Extended Results Made Available to Readers for "SOFARI: High-Dimensional Manifold-Based Inference"

This Extended Results Made Available to Readers contains the proofs of all lemmas, additional technical details and real data results. All the notation is the same as defined in the main body and the Supplementary Material of the paper.

ER.1. Proofs of Lemmas

ER.1.1. Proof of Lemma 1

When $r^*=2$, the loss function (2) can be written as

$$\begin{aligned}
&L(\boldsymbol{u}_1,\boldsymbol{\eta}_1)=(2n)^{-1}\|\mathbf{Y}-\mathbf{X}\boldsymbol{u}_1\boldsymbol{v}_1^T-\mathbf{X}\boldsymbol{u}_2\boldsymbol{v}_2^T\|_F^2\\
&\text{subject to }\ \boldsymbol{u}_1^T\boldsymbol{u}_2=0 \text{ and } \left[\boldsymbol{v}_1,\boldsymbol{v}_2\right]^T\left[\boldsymbol{v}_1,\boldsymbol{v}_2\right]=\mathbf{I}_2,
\end{aligned}$$

where $\boldsymbol{\eta}_1=\left(\boldsymbol{v}_1^T,\boldsymbol{v}_2^T,\boldsymbol{u}_2^T\right)^T$. Then under the orthogonality constraint $\boldsymbol{v}_1^T\boldsymbol{v}_2=0$, it can be simplified as

$$\begin{aligned}
L=(2n)^{-1}\Big\{&\|\mathbf{Y}\|_F^2+\boldsymbol{u}_1^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}_1\boldsymbol{v}_1^T\boldsymbol{v}_1+\boldsymbol{u}_2^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}_2\boldsymbol{v}_2^T\boldsymbol{v}_2\\
&-2\boldsymbol{u}_1^T\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_1-2\boldsymbol{u}_2^T\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_2\Big\}.
\end{aligned} \tag{ER.1}$$

After some calculations with $\|\boldsymbol{v}_1\|_2=\|\boldsymbol{v}_2\|_2=1$, we can deduce that

$$\frac{\partial L}{\partial \boldsymbol{u}_1}=\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1-n^{-1}\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_1, \tag{ER.2}$$

$$\frac{\partial L}{\partial \boldsymbol{v}_1}=\boldsymbol{v}_1\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1-n^{-1}\mathbf{Y}^T\mathbf{X}\boldsymbol{u}_1, \tag{ER.3}$$

$$\frac{\partial L}{\partial \boldsymbol{u}_2}=\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2-n^{-1}\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_2, \tag{ER.4}$$

$$\frac{\partial L}{\partial \boldsymbol{v}_2}=\boldsymbol{v}_2\boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2-n^{-1}\mathbf{Y}^T\mathbf{X}\boldsymbol{u}_2. \tag{ER.5}$$

Utilizing the derivatives (ER.3)–(ER.5) with some calculations, it follows that

$$\begin{aligned}
\mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_1}\Big|_{\boldsymbol{\eta}_1^*}&=\mathbf{M}_1\left\{\boldsymbol{v}_1^*\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1-\boldsymbol{u}_1^*)-\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1-n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1\right\}-n^{-1}\mathbf{M}_3\mathbf{X}^T\mathbf{E}\boldsymbol{v}_2^*\\
&\quad+\mathbf{M}_2\left\{-\boldsymbol{v}_1^*\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*-n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_2^*\right\}\\
&=(\mathbf{M}_1\boldsymbol{v}_1^*\boldsymbol{u}_1^T-\mathbf{M}_1\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T})\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1-\boldsymbol{u}_1^*)-(\mathbf{M}_1\boldsymbol{v}_2^*+\mathbf{M}_2\boldsymbol{v}_1^*)\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*\\
&\quad-n^{-1}(\mathbf{M}_1\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1+\mathbf{M}_3\mathbf{X}^T\mathbf{E}\boldsymbol{v}_2^*+\mathbf{M}_2\mathbf{E}^T\mathbf{X}\boldsymbol{u}_2^*)\\
&=(\mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^T-\mathbf{M}_1\boldsymbol{v}_2\boldsymbol{u}_2^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1-\boldsymbol{u}_1^*)-(\mathbf{M}_1\boldsymbol{v}_2^*+\mathbf{M}_2\boldsymbol{v}_1^*)\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*+\boldsymbol{\delta}_1',
\end{aligned}$$

where we set

$$\boldsymbol{\delta}_1' = -\left\{\mathbf{M}_1(\boldsymbol{v}_1 - \boldsymbol{v}_1^*)\boldsymbol{u}_1^T - \mathbf{M}_1(\boldsymbol{v}_2\boldsymbol{u}_2^T - \boldsymbol{v}_2^*\boldsymbol{u}_2^{*T})\right\}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1 - \boldsymbol{u}_1^*) \\ - n^{-1}(\mathbf{M}_1\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1 + \mathbf{M}_2\mathbf{E}^T\mathbf{X}\boldsymbol{u}_2^* + \mathbf{M}_3\mathbf{X}^T\mathbf{E}\boldsymbol{v}_2^*). \tag{ER.6}$$

Together with the derivative (ER.2), it holds that

$$\widetilde{\psi}(\boldsymbol{u}_1, \boldsymbol{\eta}_1^*) = \frac{\partial L}{\partial \boldsymbol{u}_1}\Big|_{\boldsymbol{\eta}_1^*} - \mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_1}\Big|_{\boldsymbol{\eta}_1^*} \\ = (\mathbf{I}_p - \mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^T + \mathbf{M}_1\boldsymbol{v}_2\boldsymbol{u}_2^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1 - \boldsymbol{u}_1^*) + (\mathbf{M}_1\boldsymbol{v}_2^* + \mathbf{M}_2\boldsymbol{v}_1^*)\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^* + \boldsymbol{\delta}_1, \tag{ER.7}$$

where $\boldsymbol{\delta}_1 = -\boldsymbol{\delta}_1' - n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*$.

Therefore, combining (ER.6) and (ER.7), we can obtain that

$$\widetilde{\psi}(\boldsymbol{u}_1, \boldsymbol{\eta}_1^*) = (\mathbf{I}_p - \mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^T + \mathbf{M}_1\boldsymbol{v}_2\boldsymbol{u}_2^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1 - \boldsymbol{u}_1^*) + (\mathbf{M}_1\boldsymbol{v}_2^* + \mathbf{M}_2\boldsymbol{v}_1^*)\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^* \\ + \boldsymbol{\delta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\delta} = \mathbf{M}_1\left\{(\boldsymbol{v}_1 - \boldsymbol{v}_1^*)\boldsymbol{u}_1^T - (\boldsymbol{v}_2\boldsymbol{u}_2^T - \boldsymbol{v}_2^*\boldsymbol{u}_2^{*T})\right\}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1 - \boldsymbol{u}_1^*)$ and

$$\boldsymbol{\epsilon} = n^{-1}\left\{\mathbf{M}_1\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1 + \mathbf{M}_2\mathbf{E}^T\mathbf{X}\boldsymbol{u}_2^* + \mathbf{M}_3\mathbf{X}^T\mathbf{E}\boldsymbol{v}_2^*\right\} - n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*.$$

This completes the proof of Lemma 1.

ER.1.2. Proof of Lemma 2

For the rank-2 case with strongly orthogonal factors, since the technical arguments for the inference of $\boldsymbol{u}_1^*$ and $\boldsymbol{u}_2^*$ are similar, for simplicity we will present the proof only for $\boldsymbol{u}_1^*$ here. When we use the initial estimates $(\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{u}}_2, \widetilde{\boldsymbol{v}}_1, \widetilde{\boldsymbol{v}}_2)$ satisfying Definition 2, by Lemma EC.16 the first-order Taylor expansion of $\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{\eta}}_1)$ at $\boldsymbol{\eta}_1^*$ is given by

$$\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{\eta}}_1) = \widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \boldsymbol{\eta}_1^*) + (-n^{-1}\mathbf{X}^T\mathbf{Y} - \widetilde{\boldsymbol{u}}_1^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_1\mathbf{M}_1)(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\exp^{-1}_{\boldsymbol{v}_1^*}(\widetilde{\boldsymbol{v}}_1) \\ + (n^{-1}\mathbf{M}_2\mathbf{X}^T\mathbf{Y} - \widetilde{\boldsymbol{u}}_2^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\mathbf{M}_3)(\mathbf{I}_q - \boldsymbol{v}_2^*\boldsymbol{v}_2^{*T})\exp^{-1}_{\boldsymbol{v}_2^*}(\widetilde{\boldsymbol{v}}_2) \\ + (-\mathbf{M}_2\widehat{\boldsymbol{\Sigma}} - 2\mathbf{M}_3\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}} + n^{-1}\mathbf{M}_3\mathbf{Y}^T\mathbf{X})(\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*) + \boldsymbol{r}_{\boldsymbol{v}_1^*} + \boldsymbol{r}_{\boldsymbol{u}_2^*} + \boldsymbol{r}_{\boldsymbol{v}_2^*},$$

where the Taylor remainder terms satisfy that

$$\|\boldsymbol{r}_{\boldsymbol{v}_1^*}\|_2 = O(\|\exp^{-1}_{\boldsymbol{v}_1^*}(\widetilde{\boldsymbol{v}}_1)\|_2^2),\ \|\boldsymbol{r}_{\boldsymbol{u}_2^*}\|_2 = O(\|\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*\|_2^2),\ \|\boldsymbol{r}_{\boldsymbol{v}_2^*}\|_2 = O(\|\exp^{-1}_{\boldsymbol{v}_2^*}(\widetilde{\boldsymbol{v}}_2)\|_2^2).$$

Moreover, when the construction of $\mathbf{M} = [\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3]$ is given as

$$\mathbf{M}_1 = -(\widetilde{\boldsymbol{u}}_1^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_1)^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T,\ \mathbf{M}_2 = \mathbf{0},\ \mathbf{M}_3 = \mathbf{0},$$

it is immediate to see that

$$\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{\eta}}_1) = \widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \boldsymbol{\eta}_1^*) + (-n^{-1}\mathbf{X}^T\mathbf{Y} + \widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T)(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\exp^{-1}_{\boldsymbol{v}_1^*}(\widetilde{\boldsymbol{v}}_1)$$
$$+ \boldsymbol{r}_{\boldsymbol{v}_1^*} + \boldsymbol{r}_{\boldsymbol{u}_2^*} + \boldsymbol{r}_{\boldsymbol{v}_2^*}. \qquad \text{(ER.8)}$$

We aim to bound the difference between $\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{\eta}}_1)$ and $\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \boldsymbol{\eta}_1^*)$, which will be divided into two parts.

