x}_1, \boldsymbol{z}_1^{[1:r_{12}]}) \mathbf{A}_C \sum_{k=1}^{2} \boldsymbol{z}_k^{[1:r_{12}]}.$$

This is true because

$$\begin{aligned}
&\mathrm{cov}(\boldsymbol{x}_1, \widetilde{\boldsymbol{z}}_1^{[1:r_{12}]}) \mathbf{A}_C \sum_{k=1}^{2} \widetilde{\boldsymbol{z}}_k^{[1:r_{12}]} \\
&= \mathrm{cov}(\boldsymbol{x}_1, \boldsymbol{z}_1) (\mathbf{Q}_1^{[1:r_{12},:]})^\top \mathbf{A}_C \sum_{k=1}^{2} \mathbf{Q}_k^{[1:r_{12},:]} \boldsymbol{z}_k \\
&= \mathrm{cov}(\boldsymbol{x}_1, \boldsymbol{z}_1^{[1:r_{12}]}) \left(\mathrm{diag}(\mathbf{M}_{11}, \ldots, \mathbf{M}_{1m})\right)^\top \mathbf{A}_C \sum_{k=1}^{2} \mathrm{diag}(\mathbf{M}_{11}, \ldots, \mathbf{M}_{1m}) \boldsymbol{z}_k^{[1:r_{12}]} \\
&= \mathrm{cov}(\boldsymbol{x}_1, \boldsymbol{z}_1^{[1:r_{12}]}) \mathbf{A}_C \sum_{k=1}^{2} \boldsymbol{z}_k^{[1:r_{12}]}.
\end{aligned}$$

□

Proof of Theorem 3. Under Assumption 1, by the proof of Theorem 4.1 in Wang and Fan (2017) (see the bound for their Δ_{L1}), we have

$$\delta_{\boldsymbol{\Sigma}_k} := \|\widehat{\boldsymbol{\Sigma}}_k - \boldsymbol{\Sigma}_k\|_2 = O_P(\lambda_{k1}/\sqrt{n}). \tag{S.5}$$

From Weyl's inequality [see Theorem 3.3.16(c) in Horn and Johnson (1994)],

$$|\lambda_{k\ell} - \lambda_\ell(\boldsymbol{\Sigma}_k)| \le \|\mathrm{cov}(\boldsymbol{e}_k)\|_2 \le s_0 \quad \text{for} \quad 1 \le \ell \le r_k.$$

This implies

$$\lambda_{k\ell}/\lambda_\ell(\mathbf{\Sigma}_k) \to 1 \quad \text{for} \quad 1 \le \ell \le r_k. \tag{S.6}$$

Together with the assumption that $\lambda_{k1}/\lambda_{k,r_k}$ is bounded from above and below, we have

$$\lambda_\ell(\mathbf{\Sigma}_k) \asymp \lambda_m(\mathbf{\Sigma}_k) \quad \text{for} \quad 1 \le \ell, m \le r_k. \tag{S.7}$$

By Weyl's inequality, (S.5) and (S.6),

$$\left| \|\widetilde{\mathbf{X}}_k\|_2^2/n - \lambda_1(\mathbf{\Sigma}_k) \right| = \left| \lambda_1(\widehat{\mathbf{\Sigma}}_k) - \lambda_1(\mathbf{\Sigma}_k) \right| \le \delta_{\mathbf{\Sigma}_k} = O_P(\lambda_1(\mathbf{\Sigma}_k)/\sqrt{n}). \tag{S.8}$$

Thus,

$$\frac{\|\widetilde{\mathbf{X}}_k\|_2}{\sqrt{n\lambda_1(\mathbf{\Sigma}_k)}} = 1 + o_P(1). \tag{S.9}$$

Under Assumption 1, by the proof of Theorem C.1 in Wang and Fan (2017) (see the bound for their $\max_{i \le p} T^{-1} \sum_{t=1}^T |\widehat{u}_{it} - u_{it}|^2$), we have

$$\left\| \mathbf{X}_k - \mathbf{U}_{k1}^{[:,1:r_k]} \operatorname{diag}(\sigma_1(\mathbf{Y}_k), \ldots, \sigma_{r_k}(\mathbf{Y}_k))(\mathbf{U}_{k2}^{[:,1:r_k]})^\top \right\|_F = O_P(\sqrt{p_k \log p_k}). \tag{S.10}$$

Also under Assumption 1, by the proof of Theorem 3.1 and the argument in the third paragraph on page 1355 in Wang and Fan (2017), for $1 \le \ell \le r_k$,

$$\left| \frac{\sigma_\ell^2(\mathbf{Y}_k)}{n\lambda_{k\ell}} - \frac{[\widehat{\sigma}_\ell^S(\mathbf{Y}_k)]^2}{n\lambda_{k\ell}} \right| = O_P\left(\frac{\tau_k p_k}{n\lambda_{k\ell}} \right) = O_P\left(\frac{p_k}{n\lambda_{k\ell}} + \frac{1}{n} \right) = O_P(1)$$

and

$$\frac{[\widehat{\sigma}_\ell^S(\mathbf{Y}_k)]^2}{n\lambda_{k\ell}} - 1 = O_P\left(\frac{1}{\sqrt{n}} \right).$$

Hence,

$$\left| \frac{\sigma_\ell(\mathbf{Y}_k)}{\sqrt{n\lambda_{k\ell}}} - \frac{\widehat{\sigma}_\ell^S(\mathbf{Y}_k)}{\sqrt{n\lambda_{k\ell}}} \right| = O_P\left(\left| \frac{\sigma_\ell^2(\mathbf{Y}_k)}{n\lambda_{k\ell}} - \frac{[\widehat{\sigma}_\ell^S(\mathbf{Y}_k)]^2}{n\lambda_{k\ell}} \right| \right) = O_P\left(\frac{p_k}{n\lambda_{k\ell}} + \frac{1}{n} \right). \tag{S.11}$$

By (S.11) and (S.10), we have

$$
\begin{aligned}
\left\|\widetilde{\mathbf{X}}_k-\mathbf{X}_k\right\|_2 &\leq \left\|\widetilde{\mathbf{X}}_k-\mathbf{X}_k\right\|_F \\
&\leq \left\|\mathbf{U}_{k1}^{[:,1:r_k]}\operatorname{diag}\left(\widehat{\sigma}_1^S(\mathbf{Y}_k)-\sigma_1(\mathbf{Y}_k),\ldots,\widehat{\sigma}_{r_k}^S(\mathbf{Y}_k)-\sigma_{r_k}(\mathbf{Y}_k)\right)(\mathbf{U}_{k2}^{[:,1:r_k]})^\top\right\|_F \\
&\quad+\left\|\mathbf{X}_k-\mathbf{U}_{k1}^{[:,1:r_k]}\operatorname{diag}(\sigma_1(\mathbf{Y}_k),\ldots,\sigma_{r_k}(\mathbf{Y}_k))(\mathbf{U}_{k2}^{[:,1:r_k]})^\top\right\|_F \\
&\leq \sqrt{r_k}\max_{1\leq\ell\leq r_k}\left|\widehat{\sigma}_\ell^S(\mathbf{Y}_k)-\sigma_\ell(\mathbf{Y}_k)\right|+\left\|\mathbf{X}_k-\sum_{\ell=1}^{r_k}\sigma_\ell(\mathbf{Y}_k)\mathbf{U}_{k1}^{[:,\ell]}(\mathbf{U}_{k2}^{[:,\ell]})^\top\right\|_F \\
&= O_P\left(\frac{p_k}{\sqrt{n\lambda_{k1}}}+\sqrt{\frac{\lambda_{k1}}{n}}+\sqrt{p_k\log p_k}\right) \\
&= O_P\left(\sqrt{\frac{\lambda_1(\boldsymbol{\Sigma}_k)}{n}}+\sqrt{p_k\log p_k}\right). \qquad \text{(S.12)}
\end{aligned}
$$

It is easy to show $\mathbb{E}(f_{k\ell}^4)$ is upper bounded for all $1\leq\ell\leq r_k$. Thus, $\operatorname{var}(f_{k\ell}f_{km})$ is upper bounded for all $1\leq\ell,m\leq r_k$. Then from the central limit theorem, $\|\mathbf{F}_k\mathbf{F}_k^\top/n-\mathbf{I}_{r_k\times r_k}\|_{\max}=O_P(1/\sqrt{n})$. Hence,

$$
\frac{1}{n}\|\mathbf{F}_k\|_F^2=\text{trace}\left(\frac{1}{n}\mathbf{F}_k\mathbf{F}_k^\top\right)=r_k+O_P(1/\sqrt{n}).
$$

Then by Lemma 1 in Lam and Fan (2009), the fact $\sigma_\ell(\mathbf{B}_k)=\lambda_\ell^{1/2}(\boldsymbol{\Sigma}_k)$ for $1\leq\ell\leq r_k$, and (S.7), there exists a constant $\kappa_3\in(0,1]$ such that

$$
\begin{aligned}
\kappa_3\sqrt{r_k}+o_P(1) &\leq \frac{\sqrt{\lambda_{r_k}(\boldsymbol{\Sigma}_k)}\|\mathbf{F}_k\|_F}{\sqrt{n\lambda_1(\boldsymbol{\Sigma}_k)}} \\
&\leq \frac{\|\mathbf{X}_k\|_F}{\sqrt{n\lambda_1(\boldsymbol{\Sigma}_k)}} \\
&= \frac{\|\mathbf{B}_k\mathbf{F}_k\|_F}{\sqrt{n\lambda_1(\boldsymbol{\Sigma}_k)}} \leq \frac{\sqrt{\lambda_1(\boldsymbol{\Sigma}_k)}\|\mathbf{F}_k\|_F}{\sqrt{n\lambda_1(\boldsymbol{\Sigma}_k)}}=\sqrt{r_k}+o_P(1). \qquad \text{(S.13)}
\end{aligned}
$$

By $\|\mathbf{X}_k\|_2\leq\|\mathbf{X}_k\|_F\leq\sqrt{r_k}\|\mathbf{X}_k\|_2$, we have

$$
\kappa_3+o_P(1)\leq\frac{\|\mathbf{X}_k\|_2}{\sqrt{n\lambda_1(\boldsymbol{\Sigma}_k)}}\leq\sqrt{r_k}+o_P(1). \qquad \text{(S.14)}
$$

From (S.12), (S.13), (S.9) and $\|\widetilde{\mathbf{X}}_k\|_F \le \sqrt{r_k}\|\widetilde{\mathbf{X}}_k\|_2$, we obtain

$$\begin{aligned}\delta_{\mathbf{X}_k,2} &:= \left\|\widetilde{\mathbf{X}}_k - \mathbf{X}_k\right\|_2 \le \delta_{\mathbf{X}_k,F} := \left\|\widetilde{\mathbf{X}}_k - \mathbf{X}_k\right\|_F \\ &= O_P\left(\min\left\{\sqrt{\frac{\lambda_1(\mathbf{\Sigma}_k)}{n}} + \sqrt{p_k \log p_k}, \sqrt{n\lambda_1(\mathbf{\Sigma}_k)}\right\}\right). \end{aligned} \tag{S.15}$$

From Weyl's inequality and (S.8), for all $1 \le \ell \le p_k$,

$$|\lambda_\ell(\widehat{\mathbf{\Sigma}}_k) - \lambda_\ell(\mathbf{\Sigma}_k)| \le \|\widehat{\mathbf{\Sigma}}_k - \mathbf{\Sigma}_k\|_2 = O_P(\lambda_1(\mathbf{\Sigma}_k)/\sqrt{n}). \tag{S.16}$$

Then by (S.7),

$$\lambda_{r_k}(\widehat{\mathbf{\Sigma}}_k) \ge \lambda_{r_k}(\mathbf{\Sigma}_k) - |\lambda_{r_k}(\widehat{\mathbf{\Sigma}}_k) - \lambda_{r_k}(\mathbf{\Sigma}_k)| \ge (1 - o_P(1))\lambda_{r_k}(\mathbf{\Sigma}_k). \tag{S.17}$$

It follows that $\widetilde{r}_k = r_k$ with probability tending to 1 as $n \to \infty$. Due to Lemma S.1, we simply assume $\widetilde{r}_k = r_k$ in the rest of the proof.

By the mean value theorem and (S.16), uniformly for $\ell = 1, \ldots, r_k$, we have

$$|\lambda_\ell^{1/2}(\widehat{\mathbf{\Sigma}}_k) - \lambda_\ell^{1/2}(\mathbf{\Sigma}_k)| \le \frac{1}{2}[(1 - o_P(1))\lambda_{r_k}(\mathbf{\Sigma}_k)]^{-1/2}|\lambda_\ell(\widehat{\mathbf{\Sigma}}_k) - \lambda_\ell(\mathbf{\Sigma}_k)| = O_P(\lambda_1^{1/2}(\mathbf{\Sigma}_k)n^{-1/2}), \tag{S.18}$$

$$|\lambda_\ell^{-1/2}(\widehat{\mathbf{\Sigma}}_k) - \lambda_\ell^{-1/2}(\mathbf{\Sigma}_k)| \le \frac{1}{2}[(1 - o_P(1))\lambda_{r_k}(\mathbf{\Sigma}_k)]^{-3/2}|\lambda_\ell(\widehat{\mathbf{\Sigma}}_k) - \lambda_\ell(\mathbf{\Sigma}_k)| = O_P(\lambda_1^{-1/2}(\mathbf{\Sigma}_k)n^{-1/2}), \tag{S.19}$$

$$|\lambda_\ell^{-1}(\widehat{\mathbf{\Sigma}}_k) - \lambda_\ell^{-1}(\mathbf{\Sigma}_k)| \le [(1 - o_P(1))\lambda_{r_k}(\mathbf{\Sigma}_k)]^{-2}|\lambda_\ell(\widehat{\mathbf{\Sigma}}_k) - \lambda_\ell(\mathbf{\Sigma}_k)| = O_P(\lambda_1^{-1}(\mathbf{\Sigma}_k)n^{-1/2}).$$

By the uniqueness given in Theorem 2, we let $\mathbf{V}_k$ satisfy $(\widehat{\mathbf{V}}_k^{[:,\ell]})^\top \mathbf{V}_k^{[:,\ell]} \ge 0$ for all $k = 1, 2$ and $\ell = 1, \ldots, r_k$. By Corollary 1 in Yu et al. (2015), (S.6), (S.7) and $\min_{\ell \le r_k}(\lambda_{k\ell} - \lambda_{k,\ell+1})/\lambda_{k\ell} \ge \delta_0$, we have

$$\begin{aligned}\|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|_F &= O\left(\|\widehat{\mathbf{\Sigma}}_k - \mathbf{\Sigma}_k\|_2 \Big/ \min_{\ell \le r_k}\{\lambda_\ell(\mathbf{\Sigma}_k) - \lambda_{\ell+1}(\mathbf{\Sigma}_k)\}\right) \\ &= O_P(1/\sqrt{n}). \end{aligned} \tag{S.20}$$

Note that

$$\|\widehat{\mathbf{A}}\widehat{\mathbf{B}} - \mathbf{A}\mathbf{B}\|_2 \begin{cases} = \|\widehat{\mathbf{A}}\widehat{\mathbf{B}} - \widehat{\mathbf{A}}\mathbf{B} + \widehat{\mathbf{A}}\mathbf{B} - \mathbf{A}\mathbf{B}\|_2 \le \|\widehat{\mathbf{A}}\|_2\|\widehat{\mathbf{B}} - \mathbf{B}\|_2 + \|\mathbf{B}\|_2\|\widehat{\mathbf{A}} - \mathbf{A}\|_2, \\ = \|\widehat{\mathbf{B}}^\top\widehat{\mathbf{A}}^\top - \mathbf{B}^\top\mathbf{A}^\top\|_2 \le \|\widehat{\mathbf{B}}\|_2\|\widehat{\mathbf{A}} - \mathbf{A}\|_2 + \|\mathbf{A}\|_2\|\widehat{\mathbf{B}} - \mathbf{B}\|_2. \end{cases} \tag{S.21}$$

Now we consider the error bounds for the columns of $\widehat{\mathbf{U}}_{\theta k}$. We first consider $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_2$. By (S.21), (S.20), (S.19) and (S.7), we have

$$\begin{aligned} \delta_1^{(k)} &:= \|\widehat{\boldsymbol{\Lambda}}_k^{-1/2}\widehat{\mathbf{V}}_k^\top - \boldsymbol{\Lambda}_k^{-1/2}\mathbf{V}_k^\top\|_2 \\ &\le \lambda_{r_k}^{-1/2}(\boldsymbol{\Sigma}_k)\|\mathbf{V}_k - \widehat{\mathbf{V}}_k\|_F + \|\widehat{\mathbf{V}}_k^\top\|_2 \max_{1\le\ell\le r_k} |\lambda_\ell^{-1/2}(\widehat{\boldsymbol{\Sigma}}_k) - \lambda_\ell^{-1/2}(\boldsymbol{\Sigma}_k)| \\ &= O_P(\lambda_1^{-1/2}(\boldsymbol{\Sigma}_k)n^{-1/2}). \end{aligned} \tag{S.22}$$