(1). Upper bounds on $\|\exp^{-1}_{\boldsymbol{v}_1^*}(\widetilde{\boldsymbol{v}}_1)\|_0$**,** $\|\exp^{-1}_{\boldsymbol{v}_1^*}(\widetilde{\boldsymbol{v}}_1)\|_2$**,** $\|\exp^{-1}_{\boldsymbol{v}_2^*}(\widetilde{\boldsymbol{v}}_2)\|_0$**, and** $\|\exp^{-1}_{\boldsymbol{v}_2^*}(\widetilde{\boldsymbol{v}}_2)\|_2$. Denote by $\boldsymbol{\xi}_1 = \exp^{-1}_{\boldsymbol{v}_1^*}(\widetilde{\boldsymbol{v}}_1)$ the tangent vector, so that $\exp_{\boldsymbol{v}_1^*}(\boldsymbol{\xi}_1) = \widetilde{\boldsymbol{v}}_1$. Since $\boldsymbol{r}_{\boldsymbol{v}_1^*} = O(\|\boldsymbol{\xi}_1\|_2^2)$, if $\boldsymbol{\xi}_1 = \mathbf{0}$ we need only to bound term $\boldsymbol{r}_{\boldsymbol{u}_2^*} + \boldsymbol{r}_{\boldsymbol{v}_2^*}$ in (ER.8) above. Without loss of generality, let us assume that $\boldsymbol{\xi}_1 \neq \mathbf{0}$. Observe that the q-dimensional tangent vector $\boldsymbol{\xi}_1 \in T_{\boldsymbol{v}_1^*}\operatorname{St}(1,q)$, where $T_{\boldsymbol{v}_1^*}\operatorname{St}(1,q)$ is the tangent space of the Stiefel manifold $\operatorname{St}(1,q)$ at $\boldsymbol{v}_1^*$. Then from Lemma EC.29, we have the explicit form of the corresponding geodesic

$$\gamma(t; \boldsymbol{v}_1^*, \boldsymbol{\xi}_1) = \boldsymbol{v}_1^* \cdot \cos(\|\boldsymbol{\xi}_1\|_2 t) + \frac{\boldsymbol{\xi}_1}{\|\boldsymbol{\xi}_1\|_2} \cdot \sin(\|\boldsymbol{\xi}_1\|_2 t).$$

Hence, in view of the definition of the exponential map in (ER.139), it holds that

$$\begin{aligned} \widetilde{\boldsymbol{v}}_1 &= \exp_{\boldsymbol{v}_1^*}(\boldsymbol{\xi}_1) = \gamma(1; \boldsymbol{v}_1^*, \boldsymbol{\xi}_1) \\ &= \boldsymbol{v}_1^* \cdot \cos(\|\boldsymbol{\xi}_1\|_2) + \frac{\boldsymbol{\xi}_1}{\|\boldsymbol{\xi}_1\|_2} \cdot \sin(\|\boldsymbol{\xi}_1\|_2). \end{aligned} \qquad \text{(ER.9)}$$

Moreover, we claim that $\sin(\|\boldsymbol{\xi}_1\|_2) \neq 0$ when $\boldsymbol{\xi}_1 \neq \mathbf{0}$. Otherwise, if $\sin(\|\boldsymbol{\xi}_1\|_2) = 0$ it implies that $\cos(\|\boldsymbol{\xi}_1\|_2) = \pm 1$ and then $\widetilde{\boldsymbol{v}}_1 = \pm \boldsymbol{v}_1^*$ by (ER.9). When $\widetilde{\boldsymbol{v}}_1 = \boldsymbol{v}_1^*$, we have $\boldsymbol{\xi}_1 = \mathbf{0}$, which is a contradiction. On the other hand, if $\widetilde{\boldsymbol{v}}_1 = -\boldsymbol{v}_1^*$, we have $\|\widetilde{\boldsymbol{v}}_1 - \boldsymbol{v}_1^*\|_2 = \|2\boldsymbol{v}_1^*\|_2 = 2$. Then $\widetilde{\boldsymbol{v}}_1$ is not a consistent estimator of $\boldsymbol{v}_1^*$, which is a contradiction to Definition 2. Thus, we have that $\sin(\|\boldsymbol{\xi}_1\|_2) \neq 0$. Then it follows from (ER.9) that

$$\boldsymbol{\xi}_1 = (\widetilde{\boldsymbol{v}}_1 - \boldsymbol{v}_1^* \cos(\|\boldsymbol{\xi}_1\|_2)) \cdot \frac{\|\boldsymbol{\xi}_1\|_2}{\sin(\|\boldsymbol{\xi}_1\|_2)}. \qquad \text{(ER.10)}$$

Since $\|\boldsymbol{\xi}_1\|_2 / \sin(\|\boldsymbol{\xi}_1\|_2) \neq 0$, we can deduce that

$$\begin{aligned} \|\boldsymbol{\xi}_1\|_0 &= \|\widetilde{\boldsymbol{v}}_1 - \boldsymbol{v}_1^* \cos(\|\boldsymbol{\xi}_1\|_2)\|_0 = \|(\widetilde{\boldsymbol{v}}_1 - \boldsymbol{v}_1^*) + \boldsymbol{v}_1^*(1 - \cos(\|\boldsymbol{\xi}_1\|_2))\|_0 \\ &\le \|\widetilde{\boldsymbol{v}}_1 - \boldsymbol{v}_1^*\|_0 + \|\boldsymbol{v}_1^*(1 - \cos(\|\boldsymbol{\xi}_1\|_2))\|_0 \\ &\le c(r^* + s_u + s_v), \end{aligned} \qquad \text{(ER.11)}$$

where the last inequality above follows from Lemma EC.17 and $\|\boldsymbol{v}_1^*\|_0 = s_v$.

We next derive the upper bound on $\|\boldsymbol{\xi}_1\|_2$. Since $\boldsymbol{\xi}_1 \in T_{\boldsymbol{v}_1^*}\operatorname{St}(1,q)$, from (ER.146) in Section ER.2.2 we see that $\boldsymbol{v}_1^{*T}\boldsymbol{\xi}_1 = 0$. An application of Lemma 3 in Chen and Huang (2012) leads to

$$\|\boldsymbol{\xi}_1\|_2 = O(\|\widetilde{\boldsymbol{v}}_1 - \boldsymbol{v}_1^*\|_2).$$

Together with Lemma EC.17, it yields that

$$\|\boldsymbol{\xi}_1\|_2 \le c\|\widetilde{\boldsymbol{v}}_1 - \boldsymbol{v}_1^*\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}/d_1^*. \tag{ER.12}$$

Further, applying similar arguments to $\boldsymbol{\xi}_2 = \exp_{\boldsymbol{v}_2^*}^{-1}(\widetilde{\boldsymbol{v}}_2)$, we can obtain that

$$\|\exp_{\boldsymbol{v}_2^*}^{-1}(\widetilde{\boldsymbol{v}}_2)\|_0 \le c(r^* + s_u + s_v), \tag{ER.13}$$

$$\|\exp_{\boldsymbol{v}_2^*}^{-1}(\widetilde{\boldsymbol{v}}_2)\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}/d_2^*. \tag{ER.14}$$

(2). The upper bound on $|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{\eta}}_1) - \widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \boldsymbol{\eta}_1^*))|$. By the Taylor expansion of $\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{\eta}}_1)$ in (ER.8), it holds that

$$\begin{aligned}|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{\eta}}_1) - \widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1, \boldsymbol{\eta}_1^*))| &\le |\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(-n^{-1}\mathbf{X}^T\mathbf{Y} + \widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T)(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \\ &\quad + |\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(\boldsymbol{r}_{\boldsymbol{v}_1^*} + \boldsymbol{r}_{\boldsymbol{u}_2^*} + \boldsymbol{r}_{\boldsymbol{v}_2^*})|.\end{aligned} \tag{ER.15}$$

Let us first bound term $|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(-n^{-1}\mathbf{X}^T\mathbf{Y} + \widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T)(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1|$. Notice that $n^{-1}\mathbf{X}^T\mathbf{Y} = \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*\boldsymbol{v}_1^{*T} + \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T} + n^{-1}\mathbf{X}^T\mathbf{E}$. Along with $\boldsymbol{v}_1^{*T}\boldsymbol{v}_1^* = 1$, it gives that

$$\begin{aligned}&(-n^{-1}\mathbf{X}^T\mathbf{Y} + \widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T)(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T}) \\ &= (-\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*\boldsymbol{v}_1^{*T} + \widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T - \boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}) - n^{-1}\mathbf{X}^T\mathbf{E})(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T}) \\ &= (\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T - \boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}) - n^{-1}\mathbf{X}^T\mathbf{E})(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T}).\end{aligned}$$

Denote by $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T - \boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}) - n^{-1}\mathbf{X}^T\mathbf{E}$. It follows that

$$|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(-n^{-1}\mathbf{X}^T\mathbf{Y} + \widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T)(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| = |\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widehat{\boldsymbol{\Delta}}(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1|. \tag{ER.16}$$

Recall that $\widetilde{\mathbf{W}}_1 = \widehat{\boldsymbol{\Theta}}\{\mathbf{I}_p + (\widetilde{z}_{11} - \widetilde{z}_{22})^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{u}}_2^T\}$, where $\widetilde{z}_{11} = \widetilde{\boldsymbol{u}}_1^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_1$ and $\widetilde{z}_{22} = \widetilde{\boldsymbol{u}}_2^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2$. Denote by $\widetilde{\boldsymbol{w}}_i^T = \widehat{\boldsymbol{\theta}}_i^T\{\mathbf{I}_p + (\widetilde{z}_{11} - \widetilde{z}_{22})^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{u}}_2^T\}$ the ith row of $\widetilde{\mathbf{W}}_1$ for $i = 1, \cdots, p$. In light of Lemma EC.19, it holds that

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i\|_0 \le 2\max\{s_{\max}, 3(r^* + s_u + s_v)\} \ \text{ and } \ \max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i\|_2 \le c. \tag{ER.17}$$

Then we have that

$$\begin{aligned}|\widetilde{\boldsymbol{w}}_i^T\widehat{\boldsymbol{\Delta}}(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| &\le |\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \\ &\quad + |\widetilde{\boldsymbol{w}}_i^T\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T - \boldsymbol{u}_2^*\boldsymbol{v}_2^{*T})(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1|.\end{aligned} \tag{ER.18}$$

For the first term on the right-hand side of (ER.18) above, it follows from the sparsity of $\boldsymbol{v}_1^*$, and $\boldsymbol{\xi}_1$ that

$$
\begin{aligned}
&|\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \le |\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{\xi}_1| + |\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*\boldsymbol{v}_1^{*T}\boldsymbol{\xi}_1| \\
&\le \|\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}\|_{2,s}\|\boldsymbol{\xi}_1\|_2 + \|\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}\|_{2,s}\|\boldsymbol{v}_1^*\boldsymbol{v}_1^{*T}\boldsymbol{\xi}_1\|_2 \\
&\le 2\|\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}\|_{2,s}\|\boldsymbol{\xi}_1\|_2, \qquad \text{(ER.19)}
\end{aligned}
$$

where $s = c(r^* + s_u + s_v)$ and the last inequality above is due to $\|\boldsymbol{v}_1^*\boldsymbol{v}_1^{*T}\boldsymbol{\xi}_1\|_2 \le \|\boldsymbol{v}_1^*\|_2|\boldsymbol{v}_1^{*T}\boldsymbol{\xi}_1| \le \|\boldsymbol{\xi}_1\|_2$ for $\|\boldsymbol{v}_1^*\|_2 = 1$. Note that here, for an arbitrary vector $\mathbf{x}$, $\|\mathbf{x}\|_{2,s}^2 = \max_{|S|\le s}\sum_{i\in S} x_i^2$ with S standing for an index set.