By (S.21), (S.17), (S.22), (S.14), (S.15) and (S.7),

$$\begin{aligned} \delta_{Z_k} &:= \|\widehat{\boldsymbol{\Lambda}}_k^{-1/2}\widehat{\mathbf{V}}_k^\top\widetilde{\mathbf{X}}_k - \boldsymbol{\Lambda}_k^{-1/2}\mathbf{V}_k^\top\mathbf{X}_k\|_2 \\ &\le \delta_1^{(k)}\|\mathbf{X}_k\|_2 + \delta_{\mathbf{X}_k,2}\|\widehat{\boldsymbol{\Lambda}}_k^{-1/2}\|_2 \\ &= O_P(\delta_1^{(k)}\|\mathbf{X}_k\|_2 + \delta_{\mathbf{X}_k,2}\lambda_{r_k}^{-1/2}(\boldsymbol{\Sigma}_k)) \\ &= O_P\left(\min\left\{1 + \sqrt{p_k\lambda_1^{-1}(\boldsymbol{\Sigma}_k)\log p_k}, \sqrt{n}\right\}\right). \end{aligned}$$

Define $\mathbf{Z}_k^* = \boldsymbol{\Lambda}_k^{-1/2}\mathbf{V}_k^\top\mathbf{X}_k$. Then by (S.21), (S.17), (S.9), (S.14) and (S.7), we have

$$\begin{aligned} \left\|\frac{1}{n}\widehat{\mathbf{Z}}_1^*(\widehat{\mathbf{Z}}_2^*)^\top - \frac{1}{n}\mathbf{Z}_1^*(\mathbf{Z}_2^*)^\top\right\|_2 &\le \frac{1}{n}\left[\delta_{Z_1}\lambda_{r_2}^{-1/2}(\boldsymbol{\Sigma}_2)\|\mathbf{X}_2\|_2 + \delta_{Z_2}\lambda_{r_1}^{-1/2}(\widehat{\boldsymbol{\Sigma}}_1)\|\widetilde{\mathbf{X}}_1\|_2\right] \\ &= O_P\left(\min\left\{\frac{1}{\sqrt{n}} + \sum_{k=1}^2\sqrt{\frac{p_k\log p_k}{n\lambda_1(\boldsymbol{\Sigma}_k)}}, 1\right\}\right). \end{aligned}$$

Since $\boldsymbol{z}_k^*$ and $\boldsymbol{f}_k$ are both orthonormal bases of $\mathrm{span}(\boldsymbol{x}_k^\top)$, $\boldsymbol{z}_k^* = \mathbf{Q}_{zf_k}\boldsymbol{f}_k$ with a $r_k \times r_k$ orthogonal matrix $\mathbf{Q}_{zfk}$. Since $\mathbb{E}(f_{k\ell}^4)$ is upper bounded for all $\ell \le r_k$ and $k \le 2$, $\mathrm{var}(f_{1\ell}f_{2m})$ is upper bounded for all $\ell < r_1$ and $m \le r_2$. Then, we can use the central limit theorem to obtain

$$\begin{aligned} \left\|\frac{1}{n}\mathbf{Z}_1^*(\mathbf{Z}_2^*)^\top - \boldsymbol{\Theta}\right\|_2 &= \left\|\mathbf{Q}_{zf_1}\left(\frac{1}{n}\mathbf{F}_1\mathbf{F}_2^\top - \mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_2)\right)\mathbf{Q}_{zf_2}^\top\right\|_2 \\ &\le \|\mathbf{Q}_{zf_1}\|_2\left\|\frac{1}{n}\mathbf{F}_1\mathbf{F}_2^\top - \mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_2)\right\|_2\|\mathbf{Q}_{zf_2}^\top\|_2 = \left\|\frac{1}{n}\mathbf{F}_1\mathbf{F}_2^\top - \mathbb{E}(\boldsymbol{f}_1\boldsymbol{f}_2^\top)\right\|_2 = O_P\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Therefore,

$$\begin{aligned}
\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_2 &\leq \left\| \frac{1}{n}\widehat{\mathbf{Z}}_1^*(\widehat{\mathbf{Z}}_2^*)^\top - \frac{1}{n}\mathbf{Z}_1^*(\mathbf{Z}_2^*)^\top \right\|_2 + \left\| \frac{1}{n}\mathbf{Z}_1^*(\mathbf{Z}_2^*)^\top - \mathbf{\Theta} \right\|_2 \\
&\lesssim_P \min\left\{ \frac{1}{\sqrt{n}} + \sum_{k=1}^{2} \sqrt{\frac{p_k \log p_k}{n\lambda_1(\mathbf{\Sigma}_k)}}, 1 \right\} \\
&= \delta_\theta.
\end{aligned}$$

Here and also in the following text, for simplicity, we write $A \lesssim_P B$ if and only if $A = O_P(B)$. From Weyl's inequality, we have the bound for canonical correlation estimators

$$\max_{1\leq \ell \leq r_{\min}} |\sigma_\ell(\widehat{\mathbf{\Theta}}) - \sigma_\ell(\mathbf{\Theta})| \leq \|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_2 \lesssim_P \delta_\theta. \tag{S.23}$$

Using (S.21), we obtain

$$\begin{aligned}
&\max\{\|\widehat{\mathbf{\Theta}}\widehat{\mathbf{\Theta}}^\top - \mathbf{\Theta}\mathbf{\Theta}^\top\|_2, \|\widehat{\mathbf{\Theta}}^\top\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^\top\mathbf{\Theta}\|_2\} \\
&\leq (\|\widehat{\mathbf{\Theta}}\|_2 + \|\mathbf{\Theta}\|_2)\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_2 \\
&\lesssim_P (2\|\mathbf{\Theta}\|_2 + \delta_\theta)\delta_\theta \lesssim_P (\sigma_1(\mathbf{\Theta}) + \delta_\theta)\delta_\theta \\
&\lesssim_P \delta_\theta.
\end{aligned}$$

Let $\{\widetilde{\mathbf{U}}_{\theta k}\}_{k=1,2}$ be one pair of orthogonal matrices such that $\mathbf{\Theta} = \widetilde{\mathbf{U}}_{\theta 1}\mathbf{\Lambda}_\theta\widetilde{\mathbf{U}}_{\theta 2}^\top$. Define $\sigma_{\theta,1} > \cdots > \sigma_{\theta,r_\theta}$ to be the distinct nonzero singular values of $\mathbf{\Theta}$, and $\sigma_{\theta,r_\theta+1} = 0$. By Lemma 1 in Lam and Fan (2009) and Theorem 2 in Yu et al. (2015), there exists a matrix $\mathbf{Q}_k = \mathrm{diag}(\mathbf{Q}_{k1}, \ldots, \mathbf{Q}_{kr_\theta})$, where $\mathbf{Q}_{k\ell}$ is an orthogonal matrix with column dimension equal to the repetition number of $\sigma_{\theta,\ell}$, such that

$$\begin{aligned}
\|\widehat{\mathbf{U}}_{\theta k}^{[:,1:r_{12}]} - \widetilde{\mathbf{U}}_{\theta k}^{[:,1:r_{12}]}\mathbf{Q}_k\|_F &\leq \|\widehat{\mathbf{U}}_{\theta k}^{[:,1:r_{12}]}\mathbf{Q}_k^\top - \widetilde{\mathbf{U}}_{\theta k}^{[:,1:r_{12}]}\|_F\|\mathbf{Q}_k\|_2 \\
&\lesssim_P \min\left\{ \delta_\theta \Big/ \min_{1\leq\ell\leq r_\theta}\{\sigma_{\theta,\ell}^2 - \sigma_{\theta,\ell+1}^2\}, 1 \right\} \\
&\lesssim_P \delta_\theta.
\end{aligned}$$

Note that $\widetilde{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]}\mathbf{Q}_1\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}\mathbf{Q}_1^\top(\widetilde{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]})^\top = \widetilde{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]}\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}(\widetilde{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]})^\top = \mathbf{\Theta}$. By the uniqueness given in Theorem 2, we let $\mathbf{U}_{\theta k} = (\widetilde{\mathbf{U}}_{\theta k}^{[:,1:r_{12}]}\mathbf{Q}_1, \widetilde{\mathbf{U}}_{\theta k}^{[:,(r_{12}+1):r_k]})$. Define $\mathbf{U}_{\theta 2}^\star =$

$(\widetilde{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]}\mathbf{Q}_2, \widetilde{\mathbf{U}}_{\theta 2}^{[:,(r_{12}+1):r_2]})$. We have

$$\|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\|_F \lesssim_P \delta_\theta \tag{S.24}$$

and

$$\|\widehat{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]}\|_F \lesssim_P \delta_\theta. \tag{S.25}$$

Then by (S.21) and (S.23),

$$\begin{aligned}
&\left\|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]}\widehat{\mathbf{\Lambda}}_\theta^{[1:r_{12},1:r_{12}]}(\widehat{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]})^\top - \mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}(\mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\right\|_2 \\
&\le \|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]}\widehat{\mathbf{\Lambda}}_\theta^{[1:r_{12},1:r_{12}]} - \mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}\|_2\|(\widehat{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]})^\top\|_2 \\
&\qquad + \|\mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}\|_2\|(\widehat{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]})^\top - (\mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\|_2 \\
&\le \|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\|_2\|\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}\|_2 + \|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]}\|_2\|\widehat{\mathbf{\Lambda}}_\theta^{[1:r_{12},1:r_{12}]} - \mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}\|_2 \\
&\qquad + \|\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}\|_2\|(\widehat{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]})^\top - (\mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\|_2 \\
&\lesssim_P \sigma_1(\mathbf{\Theta})\delta_\theta + \delta_\theta \\
&\lesssim_P \delta_\theta.
\end{aligned}$$

By the above inequality, the inequality $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_2 \lesssim_P \delta_\theta$, and the triangular inequality of matrix norms, we have

$$\|\mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}(\mathbf{U}_{\theta 2}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\|_2 \lesssim_P \delta_\theta.$$

It follows that

$$\begin{aligned}
&\|\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}(\mathbf{U}_{\theta 2}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\|_F \\
&\le \sqrt{r_{12}}\|\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}(\mathbf{U}_{\theta 2}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\|_2 \\
&\le \sqrt{r_{12}}\|(\mathbf{U}_{\theta 1}^{[:,1:r_{12}]})^\top\|_2\|\mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\mathbf{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}(\mathbf{U}_{\theta 2}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\|_2 \\
&\lesssim_P \delta_\theta.
\end{aligned} \tag{S.26}$$

Define $\mathbf{\Gamma}_2^\star = \mathbf{V}_2\mathbf{\Lambda}_2^{-1/2}\mathbf{U}_{\theta 2}^\star$. Note that

$$\begin{aligned}
\mathbf{C}_1 = {} & \mathbf{\Sigma}_1\mathbf{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\mathbf{\Gamma}_1^{[:,1:r_{12}]})^\top\mathbf{X}_1 + \mathbf{\Sigma}_1\mathbf{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\mathbf{\Gamma}_2^{\star[:,1:r_{12}]})^\top\mathbf{X}_2 \\
& + \mathbf{\Sigma}_1\mathbf{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\mathbf{\Gamma}_2^{[:,1:r_{12}]} - \mathbf{\Gamma}_2^{\star[:,1:r_{12}]})^\top\mathbf{X}_2.
\end{aligned} \tag{S.27}$$

Let $\widehat{\boldsymbol{\Gamma}}_k = \widehat{\mathbf{V}}_k \widehat{\boldsymbol{\Lambda}}_k^{-1/2} \widehat{\mathbf{U}}_{\theta k}$ for $k = 1, 2$. By (S.21), (S.22) and (S.24), we have

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]} - \boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\|_2 \\
&= \|\widehat{\mathbf{V}}_1 \widehat{\boldsymbol{\Lambda}}_1^{-1/2} \widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]} - \mathbf{V}_1 \boldsymbol{\Lambda}_1^{-1/2} \mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\|_2 \\
&\leq \|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]}\|_2 \|\widehat{\mathbf{V}}_1 \widehat{\boldsymbol{\Lambda}}_1^{-1/2} - \mathbf{V}_1 \boldsymbol{\Lambda}_1^{-1/2}\|_2 + \|\mathbf{V}_1 \boldsymbol{\Lambda}_1^{-1/2}\|_2 \|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\|_F \\
&\lesssim_P \lambda_1^{-1/2}(\boldsymbol{\Sigma}_1) n^{-1/2} + \lambda_1^{-1/2}(\boldsymbol{\Sigma}_1) \delta_\theta \\
&=: \delta_\gamma^{(1)} \qquad \text{(S.28)}
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Gamma}}_2^{[:,1:r_{12}]} - \boldsymbol{\Gamma}_2^{\star[:,1:r_{12}]}\|_2 \\
&= \|\widehat{\mathbf{V}}_2 \widehat{\boldsymbol{\Lambda}}_2^{-1/2} \widehat{\mathbf{U}}_{\theta 2}^{[:,1:r_{12}]} - \mathbf{V}_2 \boldsymbol{\Lambda}_2^{-1/2} \mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]}\|_2 \\
&\lesssim_P \lambda_1^{-1/2}(\boldsymbol{\Sigma}_2) n^{-1/2} + \lambda_1^{-1/2}(\boldsymbol{\Sigma}_2) \delta_\theta \\
&=: \delta_\gamma^{(2)}. \qquad \text{(S.29)}
\end{aligned}
$$

By (S.21), (S.20) and (S.18),

$$
\begin{aligned}
&\|\widehat{\mathbf{V}}_1 \widehat{\boldsymbol{\Lambda}}_1^{1/2} - \mathbf{V}_1 \boldsymbol{\Lambda}_1^{1/2}\|_2 \\
&\leq \lambda_1^{1/2}(\boldsymbol{\Sigma}_1) \|\mathbf{V}_1 - \widehat{\mathbf{V}}_1\|_F + \|\widehat{\mathbf{V}}_1^\top\|_2 \max_{1 \leq \ell \leq r_1} |\lambda_\ell^{1/2}(\widehat{\boldsymbol{\Sigma}}_1) - \lambda_\ell^{1/2}(\boldsymbol{\Sigma}_1)| \\
&= O_P(\lambda_1^{1/2}(\boldsymbol{\Sigma}_1) n^{-1/2}). \qquad \text{(S.30)}
\end{aligned}
$$

Then by (S.24),

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Sigma}}_1 \widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]} - \boldsymbol{\Sigma}_1 \boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\|_2 \\
&= \|\widehat{\mathbf{V}}_1 \widehat{\boldsymbol{\Lambda}}_1^{1/2} \widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]} - \mathbf{V}_1 \boldsymbol{\Lambda}_1^{1/2} \mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\|_2 \\
&\leq \|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]}\|_2 \|\widehat{\mathbf{V}}_1 \widehat{\boldsymbol{\Lambda}}_1^{1/2} - \mathbf{V}_1 \boldsymbol{\Lambda}_1^{1/2}\|_2 + \|\mathbf{V}_1 \boldsymbol{\Lambda}_1^{1/2}\|_2 \|\widehat{\mathbf{U}}_{\theta 1}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 1}^{[:,1:r_{12}]}\|_2 \\
&\lesssim_P \lambda_1^{1/2}(\boldsymbol{\Sigma}_1) n^{-1/2} + \lambda_1^{1/2}(\boldsymbol{\Sigma}_1) \delta_\theta \\
&=: \delta_2. \qquad \text{(S.31)}
\end{aligned}
$$

Now consider the error bound for $\widehat{\mathbf{A}}_C^{(r)}$. Let $f(x) = \frac{1-x}{1+x}$. We notice that the derivative of $f^{\frac{1}{2}}(x)$ is unbound near $x = 1$. Thus, rather than using the mean value theorem directly for

$|f^{\frac{1}{2}}(\sigma_\ell(\widehat{\boldsymbol{\Theta}})) - f^{\frac{1}{2}}(\sigma_\ell(\boldsymbol{\Theta}))|$, we use the following technique:

$$\begin{aligned}\left|f^{\frac{1}{2}}(\sigma_\ell(\widehat{\boldsymbol{\Theta}})) - f^{\frac{1}{2}}(\sigma_\ell(\boldsymbol{\Theta}))\right|^2 &\le \left|f^{\frac{1}{2}}(\sigma_\ell(\widehat{\boldsymbol{\Theta}})) - f^{\frac{1}{2}}(\sigma_\ell(\boldsymbol{\Theta}))\right| \left|f^{\frac{1}{2}}(\sigma_\ell(\widehat{\boldsymbol{\Theta}})) + f^{\frac{1}{2}}(\sigma_\ell(\boldsymbol{\Theta}))\right| \\ &= \left|f(\sigma_\ell(\widehat{\boldsymbol{\Theta}})) - f(\sigma_\ell(\boldsymbol{\Theta}))\right| \\ &\le \sup_{0\le x\le 1} |f'(x)| \left|\sigma_\ell(\widehat{\boldsymbol{\Theta}}) - \sigma_\ell(\boldsymbol{\Theta})\right| \\ &\le \sup_{0\le x\le 1} \frac{2}{(x+1)^2} \left|\sigma_\ell(\widehat{\boldsymbol{\Theta}}) - \sigma_\ell(\boldsymbol{\Theta})\right| \\ &\lesssim_P \delta_\theta,\end{aligned}$$

where the last inequality holds uniformly for all $\ell = 1, \dots, r_{\min}$ due to (S.23). Hence,

$$\max_{1\le\ell\le r_{\min}} |\widehat{a}_\ell - a_\ell| \lesssim_P \delta_\theta^{1/2}. \tag{S.32}$$