From (ER.17) and the fact that $n^{-1}\|\mathbf{X}^T\mathbf{E}\|_{\max} \le c\{n^{-1}\log(pq)\}^{1/2}$, it holds that

$$
\begin{aligned}
\|\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}\|_{\max} &\le \|\widetilde{\boldsymbol{w}}_i\|_1\|n^{-1}\mathbf{X}^T\mathbf{E}\|_{\max} \\
&\le c\max\{s_{\max}, (r^* + s_u + s_v)\}^{1/2}\{n^{-1}\log(pq)\}^{1/2}.
\end{aligned}
$$

Then it follows that

$$
\|\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}\|_{2,s} \le c\max\{s_{\max}, (r^* + s_u + s_v)\}^{1/2}(r^* + s_u + s_v)^{1/2}\{n^{-1}\log(pq)\}^{1/2},
$$

which together with (ER.12) and (ER.19) entails that

$$
\begin{aligned}
&|\widetilde{\boldsymbol{w}}_i^T n^{-1}\mathbf{X}^T\mathbf{E}(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \\
&\le c\max\{s_{\max}, (r^* + s_u + s_v)\}^{1/2}(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}/d_1^*. \qquad \text{(ER.20)}
\end{aligned}
$$

We next bound term $|\widetilde{\boldsymbol{w}}_i^T\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T - \boldsymbol{u}_2^*\boldsymbol{v}_2^{*T})(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1|$ on the right-hand side of (ER.18) above. Observe that

$$
\begin{aligned}
\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T - \boldsymbol{u}_2^*\boldsymbol{v}_2^{*T})\|_2 &\le \|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*)\boldsymbol{v}_2^{*T}\|_2 + \|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2(\widetilde{\boldsymbol{v}}_2 - \boldsymbol{v}_2^*)^T\|_2 \\
&\le \|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*)\|_2\|\boldsymbol{v}_2^*\|_2 + \|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\|_2\|\widetilde{\boldsymbol{v}}_2 - \boldsymbol{v}_2^*\|_2 \\
&\le c(r^* + s_u + s_v)^{1/2}\eta_n^2\left\{n^{-1}\log(pq)\right\}^{1/2},
\end{aligned}
$$

where the last inequality above uses $\|\boldsymbol{v}_1^*\|_2 = 1$ and Lemma EC.17. It follows from (ER.12) and $\|\boldsymbol{v}_1^*\|_2 = 1$ that

$$
\begin{aligned}
\|(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1\|_2 &\le \|\boldsymbol{\xi}_1\|_2 + \|\boldsymbol{v}_1^*\boldsymbol{v}_1^{*T}\boldsymbol{\xi}_1\|_2 \le 2\|\boldsymbol{\xi}_1\|_2 \\
&\le c(r^* + s_u + s_v)^{1/2}\eta_n^2\left\{n^{-1}\log(pq)\right\}^{1/2}/d_1^*.
\end{aligned}
$$

Together with (ER.17), it holds that

$$
\begin{aligned}
&|\widetilde{\boldsymbol{w}}_i^T\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T - \boldsymbol{u}_2^*\boldsymbol{v}_2^{*T})(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \qquad \text{(ER.21)} \\
&\le \|\widetilde{\boldsymbol{w}}_i\|_2\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T - \boldsymbol{u}_2^*\boldsymbol{v}_2^{*T})\|_2\|(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1\|_2 \\
&\le c(r^* + s_u + s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\}/d_1^*. \qquad \text{(ER.22)}
\end{aligned}
$$

Combining (ER.18), (ER.20), and (ER.21), we can obtain that

$$|\widetilde{\boldsymbol{w}}_i^T \widehat{\boldsymbol{\Delta}}(\mathbf{I}_q - \boldsymbol{v}_1^* \boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \le c \max\{s_{\max}^{1/2}, (r^* + s_u + s_v)^{1/2}, \eta_n^2\}(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}/d_1^*.$$

Applying (ER.17) again results in

$$\max_{1\le i\le p} |\widetilde{\boldsymbol{w}}_i^T \widehat{\boldsymbol{\Delta}}(\mathbf{I}_q - \boldsymbol{v}_1^* \boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \le c \max\{s_{\max}^{1/2}, (r^* + s_u + s_v)^{1/2}, \eta_n^2\}(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}/d_1^*.$$

Thus, for each vector $\boldsymbol{a} \in \mathbb{R}^p$ we have that

$$\begin{aligned}
&|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \widehat{\boldsymbol{\Delta}}(\mathbf{I}_q - \boldsymbol{v}_1^* \boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \le \|\boldsymbol{a}\|_1 \|\widetilde{\mathbf{W}}_1 \widehat{\boldsymbol{\Delta}}(\mathbf{I}_q - \boldsymbol{v}_1^* \boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1\|_{\max} \\
&\le \|\boldsymbol{a}\|_0^{1/2} \|\boldsymbol{a}\|_2 \max_{1\le i\le p} |\widetilde{\boldsymbol{w}}_i^T \widehat{\boldsymbol{\Delta}}(\mathbf{I}_q - \boldsymbol{v}_1^* \boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1| \\
&\le c\|\boldsymbol{a}\|_0^{1/2} \|\boldsymbol{a}\|_2 \max\{s_{\max}^{1/2}, (r^* + s_u + s_v)^{1/2}, \eta_n^2\}(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}/d_1^*. \qquad \text{(ER.23)}
\end{aligned}$$

It remains to bound term $|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 (\boldsymbol{r}_{\boldsymbol{v}_1^*} + \boldsymbol{r}_{\boldsymbol{u}_2^*} + \boldsymbol{r}_{\boldsymbol{v}_2^*})|$ above. Let us recall that the Taylor remainder terms $\boldsymbol{r}_{\boldsymbol{v}_1^*}, \boldsymbol{r}_{\boldsymbol{u}_2^*}$, and $\boldsymbol{r}_{\boldsymbol{v}_2^*}$ satisfy that

$$\|\boldsymbol{r}_{\boldsymbol{v}_1^*}\|_2 = O(\|\exp_{\boldsymbol{v}_1^*}^{-1}(\widetilde{\boldsymbol{v}}_1)\|_2^2), \ \|\boldsymbol{r}_{\boldsymbol{u}_2^*}\|_2 = O(\|\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*\|_2^2), \ \|\boldsymbol{r}_{\boldsymbol{v}_2^*}\|_2 = O(\|\exp_{\boldsymbol{v}_2^*}^{-1}(\widetilde{\boldsymbol{v}}_2)\|_2^2).$$

Based on Lemma EC.19 that $\|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1\|_2 \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2$, from (ER.12) we have that

$$\begin{aligned}
&|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \boldsymbol{r}_{\boldsymbol{v}_1^*}| \le \|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1\|_2 \|\boldsymbol{r}_{\boldsymbol{v}_1^*}\|_2 \\
&\le c\|\boldsymbol{a}\|_0^{1/2} \|\boldsymbol{a}\|_2 (r^* + s_u + s_v)\eta_n^4\{n^{-1}\log(pq)\}/d_1^{*2}.
\end{aligned}$$

Then we apply similar arguments to $|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \boldsymbol{r}_{\boldsymbol{u}_2^*}|$ and $|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \boldsymbol{r}_{\boldsymbol{v}_2^*}|$. In view of Definition 2 and (ER.14), it holds that

$$\begin{aligned}
&\|\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*\|_2^2 \le c(r^* + s_u + s_v)\eta_n^4\{n^{-1}\log(pq)\}, \\
&\|\exp_{\boldsymbol{v}_2^*}^{-1}(\widetilde{\boldsymbol{v}}_2)\|_2^2 \le c(r^* + s_u + s_v)\eta_n^4\{n^{-1}\log(pq)\}/d_2^{*2}.
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
&|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \boldsymbol{r}_{\boldsymbol{u}_2^*}| \le \|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1\|_2 \|\boldsymbol{r}_{\boldsymbol{u}_2^*}\|_2 \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 (r^* + s_u + s_v)\eta_n^4\{n^{-1}\log(pq)\}, \\
&|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \boldsymbol{r}_{\boldsymbol{v}_2^*}| \le \|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1\|_2 \|\boldsymbol{r}_{\boldsymbol{v}_2^*}\|_2 \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 (r^* + s_u + s_v)\eta_n^4\{n^{-1}\log(pq)\}/d_2^{*2}.
\end{aligned}$$

Since the nonzero eigenvalues d_i^{*2} are at the constant level by Condition 4, it follows that

$$\begin{aligned}
|\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 (\boldsymbol{r}_{\boldsymbol{v}_1^*} + \boldsymbol{r}_{\boldsymbol{u}_2^*} + \boldsymbol{r}_{\boldsymbol{v}_2^*})| &\le |\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \boldsymbol{r}_{\boldsymbol{v}_1^*}| + |\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \boldsymbol{r}_{\boldsymbol{u}_2^*}| + |\boldsymbol{a}^T \widetilde{\mathbf{W}}_1 \boldsymbol{r}_{\boldsymbol{v}_2^*}| \\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 (r^* + s_u + s_v)\eta_n^4\{n^{-1}\log(pq)\}. \qquad \text{(ER.24)}
\end{aligned}$$

Combining (ER.15), (ER.16), (ER.23), and (ER.24) yields that

$$
\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1,\widetilde{\boldsymbol{\eta}}_1)-\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1,\boldsymbol{\eta}_1^*))| \\
&\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2\max\{s_{\max}^{1/2},(r^*+s_u+s_v)^{1/2},\eta_n^2\}(r^*+s_u+s_v)\eta_n^2\{n^{-1}\log(pq)\}. \qquad \text{(ER.25)}
\end{aligned}
$$

Thus, for $\boldsymbol{a}\in\mathcal{A}=\{\boldsymbol{a}\in\mathbb{R}^p:\|\boldsymbol{a}\|_0\leq m,\|\boldsymbol{a}\|_2=1\}$, we can obtain that

$$
\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1,\widetilde{\boldsymbol{\eta}}_1)-\widetilde{\psi}_1(\widetilde{\boldsymbol{u}}_1,\boldsymbol{\eta}_1^*))| \\
&\leq cm^{1/2}\max\{s_{\max}^{1/2},(r^*+s_u+s_v)^{1/2},\eta_n^2\}(r^*+s_u+s_v)\eta_n^2\{n^{-1}\log(pq)\}, \qquad \text{(ER.26)}
\end{aligned}
$$

which completes the proof of Lemma 2.

ER.1.3. Proof of Lemma 3

Notice that

$$
\begin{aligned}
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\boldsymbol{\delta}}_1| &= |\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{z}_{11}^{-1}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T\left\{(\widetilde{\boldsymbol{v}}_1-\boldsymbol{v}_1^*)\widetilde{\boldsymbol{u}}_1^T-(\widetilde{\boldsymbol{v}}_2\widetilde{\boldsymbol{u}}_2^T-\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T})\right\}\widehat{\mathbf{\Sigma}}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)| \\
&\leq |\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{z}_{11}^{-1}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T(\widetilde{\boldsymbol{v}}_1-\boldsymbol{v}_1^*)||\widetilde{\boldsymbol{u}}_1^T\widehat{\mathbf{\Sigma}}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)| \\
&\quad + |\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{z}_{11}^{-1}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T(\widetilde{\boldsymbol{v}}_2\widetilde{\boldsymbol{u}}_2^T-\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T})\widehat{\mathbf{\Sigma}}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)|.
\end{aligned}
$$

We aim to bound the two terms introduced above under Condition 4 that the nonzero eigenvalues d_i^{*2} are at the constant level. For the first term above, it follows from Conditions 2–4 that

$$
\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{z}_{11}^{-1}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T(\widetilde{\boldsymbol{v}}_1-\boldsymbol{v}_1^*)||\widetilde{\boldsymbol{u}}_1^T\widehat{\mathbf{\Sigma}}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)| \\
&\leq |\widetilde{z}_{11}^{-1}|\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2\|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\|_2\|\widetilde{\boldsymbol{v}}_2\|_2\|\widetilde{\boldsymbol{v}}_1-\boldsymbol{v}_1^*\|_2\|\widetilde{\boldsymbol{u}}_1\|_2\|\widehat{\mathbf{\Sigma}}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)\|_2 \\
&\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\}, \qquad \text{(ER.27)}
\end{aligned}
$$

where we have used Definition 2 with $\|\widetilde{\boldsymbol{v}}_2\|_2=1$, $\|\widetilde{\boldsymbol{u}}_2\|_2\leq c$, parts (a)–(c) of Lemma EC.17, and part (c) of Lemma EC.19.

For the second term above, let us first bound $\|\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T-\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2$. In light of part (a) of Definition 2 and part (a) of Lemma EC.17, we can deduce that

$$
\begin{aligned}
\|\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T-\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2 &\leq \|(\widetilde{\boldsymbol{u}}_2-\boldsymbol{u}_2^*)\boldsymbol{v}_2^{*T}\|_2+\|\widetilde{\boldsymbol{u}}_2(\widetilde{\boldsymbol{v}}_2-\boldsymbol{v}_2^*)^T\|_2 \\
&\leq \|\widetilde{\boldsymbol{u}}_2-\boldsymbol{u}_2^*\|_2\|\boldsymbol{v}_2^*\|_2+\|\widetilde{\boldsymbol{u}}_2\|_2\|\widetilde{\boldsymbol{v}}_2-\boldsymbol{v}_2^*\|_2 \\
&\leq c(r^*+s_u+s_v)^{1/2}\eta_n^2\left\{n^{-1}\log(pq)\right\}^{1/2}.
\end{aligned}
$$

Then similar to (ER.27), it holds that

$$
\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{z}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T(\widetilde{\boldsymbol{v}}_2\widetilde{\boldsymbol{u}}_2^T-\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T})\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)|\\
&\le|\widetilde{z}_{11}^{-1}|\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2\|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\|_2\|\widetilde{\boldsymbol{v}}_2\|_2\|\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T-\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)\|_2\\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\}.
\end{aligned}
$$

Thus, combining the above results yields that

$$
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\boldsymbol{\delta}}_1|\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\},
$$

which completes the proof of Lemma 3.

ER.1.4. Proof of Lemma 4

Recall that $\widetilde{\mathbf{M}}_1=-\widetilde{z}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T$. Under Condition 4 that d_i^{*2} are at the constant level, parts (b) and (c) of Lemma EC.17 show that $\|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\|_2\le c$ and $|\widetilde{z}_{11}^{-1}|\le c$. Since $\|\widetilde{\boldsymbol{v}}_2\|_2=1$ due to Definition 2, we can obtain that

$$
\|\widetilde{\mathbf{M}}_1\|_2\le\|\widetilde{z}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T\|_2\le|\widetilde{z}_{11}^{-1}|\|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\|_2\|\widetilde{\boldsymbol{v}}_2\|_2\le c. \tag{ER.28}
$$

It further holds that

$$
\begin{aligned}
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*|&\le\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2\|\widetilde{\mathbf{M}}_1\|_2\|\boldsymbol{v}_2^*\|_2|\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*|\\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2|\boldsymbol{l}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_1^*|,
\end{aligned}
\tag{ER.29}
$$

where we have used $\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2$ in Lemma EC.19, $\|\boldsymbol{v}_2^*\|_2=1$, and $|\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*|\le c|\boldsymbol{l}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_1^*|$. Therefore, under Condition 4 we have that

$$
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1^*|=o(\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2n^{-1/2}),
$$

which concludes the proof of Lemma 4.

ER.1.5. Proof of Lemma 5

Observe that

$$
\begin{aligned}
&|-\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\boldsymbol{\epsilon}}_1-h_1/\sqrt{n}|\\
&\le n^{-1}|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1\mathbf{E}^T\mathbf{X}\widetilde{\boldsymbol{u}}_1-\boldsymbol{a}^T\mathbf{W}_1^*\mathbf{M}_1^*\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1^*|+n^{-1}|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1-\mathbf{W}_1^*)\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*|\\
&\le n^{-1}|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)|+n^{-1}|(\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1-\boldsymbol{a}^T\mathbf{W}_1^*\mathbf{M}_1^*)\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1^*|\\
&\quad+n^{-1}|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1-\mathbf{W}_1^*)\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*|.
\end{aligned}
\tag{ER.30}
$$

We aim to bound the three terms introduced above separately. Let us first show that $\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1$, $\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1 - \boldsymbol{a}^T\mathbf{W}_1^*\mathbf{M}_1^*$, and $\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1 - \mathbf{W}_1^*)$ are all s-sparse with $s = c(r^* + s_u + s_v)$. Recall that $\widetilde{\mathbf{M}}_1 = -\widetilde{z}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T$ and $\mathbf{M}_1^* = -z_{11}^{*-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}$. It follows from part (b) of Lemma EC.17 and $\|\boldsymbol{v}_2^*\|_0 \le s_v$ that

$$\begin{aligned}\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1\|_0 &= \|(\widetilde{z}_{11}^{-1}\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{l}}_2)\cdot\widetilde{d}_2\widetilde{\boldsymbol{v}}_2^T\|_0\\ &\le \|\widetilde{d}_2\widetilde{\boldsymbol{v}}_2\|_0 \le \|\boldsymbol{v}_2^*\|_0 + \|\widetilde{d}_2(\widetilde{\boldsymbol{v}}_2 - \boldsymbol{v}_2^*)\|_0\\ &\le c(r^* + s_u + s_v) \end{aligned} \tag{ER.31}$$

and

$$\begin{aligned}\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1 - \boldsymbol{a}^T\mathbf{W}_1^*\mathbf{M}_1^*\|_0 &\le \|(\widetilde{z}_{11}^{-1}\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{l}}_2)\cdot\widetilde{d}_2\widetilde{\boldsymbol{v}}_2^T\|_0 + \|(z_{11}^{*-1}\boldsymbol{a}^T\mathbf{W}_1^*\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*)\cdot\boldsymbol{v}_2^{*T}\|_0\\ &\le \|\widetilde{d}_2\widetilde{\boldsymbol{v}}_2\|_0 + \|\boldsymbol{v}_2^*\|_0 \le c(r^* + s_u + s_v). \end{aligned} \tag{ER.32}$$

Further, from the definitions of $\widetilde{\mathbf{W}}_1$ and $\mathbf{W}_1^*$, we have that

$$\begin{aligned}\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1 - \mathbf{W}_1^*)\|_0 &= \|(\boldsymbol{a}^T\widehat{\boldsymbol{\Theta}}(\widetilde{z}_{11} - \widetilde{z}_{22})^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2)\cdot\widetilde{\boldsymbol{u}}_2^T - (\boldsymbol{a}^T\widehat{\boldsymbol{\Theta}}(z_{11}^* - z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*)\cdot\boldsymbol{u}_2^{*T}\|_0\\ &\le \|\widetilde{\boldsymbol{u}}_2^T\|_0 + \|\boldsymbol{u}_2^{*T}\|_0 \le 2\|\boldsymbol{u}_2^*\|_0 + \|\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*\|_0\\ &\le c(r^* + s_u + s_v), \end{aligned} \tag{ER.33}$$

where the last step above is due to Definition 2 and $\|\boldsymbol{u}_2^*\|_0 \le s_u$. Hence, combining (ER.30)–(ER.33) leads to

$$\begin{aligned}|-\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\boldsymbol{\epsilon}}_1 - h_1/\sqrt{n}| &\le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2\|\widetilde{\mathbf{M}}_1\|_2\|n^{-1}\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_1 - \boldsymbol{u}_1^*)\|_{2,s}\\ &\quad + \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1 - \boldsymbol{a}^T\mathbf{W}_1^*\mathbf{M}_1^*\|_2\|n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1^*\|_{2,s}\\ &\quad + \|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1 - \mathbf{W}_1^*)\|_2\|n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*\|_{2,s}\\ &=: A_1 + A_2 + A_3. \end{aligned} \tag{ER.34}$$

We will provide the upper bounds for the three terms A_1, A_2, and A_3 introduced in (ER.34) above separately.

We start with bounding $n^{-1}\|\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_1 - \boldsymbol{u}_1^*)\|_{2,s}$, $n^{-1}\|\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1^*\|_{2,s}$, and $n^{-1}\|\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*\|_{2,s}$. From $n^{-1}\|\mathbf{X}^T\mathbf{E}\|_{\max} \le c\{n^{-1}\log(pq)\}^{1/2}$ and Definition 2, we can deduce that

$$\begin{aligned}n^{-1}\|\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_1 - \boldsymbol{u}_1^*)\|_{\max} &\le n^{-1}\|\mathbf{E}^T\mathbf{X}\|_{\max}\|\widetilde{\boldsymbol{u}}_1 - \boldsymbol{u}_1^*\|_1\\ &\le n^{-1}\|\mathbf{E}^T\mathbf{X}\|_{\max}\|\widetilde{\boldsymbol{u}}_1 - \boldsymbol{u}_1^*\|_0^{1/2}\|\widetilde{\boldsymbol{u}}_1 - \boldsymbol{u}_1^*\|_2 \le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\},\\ n^{-1}\|\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1^*\|_{\max} &\le n^{-1}\|\mathbf{E}^T\mathbf{X}\|_{\max}\|\boldsymbol{u}_1^*\|_0^{1/2}\|\boldsymbol{u}_1^*\|_2 \le cs_u^{1/2}\{n^{-1}\log(pq)\}^{1/2}d_1^*,\\ n^{-1}\|\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*\|_{\max} &\le n^{-1}\|\mathbf{X}^T\mathbf{E}\|_{\max}\|\boldsymbol{v}_1^*\|_0^{1/2}\|\boldsymbol{v}_1^*\|_2 \le cs_v^{1/2}\{n^{-1}\log(pq)\}^{1/2}.\end{aligned}$$

Then it follows that

$$n^{-1}\|\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)\|_{2,s}\le s^{1/2}n^{-1}\|\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)\|_{\max}\le cs^{3/2}\eta_n^2\{n^{-1}\log(pq)\}, \tag{ER.35}$$

$$n^{-1}\|\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1^*\|_{2,s}\le cs^{1/2}s_u^{1/2}\{n^{-1}\log(pq)\}^{1/2}d_1^*, \tag{ER.36}$$

$$n^{-1}\|\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*\|_{2,s}\le cs^{1/2}s_v^{1/2}\{n^{-1}\log(pq)\}^{1/2}. \tag{ER.37}$$

Using part (c) of Lemma EC.19, $\|\boldsymbol{a}\|_0=m$, and $\|\boldsymbol{a}\|_2=1$, we can show that

$$\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2\le cm^{1/2}, \tag{ER.38}$$

$$\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1-\mathbf{W}_1^*)\|_2\le cm^{1/2}(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}. \tag{ER.39}$$

Further, for $\widetilde{\mathbf{M}}_1=-\widetilde{z}_{11}^{-1}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T$ and $\mathbf{M}_1^*=-z_{11}^{*-1}\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}$, it holds that

$$\|\widetilde{\mathbf{M}}_1\|_2\le\|\widetilde{z}_{11}^{-1}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T\|_2\le|\widetilde{z}_{11}^{-1}|\|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\|_2\|\widetilde{\boldsymbol{v}}_2\|_2\le cd_1^{*-2}d_2^*, \tag{ER.40}$$

$$\|\mathbf{M}_1^*\|_2\le\|z_{11}^{*-1}\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2\le|z_{11}^{*-1}|\|\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\|_2\|\boldsymbol{v}_2^*\|_2\le cd_1^{*-2}d_2^*, \tag{ER.41}$$

where we have used the results in parts (c) and (d) of Lemma EC.17. Hence, combining (ER.35), (ER.38), and (ER.40), for term A_1 above we can obtain that

$$\begin{aligned}A_1&=\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2\|\widetilde{\mathbf{M}}_1\|_2\|n^{-1}\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_1-\boldsymbol{u}_1^*)\|_{2,s}\\&\le cm^{1/2}(r^*+s_u+s_v)^{3/2}\eta_n^2\{n^{-1}\log(pq)\}d_2^*d_1^{*-2}.\end{aligned} \tag{ER.42}$$

With the aid of (ER.37) and (ER.39), it also holds that

$$\begin{aligned}A_3&=\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1-\mathbf{W}_1^*)\|_2\|n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*\|_{2,s}\\&\le cm^{1/2}(r^*+s_u+s_v)^{3/2}\eta_n^2\{n^{-1}\log(pq)\}.\end{aligned} \tag{ER.43}$$

It remains to bound term A_2 above. From Lemma EC.17 and $\|\boldsymbol{v}_2^*\|_2=1$, we see that

$$\|\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2\le\|\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\|_2\|\boldsymbol{v}_2^{*T}\|_2\le cd_2^*,$$

$$\begin{aligned}\|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T-\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2&\le\|\widehat{\mathbf{\Sigma}}(\widetilde{\boldsymbol{u}}_2-\boldsymbol{u}_2^*)\|_2\|\boldsymbol{v}_2^{*T}\|_2+\|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\|_2\|(\widetilde{\boldsymbol{v}}_2-\boldsymbol{v}_2^*)^T\|_2\\&\le c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned}$$

Together with the upper bounds for $|\widetilde{z}_{11}^{-1}|$ and $|\widetilde{z}_{11}^{-1}-z_{11}^{*-1}|$ in Lemma EC.17, it holds that

$$\begin{aligned}\|\widetilde{\mathbf{M}}_1-\mathbf{M}_1^*\|_2&=\|\widetilde{z}_{11}^{-1}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T-z_{11}^{*-1}\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2\\&\le|\widetilde{z}_{11}^{-1}|\|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{v}}_2^T-\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2+|\widetilde{z}_{11}^{-1}-z_{11}^{*-1}|\|\widehat{\mathbf{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{v}_2^{*T}\|_2\\&\le c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_1^{*-2},\end{aligned} \tag{ER.44}$$