From (S.21), (S.31), and (S.32),

$$\begin{aligned}&\|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]}\widehat{\mathbf{A}}_C^{(r_{12})} - \boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C\|_2 \\ &\le \|\widehat{\mathbf{A}}_C^{(r_{12})}\|_2\|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]} - \boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\|_2 + \|\mathbf{V}_1\boldsymbol{\Lambda}_1^{1/2}\mathbf{U}_{\theta1}^{[:,1:r_{12}]}\|_2\|\widehat{\mathbf{A}}_C^{(r_{12})} - \mathbf{A}_C\|_2 \\ &\lesssim_P \delta_2 + \lambda_1^{1/2}(\boldsymbol{\Sigma}_1)\delta_\theta^{1/2} \\ &=: \delta_{\sigma\gamma a}.\end{aligned} \tag{S.33}$$

Then by (S.21), (S.28), (S.17) and (S.7),

$$\begin{aligned}&\|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]}\widehat{\mathbf{A}}_C^{(r_{12})}(\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]})^\top - \boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\boldsymbol{\Gamma}_1^{[:,1:r_{12}]})^\top\|_2 \\ &\le \|(\widehat{\mathbf{U}}_{\theta1}^{[:,1:r_{12}]})^\top\widehat{\boldsymbol{\Lambda}}_1^{-1/2}\widehat{\mathbf{V}}_1^\top\|_2\|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]}\widehat{\mathbf{A}}_C^{(r_{12})} - \boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C\|_2 \\ &\qquad + \|\mathbf{V}_1\boldsymbol{\Lambda}_1^{1/2}\mathbf{U}_{\theta1}^{[:,1:r_{12}]}\mathbf{A}_C\|_2\|(\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]})^\top - (\boldsymbol{\Gamma}_1^{[:,1:r_{12}]})^\top\|_2 \\ &\lesssim_P \lambda_{r_1}^{-1/2}(\widehat{\boldsymbol{\Sigma}}_1)\delta_{\sigma\gamma a} + \lambda_1^{1/2}(\boldsymbol{\Sigma}_1)\delta_\gamma^{(1)} \\ &\lesssim_P \lambda_{r_1}^{-1/2}(\boldsymbol{\Sigma}_1)\delta_{\sigma\gamma a} + \lambda_1^{1/2}(\boldsymbol{\Sigma}_1)\delta_\gamma^{(1)} \\ &\lesssim_P \delta_\theta + \delta_\theta^{1/2} \\ &\lesssim_P \delta_\theta^{1/2} \\ &=: \delta_{1,1}.\end{aligned} \tag{S.34}$$

Similarly, from (S.29),

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]}\widehat{\mathbf{A}}_C^{(r_{12})}(\widehat{\boldsymbol{\Gamma}}_2^{[:,1:r_{12}]})^\top - \boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\boldsymbol{\Gamma}_2^{\star[:,1:r_{12}]})^\top\|_2 \\
&\lesssim_P \lambda_{r_2}^{-1/2}(\boldsymbol{\Sigma}_2)\delta_{\sigma\gamma a} + \lambda_1^{1/2}(\boldsymbol{\Sigma}_1)\delta_\gamma^{(2)} \\
&\lesssim_P \delta_{1,1}\lambda_1^{1/2}(\boldsymbol{\Sigma}_1)\lambda_1^{-1/2}(\boldsymbol{\Sigma}_2) \\
&=: \delta_{1,2}.
\end{aligned} \tag{S.35}
$$

By (S.21) and its variant under the Frobenius norm following from Lemma 1 in Lam and Fan (2009),

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]}\widehat{\mathbf{A}}_C^{(r_{12})}(\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]})^\top\widetilde{\mathbf{X}}_1 - \boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\boldsymbol{\Gamma}_1^{[:,1:r_{12}]})^\top\mathbf{X}_1\|_{(\cdot)} \\
&\le \|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]}\widehat{\mathbf{A}}_C^{(r_{12})}(\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]})^\top - \boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\boldsymbol{\Gamma}_1^{[:,1:r_{12}]})^\top\|_2\|\mathbf{X}_1\|_{(\cdot)} \\
&\qquad + \|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]}\widehat{\mathbf{A}}_C^{(r_{12})}(\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]})^\top\|_2\|\widetilde{\mathbf{X}}_1 - \mathbf{X}_1\|_{(\cdot)} \\
&\lesssim_P \delta_{1,1}\|\mathbf{X}_1\|_{(\cdot)} + \frac{1}{2}\lambda_1^{1/2}(\widehat{\boldsymbol{\Sigma}}_1)\lambda_{r_1}^{-1/2}(\widehat{\boldsymbol{\Sigma}}_1)\delta_{\mathbf{X}_1,(\cdot)} \\
&\lesssim_P \delta_{1,1}\|\mathbf{X}_1\|_{(\cdot)} + \lambda_1^{1/2}(\boldsymbol{\Sigma}_1)\lambda_{r_1}^{-1/2}(\boldsymbol{\Sigma}_1)\delta_{\mathbf{X}_1,(\cdot)} \\
&=: \delta_{C,(\cdot)}^{(1)}
\end{aligned} \tag{S.36}
$$

and similarly,

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r_{12}]}\widehat{\mathbf{A}}_C^{(r_{12})}(\widehat{\boldsymbol{\Gamma}}_2^{[:,1:r_{12}]})^\top\widetilde{\mathbf{X}}_2 - \boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\boldsymbol{\Gamma}_2^{\star[:,1:r_{12}]})^\top\mathbf{X}_2\|_{(\cdot)} \\
&\lesssim_P \delta_{1,2}\|\mathbf{X}_2\|_{(\cdot)} + \lambda_1^{1/2}(\boldsymbol{\Sigma}_1)\lambda_{r_2}^{-1/2}(\boldsymbol{\Sigma}_2)\delta_{\mathbf{X}_2,(\cdot)} \\
&=: \delta_{C,(\cdot)}^{(2)}.
\end{aligned} \tag{S.37}
$$

By the fact that $1 - \sqrt{\frac{1-x}{1+x}} \le 1 - \frac{1-x}{1+x} \le 2x$ for $x \in [0,1]$ and inequality (S.26), we have

$$
\|\mathbf{A}_C(\mathbf{U}_{\theta 2}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\|_F \le \|\boldsymbol{\Lambda}_\theta^{[1:r_{12},1:r_{12}]}(\mathbf{U}_{\theta 2}^{[:,1:r_{12}]} - \mathbf{U}_{\theta 2}^{\star[:,1:r_{12}]})^\top\|_F \lesssim_P \delta_\theta.
$$

It follows that

$$
\begin{aligned}
&\|\boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\boldsymbol{\Gamma}_2^{[:,1:r_{12}]}-\boldsymbol{\Gamma}_2^{\star[:,1:r_{12}]})^\top\mathbf{X}_2\|_{(.)}\\
&\le \|\mathbf{V}_1\boldsymbol{\Lambda}_1^{1/2}\mathbf{U}_{\theta1}^{[:,1:r_{12}]}\mathbf{A}_C(\mathbf{U}_{\theta2}^{[:,1:r_{12}]}-\mathbf{U}_{\theta2}^{\star[:,1:r_{12}]})^\top\boldsymbol{\Lambda}_2^{-1/2}\mathbf{V}_2^\top\|_2\|\mathbf{X}_2\|_{(\cdot)}\\
&\lesssim_P \lambda_1^{1/2}(\boldsymbol{\Sigma}_1)\lambda_{r_2}^{-1/2}(\boldsymbol{\Sigma}_2)\|\mathbf{X}_2\|_{(\cdot)}\delta_\theta\\
&\lesssim_P \delta_{1,2}\|\mathbf{X}_2\|_{(\cdot)}. \qquad \text{(S.38)}
\end{aligned}
$$

By the definition of $\widehat{\mathbf{C}}_1$, (S.27), (S.36), (S.37) and (S.38), we obtain

$$
\|\widehat{\mathbf{C}}_1-\mathbf{C}_1\|_{(\cdot)} \lesssim_P \delta_{C,(\cdot)}^{(1)}+\delta_{C,(\cdot)}^{(2)}. \qquad \text{(S.39)}
$$

Together with (S.15), (S.13) and (S.14), we obtain the claimed bound for $\|\widehat{\mathbf{C}}_1-\mathbf{C}_1\|_{(\cdot)}/\|\mathbf{C}_1\|_{(\cdot)}$.