Then a combination of (ER.38), (ER.39), (ER.41), and (ER.44) results in

$$\begin{aligned}
&\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1 - \boldsymbol{a}^T\mathbf{W}_1^*\mathbf{M}_1^*\|_2 \le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1(\widetilde{\mathbf{M}}_1 - \mathbf{M}_1^*)\|_2 + \|(\boldsymbol{a}^T\widetilde{\mathbf{W}}_1 - \boldsymbol{a}^T\mathbf{W}_1^*)\mathbf{M}_1^*\|_2 \\
&\le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2\|\widetilde{\mathbf{M}}_1 - \mathbf{M}_1^*\|_2 + \|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1 - \mathbf{W}_1^*)\|_2\|\mathbf{M}_1^*\|_2 \\
&\le cm^{1/2}(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_1^{*-2}. \qquad \text{(ER.45)}
\end{aligned}$$

With the aid of (ER.36) and (ER.45), we can deduce that

$$\begin{aligned}
A_2 &= \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\mathbf{M}}_1 - \boldsymbol{a}^T\mathbf{W}_1^*\mathbf{M}_1^*\|_2\|n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1^*\|_{2,s} \\
&\le cm^{1/2}(r^* + s_u + s_v)^{3/2}\eta_n^2\{n^{-1}\log(pq)\}d_1^{*-1}. \qquad \text{(ER.46)}
\end{aligned}$$

Therefore, combining (ER.34), (ER.42), (ER.46), and (ER.43) yields that

$$|-\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\boldsymbol{\epsilon}}_1 - h_1/\sqrt{n}| \le cm^{1/2}(r^* + s_u + s_v)^{3/2}\eta_n^2\{n^{-1}\log(pq)\}d_1^{*-1}. \qquad \text{(ER.47)}$$

Furthermore, under Condition 4 that d_1^* is at the constant level, we have that

$$|-\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\widetilde{\boldsymbol{\epsilon}}_1 - h_1/\sqrt{n}| \le cm^{1/2}(r^* + s_u + s_v)^{3/2}\eta_n^2\{n^{-1}\log(pq)\},$$

which completes the proof of Lemma 5.

ER.1.6. Proof of Lemma 6

Observe that

$$\begin{aligned}
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_k| &= |\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\left\{(\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*)\widetilde{\boldsymbol{u}}_k^T - (\widetilde{\mathbf{C}}_{-k}^T - \mathbf{C}_{-k}^{*T})\right\}\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)| \\
&\le |\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*)||\widetilde{\boldsymbol{u}}_k^T\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)| \\
&\quad + |\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widetilde{\mathbf{C}}_{-k}^T - \mathbf{C}_{-k}^{*T})\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)|.
\end{aligned}$$

From Lemma EC.21, we have that $\|\widetilde{\mathbf{M}}_k\|_2 \le cd_k^{*-2}d_1^*$. It follows from Lemma EC.22 that $\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2 \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2$. Together with parts (a) and (b) of Lemma EC.17, it holds that

$$\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*)||\widetilde{\boldsymbol{u}}_k^T\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)| \\
&\le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\|_2\|\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*\|_2\|\widetilde{\boldsymbol{u}}_k\|_2\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)\|_2 \\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^* + s_u + s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\}d_k^{*-2}d_1^*.
\end{aligned}$$

Further, by Lemma EC.21 we can obtain that

$$\|\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}^*\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2\left\{n^{-1}\log(pq)\right\}^{1/2}.$$

Then it follows that

$$
\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widetilde{\mathbf{C}}_{-k}^T-\mathbf{C}_{-k}^{*T})\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k-\boldsymbol{u}_k^*)|\\
&\leq \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\|_2\|\widetilde{\mathbf{C}}_{-k}^T-\mathbf{C}_{-k}^{*T}\|_2\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k-\boldsymbol{u}_k^*)\|_2\\
&\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\}d_k^{*-2}d_1^*.
\end{aligned}
$$

Combining the above results leads to

$$
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_k|\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\}d_k^{*-2}d_1^*.
$$

Thus, under Condition 4 that the nonzero eigenvalues d_i^{*2} are at the constant level, we can deduce that

$$
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_k|\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\},
$$

which completes the proof of Lemma 6.

ER.1.7. Proof of Lemma 7

According to the construction that $\widetilde{\mathbf{M}}_k=-\widetilde{z}_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{C}}_{-k}$, it holds that

$$
\begin{aligned}
&\widetilde{\mathbf{M}}_k\mathbf{C}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*=\sum_{j\neq k}\widetilde{\mathbf{M}}_k\boldsymbol{v}_j^*\boldsymbol{u}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*=-\sum_{j\neq k}\widetilde{z}_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\sum_{i\neq k}\widetilde{\boldsymbol{u}}_i\widetilde{\boldsymbol{v}}_i^T\boldsymbol{v}_j^*\boldsymbol{u}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*\\
&=-\sum_{j\neq k}\widetilde{z}_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\sum_{i\neq k}\widetilde{\boldsymbol{u}}_i\boldsymbol{v}_i^{*T}\boldsymbol{v}_j^*\boldsymbol{u}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*-\sum_{j\neq k}\widetilde{z}_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\sum_{i\neq k}\widetilde{\boldsymbol{u}}_i(\widetilde{\boldsymbol{v}}_i^T-\boldsymbol{v}_i^{*T})\boldsymbol{v}_j^*\boldsymbol{u}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*\\
&=-\sum_{j\neq k}\widetilde{z}_{kk}^{-1}(\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_j+\sum_{i\neq k}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_i(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)^T\boldsymbol{v}_j^*)\boldsymbol{u}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*,
\end{aligned}
\tag{ER.48}
$$

where the last step above has used $\boldsymbol{v}_i^{*T}\boldsymbol{v}_j^*=0$ for each $i\neq j$. For term $\sum_{i\neq k}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_i(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)^T\boldsymbol{v}_j^*$ above, we can deduce that

$$
\begin{aligned}
\|\sum_{i\neq k}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_i(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)^T\boldsymbol{v}_j^*\|_2&\leq\sum_{i\neq k}\|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{l}}_i\|_2\|\widetilde{d}_i(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\|_2\|\boldsymbol{v}_j^*\|_2\\
&\leq c\sum_{i\neq k}\|\widetilde{d}_i(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\|_2\leq cr^*\gamma_n,
\end{aligned}
$$

where we have used $\|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{l}}_i\|_2\leq c$ due to Condition 2, $\|\boldsymbol{v}_j^*\|_2=1$, and parts (a) and (b) of Lemma EC.17. Recall the condition that $m^{1/2}\kappa_n=o(1)$ with

$$
\kappa_n=\max\{s_{\max}^{1/2},(r^*+s_u+s_v)^{1/2},\eta_n^2\}(r^*+s_u+s_v)\eta_n^2\log(pq)/\sqrt{n}.
$$

Then an application of similar arguments as for (ER.93) leads to $r^*\gamma_n=o(1)$.

Part (b) of Lemma EC.17 shows that $\|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_j\|_2 \le cd_j^*$. Since the nonzero eigenvalues d_i^{*2} are at the constant level by Condition 4, for sufficiently large n we have that

$$\begin{aligned}\|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_j + \sum_{i\neq k}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_i(\widetilde{\boldsymbol{v}}_i - \boldsymbol{v}_i^*)^T\boldsymbol{v}_j^*\|_2 &\le \|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_j\|_2 + \|\sum_{i\neq k}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_i(\widetilde{\boldsymbol{v}}_i - \boldsymbol{v}_i^*)^T\boldsymbol{v}_j^*\|_2 \\ &\le cd_j^*.\end{aligned}$$

Together with $|\widetilde{z}_{kk}^{-1}| \le cd_k^{*-2}$ in part (c) of Lemma EC.17, it follows that

$$\begin{aligned}\|\widetilde{\mathbf{M}}_k\mathbf{C}_{-k}^{*T}\widehat{\mathbf{\Sigma}}\boldsymbol{u}_k^*\|_2 &\le \sum_{j\neq k}|\widetilde{z}_{kk}^{-1}|\big\|\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_j + \sum_{i\neq k}\widehat{\mathbf{\Sigma}}\widetilde{\boldsymbol{u}}_i(\widetilde{\boldsymbol{v}}_i - \boldsymbol{v}_i^*)^T\boldsymbol{v}_j^*\big\|_2|\boldsymbol{u}_j^{*T}\widehat{\mathbf{\Sigma}}\boldsymbol{u}_k^*| \\ &\le c\sum_{j\neq k}(d_j^*/d_k^{*2})|\boldsymbol{u}_j^{*T}\widehat{\mathbf{\Sigma}}\boldsymbol{u}_k^*| = c\sum_{j\neq k}(d_j^{*2}/d_k^*)|\boldsymbol{l}_j^{*T}\widehat{\mathbf{\Sigma}}\boldsymbol{l}_k^*|. \qquad \text{(ER.49)}\end{aligned}$$

Using Condition 4 that $\sum_{i\neq j}|\boldsymbol{l}_i^{*T}\widehat{\mathbf{\Sigma}}\boldsymbol{l}_j^*| = o(n^{-1/2})$ and the nonzero eigenvalues d_i^{*2} are at the constant level, and Lemma EC.22 that $\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2 \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2$, we can obtain that

$$\begin{aligned}|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\mathbf{C}_{-k}^{*T}\widehat{\mathbf{\Sigma}}\boldsymbol{u}_k^*| &\le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\mathbf{C}_{-k}^{*T}\widehat{\mathbf{\Sigma}}\boldsymbol{u}_k^*\|_2 \\ &= o(\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 n^{-1/2}).\end{aligned}$$

This concludes the proof of Lemma 7.

ER.1.8. Proof of Lemma 8

We proof the results using similar arguments as in the proof of Lemma 2. For the nuisance parameter $\boldsymbol{\eta}_k = [\boldsymbol{v}_1^T, \cdots, \boldsymbol{v}_{r^*}^T, \boldsymbol{u}_1^T, \cdots, \boldsymbol{u}_{k-1}^T, \boldsymbol{u}_{k+1}^T, \cdots, \boldsymbol{u}_{r^*}^T,]^T$, it follows from the definition of $\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)$ that

$$\begin{aligned}\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k) &= \frac{\partial L}{\partial \boldsymbol{u}_k} - \mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_k} \\ &= \frac{\partial L}{\partial \boldsymbol{u}_k} - \left(\mathbf{M}_k^v\frac{\partial L}{\partial \boldsymbol{v}_k} + \sum_{j\neq k}\mathbf{M}_j^u\frac{\partial L}{\partial \boldsymbol{u}_j} + \sum_{j\neq k}\mathbf{M}_j^v\frac{\partial L}{\partial \boldsymbol{v}_j}\right). \qquad \text{(ER.50)}\end{aligned}$$

By Proposition 2, we see that $\mathbf{M}_j^u = \mathbf{0}$ and $\mathbf{M}_j^v = \mathbf{0}$ for $j \in \{1, \cdots, r^*\}$ with $j \neq k$, which means that we need only to consider $\boldsymbol{v}_k$ as the nuisance parameter. In light of the derivatives (ER.71) and (ER.72), we can deduce that

$$\begin{aligned}\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k) &= \frac{\partial L}{\partial \boldsymbol{u}_k} - \mathbf{M}_k^v\frac{\partial L}{\partial \boldsymbol{v}_k} \\ &= \widehat{\mathbf{\Sigma}}\boldsymbol{u}_k - n^{-1}\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_k - \mathbf{M}_k^v(\boldsymbol{v}_k\boldsymbol{u}_k^T\widehat{\mathbf{\Sigma}}\boldsymbol{u}_k - n^{-1}\mathbf{Y}^T\mathbf{X}\boldsymbol{u}_k). \qquad \text{(ER.51)}\end{aligned}$$