Now consider the relative error bound for $\widehat{\mathbf{D}}_1$. Write $\widehat{\mathbf{C}}_1^{(r)}$ equivalently by

$$
\widehat{\mathbf{C}}_1^{(r)}=\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Gamma}}_1^{[:,1:r]}\widehat{\mathbf{A}}_C^{(r)}\sum_{k=1}^{2}(\widehat{\boldsymbol{\Gamma}}_k^{[:,1:r]})^\top\widetilde{\mathbf{X}}_k.
$$

Note that $\widetilde{\mathbf{X}}_1=\widehat{\mathbf{C}}_1^{(\widetilde{r}_{12})}+\widehat{\mathbf{D}}_1$. We have

$$
\|\widehat{\mathbf{D}}_1-\mathbf{D}_1\|_{(\cdot)} \le \|\widehat{\mathbf{C}}_1^{(\widetilde{r}_{12})}-\mathbf{C}_1\|_{(\cdot)}+\|\widetilde{\mathbf{X}}_1-\mathbf{X}_1\|_{(\cdot)}. \qquad \text{(S.40)}
$$

When $\widetilde{r}_{12}\le r_{12}$, $\widehat{\mathbf{C}}_1=\widehat{\mathbf{C}}_1^{(\widetilde{r}_{12})}$ and thus $\|\widehat{\mathbf{D}}_1-\mathbf{D}_1\|_{(\cdot)}\le\|\widehat{\mathbf{C}}_1-\mathbf{C}_1\|_{(\cdot)}+\|\widetilde{\mathbf{X}}_1-\mathbf{X}_1\|_{(\cdot)}$, immediately leading to the bound for $\|\widehat{\mathbf{D}}_1-\mathbf{D}_1\|_{(\cdot)}/\|\mathbf{D}_1\|_{(\cdot)}$. Now consider the case when $\widetilde{r}_{12}>r_{12}$. Let $r\in(r_{12},r_{\min}]$. We first look at $\|\widehat{\mathbf{C}}_1^{(r)}-\mathbf{C}_1\|_{(\cdot)}$. Define $\widetilde{\boldsymbol{\Gamma}}_k^{(r)}=\mathbf{V}_k\boldsymbol{\Lambda}_k^{-1/2}(\mathbf{U}_{\theta k}^{[:,1:r_{12}]},\widehat{\mathbf{U}}_{\theta k}^{[:,(r_{12}+1):r]})$ and $\widetilde{\boldsymbol{\Gamma}}_2^{\star(r)}=\mathbf{V}_2\boldsymbol{\Lambda}_2^{-1/2}(\mathbf{U}_{\theta2}^{\star[:,1:r_{12}]},\widehat{\mathbf{U}}_{\theta2}^{[:,(r_{12}+1):r]})$. We have

$$
\begin{aligned}
\mathbf{C}_1=\boldsymbol{\Sigma}_1\widetilde{\boldsymbol{\Gamma}}_1^{(r)}\mathbf{A}_C^{(r)}(\widetilde{\boldsymbol{\Gamma}}_1^{(r)})^\top\mathbf{X}_1+\boldsymbol{\Sigma}_1\widetilde{\boldsymbol{\Gamma}}_1^{(r)}\mathbf{A}_C^{(r)}(\widetilde{\boldsymbol{\Gamma}}_2^{\star(r)})^\top\mathbf{X}_2&\\
+\boldsymbol{\Sigma}_1\boldsymbol{\Gamma}_1^{[:,1:r_{12}]}\mathbf{A}_C(\boldsymbol{\Gamma}_2^{[:,1:r_{12}]}-\boldsymbol{\Gamma}_2^{\star[:,1:r_{12}]})^\top\mathbf{X}_2& \qquad \text{(S.41)}
\end{aligned}
$$

with $\mathbf{A}_C^{(r)}\coloneqq\mathrm{diag}(a_1,\dots,a_r)$ and $a_\ell=0$ for $\ell>r_{12}$. By (S.24) and (S.25),

$$
\begin{aligned}
\max\Big\{&\|(\widehat{\mathbf{U}}_{\theta1}^{[:,1:r_{12}]},\widehat{\mathbf{U}}_{\theta1}^{[:,(r_{12}+1):r]})-(\mathbf{U}_{\theta1}^{[:,1:r_{12}]},\widehat{\mathbf{U}}_{\theta1}^{[:,(r_{12}+1):r]})\|_F,\\
&\|(\widehat{\mathbf{U}}_{\theta2}^{[:,1:r_{12}]},\widehat{\mathbf{U}}_{\theta2}^{[:,(r_{12}+1):r]})-(\mathbf{U}_{\theta2}^{\star[:,1:r_{12}]},\widehat{\mathbf{U}}_{\theta2}^{[:,(r_{12}+1):r]})\|_F\Big\}\lesssim_P\delta_\theta. \qquad \text{(S.42)}
\end{aligned}
$$

Then following the proof lines for (S.28), (S.29), (S.31) and (S.33)-(S.35), we can obtain

$$\|\widehat{\mathbf{\Gamma}}_1^{[:,1:r]} - \widetilde{\mathbf{\Gamma}}_1^{(r)}\|_2 \lesssim_P \delta_\gamma^{(1)},$$

$$\|\widehat{\mathbf{\Gamma}}_2^{[:,1:r]} - \widetilde{\mathbf{\Gamma}}_2^{\star(r)}\|_2 \lesssim_P \delta_\gamma^{(2)},$$

$$\|\widehat{\mathbf{\Sigma}}_1\widehat{\mathbf{\Gamma}}_1^{[:,1:r]} - \mathbf{\Sigma}_1\widetilde{\mathbf{\Gamma}}_1^{(r)}\|_2 \lesssim_P \delta_2,$$

$$\|\widehat{\mathbf{\Sigma}}_1\widehat{\mathbf{\Gamma}}_1^{[:,1:r]}\widehat{\mathbf{A}}_C^{(r)} - \mathbf{\Sigma}_1\widetilde{\mathbf{\Gamma}}_1^{(r)}\mathbf{A}_C^{(r)}\|_2 \lesssim_P \delta_{a\gamma a},$$

$$\|\widehat{\mathbf{\Sigma}}_1\widehat{\mathbf{\Gamma}}_1^{[:,1:r]}\widehat{\mathbf{A}}_C^{(r)}(\widehat{\mathbf{\Gamma}}_1^{[:,1:r]})^\top - \mathbf{\Sigma}_1\widetilde{\mathbf{\Gamma}}_1^{(r)}\mathbf{A}_C^{(r)}(\widetilde{\mathbf{\Gamma}}_1^{(r)})^\top\|_2 \lesssim_P \delta_{1,1}$$

and

$$\|\widehat{\mathbf{\Sigma}}_1\widehat{\mathbf{\Gamma}}_1^{[:,1:r]}\widehat{\mathbf{A}}_C^{(r)}(\widehat{\mathbf{\Gamma}}_1^{[:,1:r]})^\top - \mathbf{\Sigma}_1\widetilde{\mathbf{\Gamma}}_1^{(r)}\mathbf{A}_C^{(r)}(\widetilde{\mathbf{\Gamma}}_2^{\star(r)})^\top\|_2 \lesssim_P \delta_{1,2}.$$

Following the derivation of (S.36) and (S.37), we can obtain

$$\|\widehat{\mathbf{\Sigma}}_1\widehat{\mathbf{\Gamma}}_1^{[:,1:r]}\widehat{\mathbf{A}}_C^{(r)}(\widehat{\mathbf{\Gamma}}_1^{[:,1:r]})^\top\widetilde{\mathbf{X}}_1 - \mathbf{\Sigma}_1\widetilde{\mathbf{\Gamma}}_1^{(r)}\mathbf{A}_C^{(r)}(\widetilde{\mathbf{\Gamma}}_1^{(r)})^\top\mathbf{X}_1\|_{(\cdot)} \lesssim_P \delta_{C,(\cdot)}^{(1)}$$

and

$$\|\widehat{\mathbf{\Sigma}}_1\widehat{\mathbf{\Gamma}}_1^{[:,1:r]}\widehat{\mathbf{A}}_C^{(r)}(\widehat{\mathbf{\Gamma}}_2^{[:,1:r]})^\top\widetilde{\mathbf{X}}_2 - \mathbf{\Sigma}_1\widetilde{\mathbf{\Gamma}}_1^{(r)}\mathbf{A}_C^{(r)}(\widetilde{\mathbf{\Gamma}}_2^{\star(r)})^\top\mathbf{X}_2\|_{(\cdot)} \lesssim_P \delta_{C,(\cdot)}^{(2)}.$$

By the above two inequalities, (S.38), the definition of $\widehat{\mathbf{C}}_1^{(r)}$, and (S.41), we obtain

$$\|\widehat{\mathbf{C}}_1^{(r)} - \mathbf{C}_1\|_{(\cdot)} \lesssim_P \delta_{C,(\cdot)}^{(1)} + \delta_{C,(\cdot)}^{(2)}, \tag{S.43}$$

which has the same bound for $\|\widehat{\mathbf{C}}_1 - \mathbf{C}_1\|_{(\cdot)}$ given in (S.39). Then using (S.40) gives

$$\|\widehat{\mathbf{D}}_1 - \mathbf{D}_1\|_{(\cdot)} \lesssim_P \delta_{C,(\cdot)}^{(1)} + \delta_{C,(\cdot)}^{(2)} \tag{S.44}$$

and the claimed bound for $\|\widehat{\mathbf{D}}_1 - \mathbf{D}_1\|_{(\cdot)}/\|\mathbf{D}_1\|_{(\cdot)}$ in the theorem.