For arbitrary fixed $\mathbf{M}_k^v$, we can see that $\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)$ is only a function of $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$, which means that we need only to do the Taylor expansion of $\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)$ with respect to $\boldsymbol{v}_k$. Similar to the proof of Lemma EC.16, we can obtain the Taylor expansion of $\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)$

$$\widetilde{\psi}_k\left(\boldsymbol{u}_k, \boldsymbol{\eta}_k\right) = \widetilde{\psi}_k\left(\boldsymbol{u}_k, \boldsymbol{\eta}_k^*\right) + \frac{\partial\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)}{\partial \boldsymbol{v}_k^T}\Big|_{\boldsymbol{v}_k^*}(\mathbf{I}_q - \boldsymbol{v}_k^*\boldsymbol{v}_k^{*T})\exp_{\boldsymbol{v}_k^*}^{-1}(\boldsymbol{v}_k) + \boldsymbol{r}_{\boldsymbol{v}_k^*},$$

where the Taylor remainder term satisfies that

$$\|\boldsymbol{r}_{\boldsymbol{v}_k^*}\|_2 = O(\|\exp^{-1}_{\boldsymbol{v}_k^*}(\boldsymbol{v}_k)\|_2^2).$$

From (ER.51), it holds that

$$\frac{\partial \widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)}{\partial \boldsymbol{v}_k^T}\Big|_{\boldsymbol{v}_k^*} = -n^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{u}_k^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k\mathbf{M}_k^v.$$

Then by Proposition 2 that $\mathbf{M}_k^v = -z_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\mathbf{C}_{-k}$ and the initial estimates in Definition 2, we have that

$$\begin{aligned}
&\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \widetilde{\boldsymbol{\eta}}_k) - \widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \boldsymbol{\eta}_k^*) \\
&= (-n^{-1}\mathbf{X}^T\mathbf{Y} + \widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{C}}_{-k})(\mathbf{I}_q - \boldsymbol{v}_k^*\boldsymbol{v}_k^{*T})\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k) + \boldsymbol{r}_{\boldsymbol{v}_k^*} \\
&= (\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}^*) - n^{-1}\mathbf{X}^T\mathbf{E})(\mathbf{I}_q - \boldsymbol{v}_k^*\boldsymbol{v}_k^{*T})\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k) + \boldsymbol{r}_{\boldsymbol{v}_k^*},
\end{aligned} \tag{ER.52}$$

where we slightly abuse the notation and denote the Taylor remainder term as

$$\|\boldsymbol{r}_{\boldsymbol{v}_k^*}\|_2 = O(\|\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k)\|_2^2).$$

We next bound term $\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \widetilde{\boldsymbol{\eta}}_k) - \widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \boldsymbol{\eta}_k^*))$ above. Observe that

$$\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \widetilde{\boldsymbol{\eta}}_k) - \widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \boldsymbol{\eta}_k^*))| \\
&\le |\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}^*) - n^{-1}\mathbf{X}^T\mathbf{E})(\mathbf{I}_q - \boldsymbol{v}_k^*\boldsymbol{v}_k^{*T})\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k)| + |\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\boldsymbol{r}_{\boldsymbol{v}_k^*}|.
\end{aligned}$$

By Lemma EC.21, it can be seen that

$$\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}^*)\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2\left\{n^{-1}\log(pq)\right\}^{1/2}. \tag{ER.53}$$

Denote by $\widetilde{\boldsymbol{w}}_{k,i}^T$ the ith row of $\widetilde{\mathbf{W}}_k$ for $i = 1, \cdots, p$. By parts (a) and (b) of Lemma EC.22, we have that

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_{k,i}\|_0 \le 2\max\{s_{\max}, 3(r^* + s_u + s_v)\} \text{ and } \max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_{k,i}\|_2 \le c. \tag{ER.54}$$

Using similar arguments as for (ER.13) and (ER.14), it holds that

$$\|\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k)\|_0 \le c(r^* + s_u + s_v), \tag{ER.55}$$

$$\|\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k)\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}/d_k^*. \tag{ER.56}$$

Based on results (ER.53)–(ER.56) above, we proceed with following the proof for the rank-2 case. With similar arguments as for (ER.19) and (ER.20), it follows that

$$\begin{aligned}
&|\widetilde{\boldsymbol{w}}_{k,i}^T n^{-1}\mathbf{X}^T\mathbf{E}(\mathbf{I}_q - \boldsymbol{v}_k^*\boldsymbol{v}_k^{*T})\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k)| \\
&\le c\max\{s_{\max}, (r^* + s_u + s_v)\}^{1/2}(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}/d_k^*.
\end{aligned}$$

Further, similar to (ER.21), we can deduce that

$$
\begin{aligned}
&|\widetilde{\boldsymbol{w}}_{k,i}^T\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}_{-k}-\mathbf{C}_{-k}^*)(\mathbf{I}_q-\boldsymbol{v}_k^*\boldsymbol{v}_k^{*T})\exp_{\boldsymbol{v}_k^*}^{-1}(\widetilde{\boldsymbol{v}}_k)|\\
&\leq c(r^*+s_u+s_v)\eta_n^4\left\{n^{-1}\log(pq)\right\}/d_k^*.
\end{aligned}
$$

Thus, similar to (ER.23), combining the above results gives that

$$
\begin{aligned}
&\left|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}_{-k}-\mathbf{C}_{-k}^*)-n^{-1}\mathbf{X}^T\mathbf{E})(\mathbf{I}_q-\boldsymbol{v}_k^*\boldsymbol{v}_k^{*T})\exp_{\boldsymbol{v}_k^*}^{-1}(\widetilde{\boldsymbol{v}}_k)\right| \quad\quad \text{(ER.57)}\\
&\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2\max\{s_{\max}^{1/2},(r^*+s_u+s_v)^{1/2},\eta_n^2\}(r^*+s_u+s_v)\eta_n^2\{n^{-1}\log(pq)\}/d_k^*.
\end{aligned}
$$

Moreover, an application of similar arguments as for (ER.24) shows that

$$
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\boldsymbol{r}_{\boldsymbol{v}_k^*}|\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\{n^{-1}\log(pq)\}/d_k^{*2}. \quad\quad \text{(ER.58)}
$$

Under Condition 4 that the nonzero eigenvalues d_i^{*2} are at the constant level, combining the above results yields that

$$
\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k,\widetilde{\boldsymbol{\eta}}_k)-\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k,\boldsymbol{\eta}_k^*))|\\
&\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2\max\{s_{\max}^{1/2},(r^*+s_u+s_v)^{1/2},\eta_n^2\}(r^*+s_u+s_v)\eta_n^2\{n^{-1}\log(pq)\}, \quad\quad \text{(ER.59)}
\end{aligned}
$$

which completes the proof of Lemma 8.

ER.1.9. Proof of Lemma 9

The proof of Lemma 9 follows similar technical arguments as in the proof of Lemma 5. We will first show that $\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k$, $\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k-\boldsymbol{a}^T\mathbf{W}_k^*\mathbf{M}_k^*$, and $\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k-\mathbf{W}_k^*)$ are s-sparse with $s=c(r^*+s_u+s_v)$. It follows from the sparsity of $\mathbf{U}^*$ and $\mathbf{V}^*$ and (ER.106) that

$$
\sum_{1\leq i\leq r^*}\|\boldsymbol{u}_i^*\|_0\leq s_u,\ \sum_{1\leq i\leq r^*}\|\boldsymbol{v}_i^*\|_0\leq s_v,\ \sum_{1\leq i\leq r^*}\|\widetilde{\boldsymbol{u}}_i\|_0\leq c(r^*+s_u+s_v).
$$

Also, in view of part (b) of Lemma EC.17, we have that

$$
\sum_{1\leq i\leq r^*}\|\widetilde{d}_i\widetilde{\boldsymbol{v}}_i\|_0\leq\sum_{1\leq i\leq r^*}\|\widetilde{d}_i(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\|_0+\sum_{1\leq i\leq r^*}\|\boldsymbol{v}_i^*\|_0\leq c(r^*+s_u+s_v).
$$

Note that $\widetilde{\mathbf{M}}_k=-\widetilde{z}_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\sum_{i\neq k}\widetilde{\boldsymbol{u}}_i\widetilde{\boldsymbol{v}}_i^T$. With similar arguments as for (ER.31)–(ER.33), we can deduce that

$$
\begin{aligned}
&\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\|_0=\|\sum_{i\neq k}(\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{z}_{ii}^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{l}}_i)\cdot\widetilde{d}_i\widetilde{\boldsymbol{v}}_i^T\|_0\leq\sum_{i\neq k}\|\widetilde{d}_i\widetilde{\boldsymbol{v}}_i\|_0\leq c(r^*+s_u+s_v),\\
&\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k-\boldsymbol{a}^T\mathbf{W}_k^*\mathbf{M}_k^*\|_0\leq\sum_{i\neq k}\|\widetilde{d}_i\widetilde{\boldsymbol{v}}_i\|_0+\sum_{i\neq k}\|\boldsymbol{v}_i^*\|_0\leq c(r^*+s_u+s_v),\\
&\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k-\mathbf{W}_k^*)\|_0\leq\sum_{i\neq k}\|\widetilde{\boldsymbol{u}}_i\|_0+\sum_{i\neq k}\|\boldsymbol{u}_i^*\|_0\leq c(r^*+s_u+s_v).
\end{aligned}
$$

Then similar to (ER.34), it holds that

$$\begin{aligned}|-\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\epsilon}}_k - h_k/\sqrt{n}| \leq &\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\|_2\|n^{-1}\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)\|_{2,s}\\ &+\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k - \boldsymbol{a}^T\mathbf{W}_k^*\mathbf{M}_k^*\|_2\|n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k^*\|_{2,s}\\ &+\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\|_2\|n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_k^*\|_{2,s}.\end{aligned} \tag{ER.60}$$

An application of similar arguments as for (ER.35)–(ER.37) gives that

$$n^{-1}\|\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)\|_{2,s} \leq cs^{3/2}\eta_n^2\{n^{-1}\log(pq)\}, \tag{ER.61}$$

$$n^{-1}\|\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k^*\|_{2,s} \leq cs^{1/2}s_u^{1/2}\{n^{-1}\log(pq)\}^{1/2}d_k^*, \tag{ER.62}$$

$$n^{-1}\|\mathbf{X}^T\mathbf{E}\boldsymbol{v}_k^*\|_{2,s} \leq cs^{1/2}s_v^{1/2}\{n^{-1}\log(pq)\}^{1/2}. \tag{ER.63}$$

From Lemma EC.21, we have that

$$\|\mathbf{M}_k^*\|_2 \leq cd_k^{*-2}d_1^*, \quad \|\widetilde{\mathbf{M}}_k\|_2 \leq cd_k^{*-2}d_1^*, \tag{ER.64}$$

$$\|\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*\|_2 \leq c(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_k^{*-3}d_1^*. \tag{ER.65}$$

Along with parts (d) and (e) of Lemma EC.22, it follows that

$$\begin{aligned}&\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\|_2 \leq cm^{1/2}(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2},\\ &\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k - \boldsymbol{a}^T\mathbf{W}_k^*\mathbf{M}_k^*\|_2 \leq \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*)\|_2 + \|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\mathbf{M}_k^*\|_2\\ &\leq \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*\|_2 + \|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\|_2\|\mathbf{M}_k^*\|_2\\ &\leq cm^{1/2}(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_k^{*-2}d_1^*.\end{aligned} \tag{ER.66}$$

Therefore, by (ER.60)–(ER.66), Lemma EC.22, and Condition 4 that the nonzero eigenvalues d_i^{*2} are at the constant level, we can obtain that

$$|-\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\epsilon}}_k - h_k/\sqrt{n}| \leq cm^{1/2}(r^* + s_u + s_v)^{3/2}\eta_n^2\{n^{-1}\log(pq)\}.$$

This completes the proof of Lemma 9.