The relative error bound for $\widehat{\mathbf{X}}_1$ immediately follows from

$$\|\widehat{\mathbf{X}}_1 - \mathbf{X}_1\|_{(\cdot)} \le \|\widehat{\mathbf{C}}_1 - \mathbf{C}_1\|_{(\cdot)} + \|\widehat{\mathbf{D}}_1 - \mathbf{D}_1\|_{(\cdot)} \lesssim_P \delta_{C,(\cdot)}^{(1)} + \delta_{C,(\cdot)}^{(2)}.$$

Similarly, we can obtain the bounds for estimated matrices of the second dataset. □

Proof of Corollary 1. For $k = 1, 2$, since $\breve{r}_k \xrightarrow{P} r_k$ and $\breve{r}_k$ is an integer, we have $\mathbb{P}(\breve{r}_k = r_k) \to 1$ as $n \to \infty$. Due to Lemma S.1, in this proof we simply assume $\breve{r}_k = r_k$. Hence, we only need to prove the relative error bounds for $\widehat{\mathbf{C}}_k^{(r)}$ and $\widehat{\mathbf{X}}_k^{(r)}$, and refer the other two bounds to Theorem 3.

When $\widetilde{r}_{12} \leq r_{12}$, we have $\widetilde{r}_{12} = \min(r_{12}, \widetilde{r}_{12}) \leq r \leq r_{\min}$ and thus $\widehat{\mathbf{C}}_k^{(r)} = \widehat{\mathbf{C}}_k^{(\widetilde{r}_{12})} = \widehat{\mathbf{C}}_k$. Then, the result stated in the corollary has been given in Theorem 3. On the other hand, when $\widetilde{r}_{12} > r_{12}$, we have $r_{12} = \min(r_{12}, \widetilde{r}_{12}) \leq r \leq r_{\min}$. Then by (S.43), we can immediately obtain the claimed result in the corollary. □

S.2 Additional Simulations

We consider Setups 1* and 2* which have the same settings as those in Setups 1 and 2, respectively, except for the noise covariance matrices $\text{cov}(\boldsymbol{e}_k) = (0.7^{|i-j|}\sigma_e^2)_{1 \leq i,j \leq p_k}$, $k = 1, 2$. Note that $\lambda_1(\text{cov}(\boldsymbol{e}_k)) \in (5.62\sigma_e^2, 5.67\sigma_e^2)$ for $100 \leq p_k \leq 1500$. Especially when $\sigma_e^2 = 16$, $\lambda_1(\text{cov}(\boldsymbol{e}_k)) \approx 90$ is quite close to 100 that is the minimum nonzero eigenvalue of $\boldsymbol{\Sigma}_k$, resulting in challenging cases for estimation (see conditions (I), (II) and (V) in Assumption 1, and also the result in (S.6)). The finite sample performance of our D-CCA estimates shown in Table S.1 and Figures S.1 and S.2 is similar to that in Table 1 and Figures 3 and 4.

Table S.1: Averages (standard errors) of D-CCA estimates for the first canonical angle/correlation.

(p_1, σ_e^2)	$\theta_1 = 0°/\rho_1 = 1$	$\theta_1 = 45°/\rho_1 = 0.707$	$\theta_1 = 60°/\rho_1 = 0.5$	$\theta_1 = 75°/\rho_1 = 0.259$
		Setup 1*		
$(100, 1)$	$4.15°(0.24°)/0.997(0.000)$	$44.7°(2.39°)/0.710(0.029)$	$59.4°(2.89°)/0.509(0.043)$	$73.5°(3.08°)/0.283(0.051)$
$(600, 1)$	$3.65°(0.22°)/0.998(0.000)$	$44.7°(2.39°)/0.710(0.029)$	$59.4°(2.89°)/0.509(0.043)$	$73.5°(3.08°)/0.283(0.051)$
$(900, 1)$	$3.65°(0.22°)/0.998(0.000)$	$44.7°(2.39°)/0.710(0.029)$	$59.4°(2.89°)/0.509(0.043)$	$73.5°(3.07°)/0.283(0.051)$
$(1500, 1)$	$3.64°(0.22°)/0.998(0.000)$	$44.7°(2.38°)/0.710(0.029)$	$59.4°(2.89°)/0.509(0.043)$	$73.5°(3.08°)/0.283(0.051)$
$(900, 0.01)$	$0.36°(0.02°)/1.000(0.000)$	$44.6°(2.38°)/0.712(0.029)$	$59.3°(2.89°)/0.510(0.043)$	$73.5°(3.08°)/0.284(0.051)$
$(900, 1)$	$3.65°(0.22°)/0.998(0.000)$	$44.7°(2.39°)/0.710(0.029)$	$59.4°(2.89°)/0.509(0.043)$	$73.5°(3.07°)/0.283(0.051)$
$(900, 9)$	$12.1°(0.81°)/0.978(0.003)$	$45.9°(2.46°)/0.696(0.031)$	$60.1°(2.92°)/0.499(0.044)$	$73.9°(3.05°)/0.277(0.051)$
$(900, 16)$	$17.6°(1.28°)/0.953(0.007)$	$47.4°(2.57°)/0.676(0.033)$	$61.1°(2.98°)/0.482(0.046)$	$74.9°(3.16°)/0.260(0.053)$
		Setup 2*		
$(100, 1)$	$3.97°(0.24°)/0.998(0.000)$	$44.5°(2.36°)/0.712(0.029)$	$59.0°(2.82°)/0.514(0.042)$	$72.7°(2.88°)/0.296(0.048)$
$(600, 1)$	$3.72°(0.23°)/0.998(0.000)$	$44.5°(2.36°)/0.712(0.029)$	$59.0°(2.82°)/0.514(0.042)$	$72.7°(2.88°)/0.296(0.048)$
$(900, 1)$	$3.72°(0.22°)/0.998(0.000)$	$44.5°(2.36°)/0.712(0.029)$	$59.0°(2.83°)/0.514(0.042)$	$72.7°(2.88°)/0.297(0.048)$
$(1500, 1)$	$3.72°(0.23°)/0.998(0.000)$	$44.5°(2.37°)/0.712(0.029)$	$59.0°(2.84°)/0.514(0.043)$	$72.7°(2.89°)/0.296(0.048)$
$(900, 0.01)$	$0.37°(0.02°)/1.000(0.000)$	$44.4°(2.35°)/0.714(0.029)$	$59.0°(2.82°)/0.515(0.042)$	$72.7°(2.89°)/0.297(0.048)$
$(900, 1)$	$3.72°(0.22°)/0.998(0.000)$	$44.5°(2.36°)/0.712(0.029)$	$59.0°(2.83°)/0.514(0.042)$	$72.7°(2.88°)/0.297(0.048)$
$(900, 9)$	$12.0°(0.79°)/0.978(0.003)$	$45.6°(2.42°)/0.698(0.030)$	$59.7°(2.86°)/0.505(0.043)$	$73.0°(2.89°)/0.292(0.048)$
$(900, 16)$	$17.3°(2.73°)/0.954(0.030)$	$47.1°(2.53°)/0.680(0.032)$	$60.7°(2.90°)/0.488(0.044)$	$74.0°(2.97°)/0.275(0.050)$