ER.1.10. Proof of Lemma EC.10

The proof of Lemma EC.10 is similar to that of Lemma 8 for the general rank case in Section ER.1.8. Notice that the nuisance parameter is $\boldsymbol{\eta}_k = \left[\boldsymbol{v}_k^T, \boldsymbol{v}_{k+1}^T \cdots, \boldsymbol{v}_{r^*}^T, \boldsymbol{u}_{k+1}^T, \cdots, \boldsymbol{u}_{r^*}^T\right]^T$. By the definition of $\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)$, we have that

$$\begin{aligned}\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k) &= \frac{\partial L}{\partial \boldsymbol{u}_k} - \mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_k}\\ &= \frac{\partial L}{\partial \boldsymbol{u}_k} - \left(\mathbf{M}_k^v\frac{\partial L}{\partial \boldsymbol{v}_k} + \sum_{j=k+1}^{r^*}\mathbf{M}_j^u\frac{\partial L}{\partial \boldsymbol{u}_j} + \sum_{j=k+1}^{r^*}\mathbf{M}_j^v\frac{\partial L}{\partial \boldsymbol{v}_j}\right).\end{aligned}$$

From Proposition 4, we see that $\mathbf{M}_j^u = \mathbf{0}$ and $\mathbf{M}_j^v = \mathbf{0}$ for each $j \in \{k+1, \cdots, r^*\}$, which means that we need only to consider $\boldsymbol{v}_k$ as the nuisance parameter. In light of the derivatives (ER.77) and (ER.78), it holds that

$$\begin{aligned}
\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k) &= \frac{\partial L}{\partial \boldsymbol{u}_k} - \mathbf{M}_k^v \frac{\partial L}{\partial \boldsymbol{v}_k} \\
&= \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k - n^{-1} \mathbf{X}^T \mathbf{Y} \boldsymbol{v}_k + \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{C}}^{(1)} \boldsymbol{v}_k \\
&\quad - \mathbf{M}_k^v (\boldsymbol{v}_k \boldsymbol{u}_k^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k - n^{-1} \mathbf{Y}^T \mathbf{X} \boldsymbol{u}_k + (\widehat{\mathbf{C}}^{(1)})^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k).
\end{aligned}$$

For each arbitrary fixed $\mathbf{M}_k^v$, we can see from the representation above that $\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)$ is only a function of $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$, which entails that we need only to do the Taylor expansion of $\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)$ with respect to $\boldsymbol{v}_k$. Similar to the proof of Lemma EC.16, we can obtain the Taylor expansion of $\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)$

$$\begin{aligned}
\widetilde{\psi}_k\left(\boldsymbol{u}_k, \boldsymbol{\eta}_k\right) &= \widetilde{\psi}_k\left(\boldsymbol{u}_k, \boldsymbol{\eta}_k^*\right) + \frac{\partial \widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)}{\partial \boldsymbol{v}_k^T}\Big|_{\boldsymbol{v}_k^*} (\mathbf{I}_q - \boldsymbol{v}_k^* \boldsymbol{v}_k^{*T}) \exp_{\boldsymbol{v}_k^*}^{-1}(\boldsymbol{v}_k) \\
&\quad + \boldsymbol{r}_{\boldsymbol{v}_k^*},
\end{aligned}$$

where the Taylor remainder term satisfies that

$$\|\boldsymbol{r}_{\boldsymbol{v}_k^*}\|_2 = O(\| \exp_{\boldsymbol{v}_k^*}^{-1}(\boldsymbol{v}_k)\|_2^2).$$

Moreover, it follows that

$$\frac{\partial \widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k)}{\partial \boldsymbol{v}_k^T}\Big|_{\boldsymbol{v}_k^*} = -n^{-1} \mathbf{X}^T \mathbf{Y} + \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{C}}^{(1)} - \boldsymbol{u}_k^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k \mathbf{M}_k^v.$$

Then from Proposition 4 that $\mathbf{M}_k^v = -z_{kk}^{-1} \widehat{\boldsymbol{\Sigma}} \mathbf{C}^{(2)}$ and the initial estimates in Definition 2, we can deduce that

$$\begin{aligned}
&\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \widetilde{\boldsymbol{\eta}}_k) - \widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \boldsymbol{\eta}_k^*) \\
&= (-n^{-1} \mathbf{X}^T \mathbf{Y} + \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{C}}^{(1)} + \widehat{\boldsymbol{\Sigma}} \widetilde{\mathbf{C}}^{(2)}) (\mathbf{I}_q - \boldsymbol{v}_k^* \boldsymbol{v}_k^{*T}) \exp_{\boldsymbol{v}_k^*}^{-1}(\widetilde{\boldsymbol{v}}_k) + \boldsymbol{r}_{\boldsymbol{v}_k^*} \\
&= (\widehat{\boldsymbol{\Sigma}} (\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}) - n^{-1} \mathbf{X}^T \mathbf{E}) (\mathbf{I}_q - \boldsymbol{v}_k^* \boldsymbol{v}_k^{*T}) \exp_{\boldsymbol{v}_k^*}^{-1}(\widetilde{\boldsymbol{v}}_k) + \boldsymbol{r}_{\boldsymbol{v}_k^*},
\end{aligned}$$

where we slightly abuse the notation and denote the Taylor remainder term as

$$\boldsymbol{r}_{\boldsymbol{v}_k^*} = O(\| \exp_{\boldsymbol{v}_k^*}^{-1}(\widetilde{\boldsymbol{v}}_k)\|_2^2).$$

We next bound term $\boldsymbol{a}^T \widetilde{\mathbf{W}}_k (\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \widetilde{\boldsymbol{\eta}}_k) - \widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \boldsymbol{\eta}_k^*))$ above, which will follow similar arguments as for (ER.53)–(ER.59). Observe that

$$\begin{aligned}
&|\boldsymbol{a}^T \widetilde{\mathbf{W}}_k (\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \widetilde{\boldsymbol{\eta}}_k) - \widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \boldsymbol{\eta}_k^*))| \\
&\le |\boldsymbol{a}^T \widetilde{\mathbf{W}}_k (\widehat{\boldsymbol{\Sigma}} (\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}^*) - n^{-1} \mathbf{X}^T \mathbf{E}) (\mathbf{I}_q - \boldsymbol{v}_k^* \boldsymbol{v}_k^{*T}) \exp_{\boldsymbol{v}_k^*}^{-1}(\widetilde{\boldsymbol{v}}_k)| \\
&\quad + |\boldsymbol{a}^T \widetilde{\mathbf{W}}_k \boldsymbol{r}_{\boldsymbol{v}_k^*}|.
\end{aligned}$$

Denote by $\widetilde{\boldsymbol{w}}_{k,i}^T$ the ith row of $\widetilde{\mathbf{W}}_k$ with $i = 1, \cdots, p$. From Lemma EC.26, we have that

$$\max_{1\le i\le p} \|\widetilde{\boldsymbol{w}}_{k,i}\|_0 \le 2\max\{s_{\max}, 3(r^* + s_u + s_v)\} \text{ and } \max_{1\le i\le p} \|\widetilde{\boldsymbol{w}}_{k,i}\|_2 \le c,$$

which has the same upper bound as in (ER.54). An application of similar arguments as for (ER.13) and (ER.14) leads to

$$\begin{aligned}
&\|\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k)\|_0 \le c(r^* + s_u + s_v),\\
&\|\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k)\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}/d_k^*.
\end{aligned}$$

Similar to (ER.53), it also holds that

$$\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}^*)\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2\left\{n^{-1}\log(pq)\right\}^{1/2}.$$

In view of the above results, we can see that the upper bound for $|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \widetilde{\boldsymbol{\eta}}_k) - \widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \boldsymbol{\eta}_k^*))|$ is similar to that for the general rank case of Lemma 8 in Section ER.1.8. Similar to (ER.52)–(ER.57), it follows that

$$\begin{aligned}
&\left|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}^*) - n^{-1}\mathbf{X}^T\mathbf{E})(\mathbf{I}_q - \boldsymbol{v}_k^*\boldsymbol{v}_k^{*T})\exp^{-1}_{\boldsymbol{v}_k^*}(\widetilde{\boldsymbol{v}}_k)\right|\\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2\max\{s_{\max}^{1/2}, (r^* + s_u + s_v)^{1/2}, \eta_n^2\}(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}/d_k^*.
\end{aligned}$$

Further, from (ER.58) we can show that

$$|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\boldsymbol{r}_{\boldsymbol{v}_k^*}| \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^* + s_u + s_v)\eta_n^4\{n^{-1}\log(pq)\}/d_k^{*2}.$$

Therefore, it holds that

$$\begin{aligned}
&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \widetilde{\boldsymbol{\eta}}_k) - \widetilde{\psi}_k(\widetilde{\boldsymbol{u}}_k, \boldsymbol{\eta}_k^*))|\\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2\max\{s_{\max}^{1/2}, (r^* + s_u + s_v)^{1/2}, \eta_n^2\}(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}\\
&\quad \times\max\{d_k^{*-1}, d_k^{*-2}\},
\end{aligned}$$

which concludes the proof of Lemma EC.10.

ER.1.11. Proof of Lemma EC.11

This proof follows similar technical arguments as in the proof of Lemma 9 in Section ER.1.9. Note that $\widetilde{\mathbf{M}}_k = -\widetilde{z}_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{C}}^{(2)}$ and $\mathbf{M}_k^* = -z_{kk}^{*-1}\widehat{\boldsymbol{\Sigma}}\mathbf{C}^{*(2)}$. Using similar arguments, we can show that $\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k$,

$\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k - \boldsymbol{a}^T\mathbf{W}_k^*\mathbf{M}_k^*$, and $\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)$ are all s-sparse with $s = c(r^* + s_u + s_v)$. An application of similar arguments as for (ER.60) and (ER.61)–(ER.63) gives that

$$\begin{aligned}
&|-\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\epsilon}}_k - h_k/\sqrt{n}| \le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\|_2\|n^{-1}\mathbf{E}^T\mathbf{X}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)\|_{2,s} \\
&\quad + \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k - \boldsymbol{a}^T\mathbf{W}_k^*\mathbf{M}_k^*\|_2\|n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k^*\|_{2,s} \\
&\quad + \|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\|_2\|n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_k^*\|_{2,s} \\
&\le cs^{3/2}\eta_n^2\{n^{-1}\log(pq)\}\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\|_2 \\
&\quad + cs^{1/2}s_v^{1/2}\{n^{-1}\log(pq)\}^{1/2}\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\|_2 \\
&\quad + cs^{1/2}s_u^{1/2}\{n^{-1}\log(pq)\}^{1/2}d_k^*\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k - \boldsymbol{a}^T\mathbf{W}_k^*\mathbf{M}_k^*\|_2.
\end{aligned}$$

For the terms above, from Lemma EC.25 we have that

$$\begin{aligned}
&\|\mathbf{M}_k^*\|_2 \le cd_k^{*-2}d_{k+1}^*, \ \|\widetilde{\mathbf{M}}_k\|_2 \le cd_k^{*-2}d_{k+1}^*, \\
&\|\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_k^{*-2},
\end{aligned}$$

Moreover, it follows from parts (d) and (e) of Lemma EC.26 that

$$\begin{aligned}
&\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\|_2 \le cm^{1/2}(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_{k+1}^*d_k^{*-2}, \\
&\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2 \le cm^{1/2}.
\end{aligned}$$

Then it holds that

$$\begin{aligned}
&\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k - \boldsymbol{a}^T\mathbf{W}_k^*\mathbf{M}_k^*\|_2 \le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k(\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*)\|_2 + \|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\mathbf{M}_k^*\|_2 \\
&\le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*\|_2 + \|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_k - \mathbf{W}_k^*)\|_2\|\mathbf{M}_k^*\|_2 \\
&\le cm^{1/2}(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_k^{*-2}.
\end{aligned}$$

Thus, combining the above terms yields that

$$|-\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\epsilon}}_k - h_k/\sqrt{n}| \le cm^{1/2}(r^* + s_u + s_v)^{3/2}\eta_n^2\{n^{-1}\log(pq)\}d_k^{*-1}.$$

This completes the proof of Lemma EC.11.

ER.1.12. Proof of Lemma EC.12

Observe that $\widetilde{\mathbf{M}}_k = -\widetilde{z}_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{C}}^{(2)}$. With the aid of similar arguments as for (ER.48)–(ER.49), it holds that

$$\|\widetilde{\mathbf{M}}_k(\mathbf{C}^{*(2)})^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*\|_2 \le c\sum_{j=k+1}^{r^*}(d_j^{*2}/d_k^*)|\boldsymbol{l}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_k^*|.$$

Together with Condition 5 that $\sum_{j=k+1}^{r^*}(d_j^{*2}/d_k^*)|\boldsymbol{l}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_k^*| = o(n^{-1/2})$ and Lemma EC.26 that $\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2 \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2$, it follows that

$$
\begin{aligned}
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\mathbf{C}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*| &\le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\mathbf{C}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*\|_2 \\
&= o(\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 n^{-1/2}),
\end{aligned}
$$

which concludes the proof of Lemma EC.12.