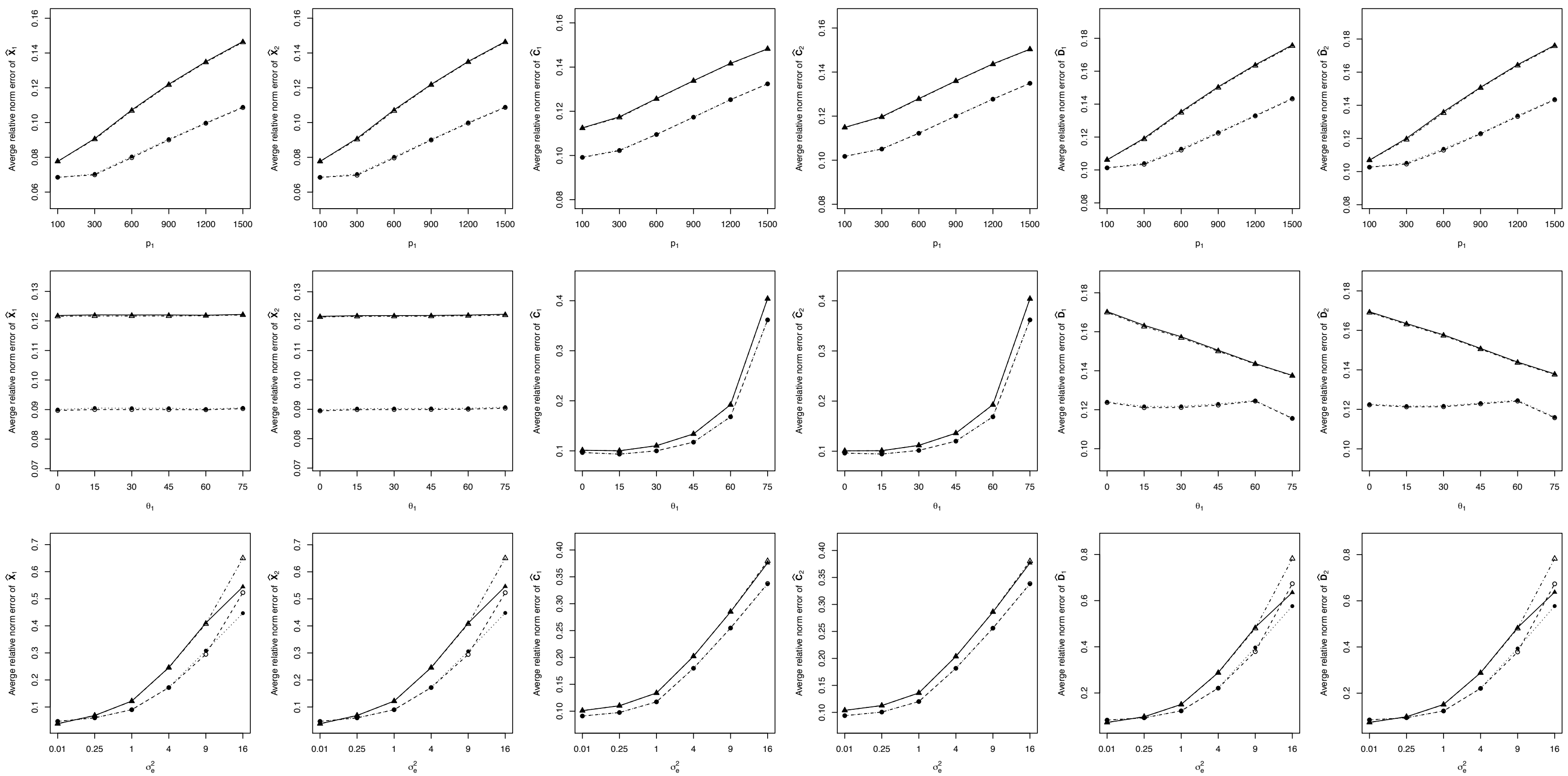

Figure S.1: Average relative errors of D-CCA estimates under Setup 1* in spectral norm (○) and Frobenius norm (△) using true r_1, r_2 and r_{12}, and those in spectral norm (●) and Frobenius norm (▲) using $\widehat{r}_1, \widehat{r}_2$ and $\widehat{r}_{12}$.

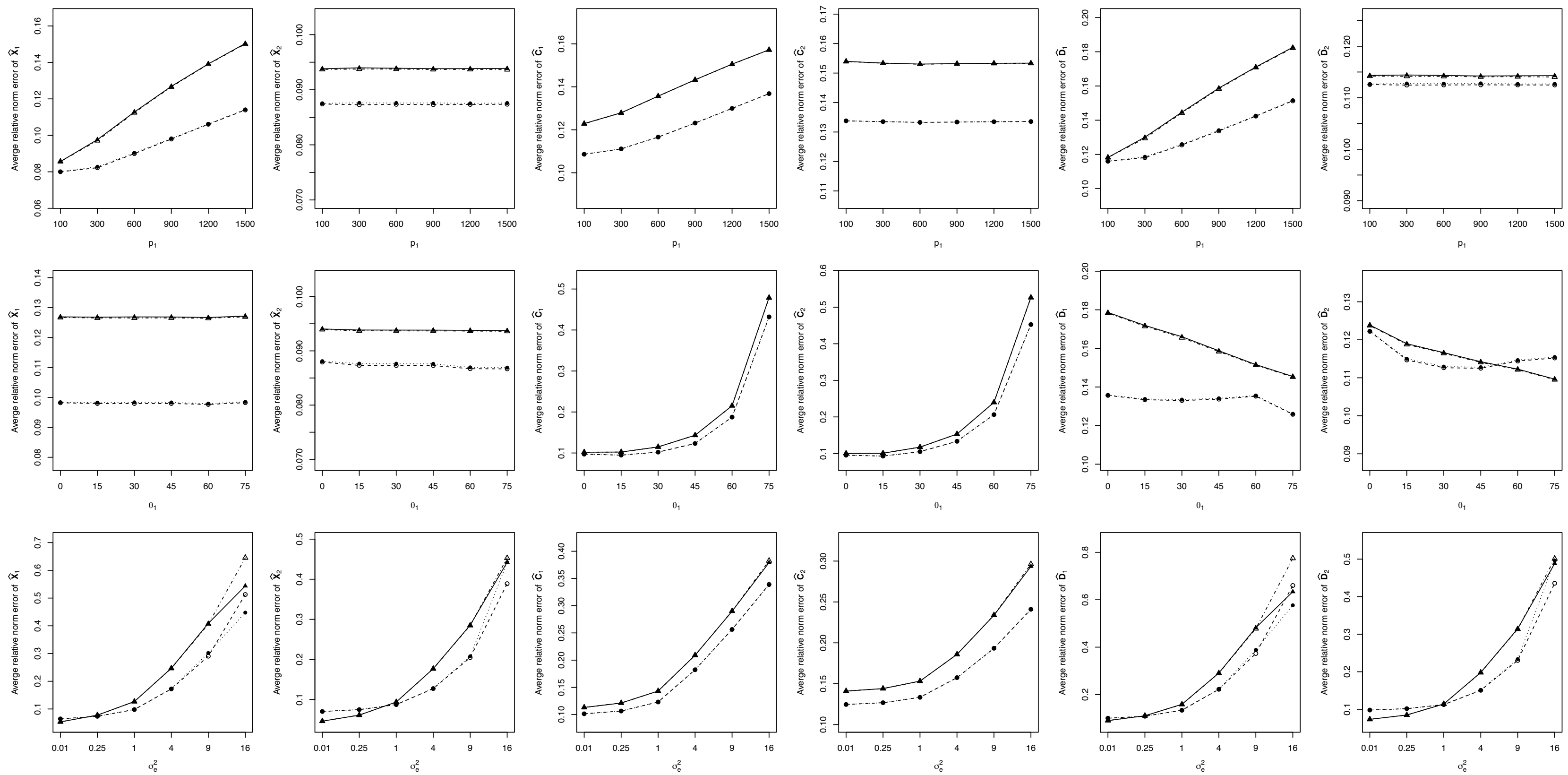

Figure S.2: Average relative errors of D-CCA estimates under Setup 2* in spectral norm (○) and Frobenius norm (△) using true r_1, r_2 and r_{12}, and those in spectral norm (●) and Frobenius norm (▲) using $\widehat{r}_1, \widehat{r}_2$ and $\widehat{r}_{12}$.

References

Hallin, M. and Liška, R. (2011), “Dynamic factors in the presence of blocks,” *Journal of Econometrics*, 163, 29–41.

Horn, R. A. and Johnson, C. R. (1994), *Topics in Matrix Analysis*, Cambridge University Press, Cambridge.

Lam, C. and Fan, J. (2009), “Sparsistency and rates of convergence in large covariance matrix estimation,” *The Annals of Statistics*, 37, 4254–4278.

Wang, W. and Fan, J. (2017), “Asymptotics of empirical eigenstructure for high dimensional spiked covariance,” *The Annals of Statistics*, 45, 1342–1374.

Yu, Y., Wang, T., and Samworth, R. J. (2015), “A useful variant of the Davis–Kahan theorem for statisticians,” *Biometrika*, 102, 315–323.