ER.1.13. Proof of Lemma EC.13

Denote by $\widetilde{\boldsymbol{\delta}}_k = \widetilde{\boldsymbol{\delta}}_{0,k} + \widetilde{\boldsymbol{\delta}}_{1,k}$ with

$$
\begin{aligned}
\widetilde{\boldsymbol{\delta}}_{0,k} &= \widetilde{\mathbf{M}}_k((\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*)\widetilde{\boldsymbol{u}}_k^T - (\widetilde{\mathbf{C}}^{(2)} - \mathbf{C}^{*(2)})^T)\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*), \\
\widetilde{\boldsymbol{\delta}}_{1,k} &= -\widetilde{\mathbf{M}}_k(\widehat{\mathbf{C}}^{(1)} - \mathbf{C}^{*(1)})^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_k.
\end{aligned}
$$

The derivation for the upper bound on $|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_{0,k}|$ is similar to that in the proof of Lemma 6 in Section ER.1.6. From Lemma EC.25, we see that $\|\widetilde{\mathbf{M}}_k\|_2 \le c d_k^{*-2} d_{k+1}^*$. Part (e) of Lemma EC.26 entails that $\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2 \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2$. Further, observe that $\widetilde{\mathbf{C}}^{(2)}$ and $\mathbf{C}^{*(2)}$ are submatrices of $\widetilde{\mathbf{C}}_{-k}$ and $\mathbf{C}_{-k}^*$, respectively. It follows from similar arguments as for (ER.103)–(ER.105) that

$$
\|\widetilde{\mathbf{C}}^{(2)} - \mathbf{C}^{*(2)}\|_2 \le c\gamma_n,
$$

where $\gamma_n = (r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}$. From the above results and Lemma EC.17, we can deduce that

$$
\begin{aligned}
|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_{0,k}| &\le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\|_2\|(\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*)\widetilde{\boldsymbol{u}}_k^T - (\widetilde{\mathbf{C}}^{(2)} - \mathbf{C}^{*(2)})^T\|_2\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)\|_2 \\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2\gamma_n d_k^{*-2}d_{k+1}^*(\|\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*\|_2\|\widetilde{\boldsymbol{u}}_k^T\|_2 + \|(\widetilde{\mathbf{C}}^{(2)} - \mathbf{C}^{*(2)})^T\|_2) \\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^* + s_u + s_v)\eta_n^4\{n^{-1}\log(pq)\}d_k^{*-2}d_{k+1}^*. \qquad \text{(ER.67)}
\end{aligned}
$$

We next bound term $|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_{1,k}|$ above. It can be seen that

$$
\begin{aligned}
\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_{1,k} = &-\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widehat{\mathbf{C}}^{(1)} - \mathbf{C}^{*(1)})^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^* \\
&+ \boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widehat{\mathbf{C}}^{(1)} - \mathbf{C}^{*(1)})^T\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k^* - \widetilde{\boldsymbol{u}}_k).
\end{aligned}
$$

Notice that $\widetilde{\boldsymbol{v}}_j^T\widetilde{\boldsymbol{v}}_i = 0$ and $\boldsymbol{v}_j^{*T}\boldsymbol{v}_i^* = 0$ for each $1 \le i \le k-1$ and $k+1 \le j \le r^*$. It holds that

$$
\widetilde{\mathbf{M}}_k\widetilde{\boldsymbol{v}}_i = -\widetilde{z}_{kk}^{-1}\widehat{\boldsymbol{\Sigma}}\sum_{j=k+1}^{r^*}\widetilde{\boldsymbol{u}}_j\widetilde{\boldsymbol{v}}_j^T\widetilde{\boldsymbol{v}}_i = \mathbf{0} \ \text{ and } \ \mathbf{M}_k^*\boldsymbol{v}_i^* = -z_{kk}^{*-1}\widehat{\boldsymbol{\Sigma}}\sum_{j=k+1}^{r^*}\boldsymbol{u}_i^*\boldsymbol{v}_j^{*T}\boldsymbol{v}_i^* = \mathbf{0}.
$$

Then we can show that

$$
\begin{aligned}
&-\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widehat{\mathbf{C}}^{(1)}-\mathbf{C}^{*(1)})^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^* = -\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\sum_{i=1}^{k-1}(\widetilde{\boldsymbol{v}}_i\widetilde{\boldsymbol{u}}_i^T-\boldsymbol{v}_i^*\boldsymbol{u}_i^{*T})\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*\\
&=\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\sum_{i=1}^{k-1}\boldsymbol{v}_i^*\boldsymbol{u}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^* = \boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\sum_{i=1}^{k-1}(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*-\widetilde{\boldsymbol{v}}_i)\boldsymbol{u}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*\\
&=\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\sum_{i=1}^{k-1}(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\boldsymbol{u}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*.
\end{aligned}
\tag{ER.68}
$$

Hence, it follows that

$$
\begin{aligned}
\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_{1,k} &= -\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k\sum_{i=1}^{k-1}(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\boldsymbol{u}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^* + \boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widehat{\mathbf{C}}^{(1)}-\mathbf{C}^{*(1)})^T\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k^*-\widetilde{\boldsymbol{u}}_k)\\
&=: B_1 + B_2.
\end{aligned}
$$

We will bound the two terms B_1 and B_2 introduced above separately.

For the first term B_1 above, we have that

$$
\begin{aligned}
&|\sum_{i=1}^{k-1}\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\boldsymbol{u}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*|\\
&\le \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\|_2\sum_{i=1}^{k-1}\|d_i^*(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\|_2|d_k^*||\boldsymbol{l}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_k^*|\\
&\le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 d_k^{*-1}d_{k+1}^*\sum_{i=1}^{k-1}\|d_i^*(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\|_2|\boldsymbol{l}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_k^*|,
\end{aligned}
$$

where the last step above has used Lemma EC.25 and part (c) of Lemma EC.26. For term $|\boldsymbol{l}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_k^*|$ with $i=1,\cdots,k$, in view of Condition 5 it holds that

$$
(d_k^{*2}/d_i^*)|\boldsymbol{l}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_k^*| = o(n^{-1/2}).
$$

Together with part (a) of Lemma EC.17, we can obtain that

$$
\sum_{i=1}^{k-1}\|d_i^*(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\|_2|\boldsymbol{l}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_k^*| \le c\gamma_n\sum_{i=1}^{k-1}\frac{d_i^*}{d_k^{*2}\sqrt{n}},
$$

where $\gamma_n = (r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}$. This further leads to

$$
|\sum_{i=1}^{k-1}\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widetilde{\boldsymbol{v}}_i-\boldsymbol{v}_i^*)\boldsymbol{u}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*| \le c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2\gamma_n\sum_{i=1}^{k-1}d_i^*d_{k+1}^*d_k^{*-3}n^{-1/2}. \tag{ER.69}
$$

For term B_2 above, let us first bound term $\|\widehat{\mathbf{C}}^{(1)}-\mathbf{C}^{*(1)}\|_2$. Observe that $\widehat{\mathbf{C}}^{(1)} = \widetilde{\mathbf{U}}^{(1)}(\widetilde{\mathbf{V}}^{(1)})^T$ and $\mathbf{C}^{*(1)} = \mathbf{U}^{*(1)}(\mathbf{V}^{*(1)})^T$, where $\widetilde{\mathbf{U}}^{(1)} = (\widetilde{\boldsymbol{u}}_1,\cdots,\widetilde{\boldsymbol{u}}_{k-1})$, $\widetilde{\mathbf{V}}^{(1)} = (\widetilde{\boldsymbol{v}}_1,\cdots,\widetilde{\boldsymbol{v}}_{k-1})$, $\mathbf{U}^{*(1)} = (\boldsymbol{u}_1^*,\cdots,\boldsymbol{u}_{k-1}^*)$, and $\mathbf{V}^{*(1)} = (\boldsymbol{v}_1^*,\cdots,\boldsymbol{v}_{k-1}^*)$. Then it follows that

$$
\|\widehat{\mathbf{C}}^{(1)}-\mathbf{C}^{*(1)}\|_2 = \|\widetilde{\mathbf{U}}^{(1)}(\widetilde{\mathbf{V}}^{(1)})^T - \mathbf{U}^{*(1)}(\mathbf{V}^{*(1)})^T\|_2
$$

$$\leq \|\widetilde{\mathbf{U}}^{(1)}(\widetilde{\mathbf{V}}^{(1)}-\mathbf{V}^{*(1)})^T\|_2+\|(\widetilde{\mathbf{U}}^{(1)}-\mathbf{U}^{*(1)})(\mathbf{V}^{*(1)})^T\|_2$$
$$\leq \|\widetilde{\mathbf{L}}^{(1)}\|_2\|\widetilde{\mathbf{D}}^{(1)}(\widetilde{\mathbf{V}}^{(1)}-\mathbf{V}^{*(1)})^T\|_2+\|\widetilde{\mathbf{U}}^{(1)}-\mathbf{U}^{*(1)}\|_2\|(\mathbf{V}^{*(1)})^T\|_2.$$

It can be seen that $\|\widetilde{\mathbf{L}}^{(1)}\|_2=\|(\mathbf{V}^{*(1)})^T\|_2=1$. For term $\|(\widetilde{\mathbf{V}}^{(1)}-\mathbf{V}^{*(1)})\widetilde{\mathbf{D}}^{(1)}\|_2$, we have that

$$\begin{aligned}\|(\widetilde{\mathbf{V}}^{(1)}-\mathbf{V}^{*(1)})\widetilde{\mathbf{D}}^{(1)}\|_2 &\leq \|\widetilde{\mathbf{V}}^{(1)}\widetilde{\mathbf{D}}^{(1)}-\mathbf{V}^{*(1)}\mathbf{D}^{*(1)}\|_2+\|\widetilde{\mathbf{V}}^{(1)}\|_2\|\widetilde{\mathbf{D}}^{(1)}-\mathbf{D}^{*(1)}\|_2\\ &\leq \|\widetilde{\mathbf{V}}^{(1)}\widetilde{\mathbf{D}}^{(1)}-\mathbf{V}^{*(1)}\mathbf{D}^{*(1)}\|_F+\|\widetilde{\mathbf{D}}^{(1)}-\mathbf{D}^{*(1)}\|_F\\ &\leq \|\widetilde{\mathbf{V}}\widetilde{\mathbf{D}}-\mathbf{V}^*\mathbf{D}^*\|_F+\|\widetilde{\mathbf{D}}-\mathbf{D}^*\|_F\\ &\leq c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2},\end{aligned}$$

where we have used Definition 2 and $\|\widetilde{\mathbf{V}}_{-k}\|_2=1$. Moreover, from Definition 2 it holds that

$$\begin{aligned}\|\widetilde{\mathbf{U}}^{(1)}-\mathbf{U}^{*(1)}\|_2 &\leq \|\widetilde{\mathbf{U}}^{(1)}-\mathbf{U}^{*(1)}\|_F \leq \|\widetilde{\mathbf{U}}-\mathbf{U}^*\|_F\\ &\leq c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned}$$

Combining the above results gives that

$$\|\widehat{\mathbf{C}}^{(1)}-\mathbf{C}^{*(1)}\|_2\leq c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.$$

It follows from part (c) of Lemma EC.17, Lemma EC.25, and Lemma EC.26 that

$$\begin{aligned}&|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\mathbf{M}}_k(\widehat{\mathbf{C}}^{(1)}-\mathbf{C}^{*(1)})^T\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k^*-\widetilde{\boldsymbol{u}}_k)|\\ &\leq \|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\|_2\|\widetilde{\mathbf{M}}_k\|_2\|\widehat{\mathbf{C}}^{(1)}-\mathbf{C}^{*(1)}\|_2\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k-\boldsymbol{u}_k^*)\|_2\\ &\leq c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\{n^{-1}\log(pq)\}d_k^{*-2}d_{k+1}^*.\end{aligned} \tag{ER.70}$$

Thus, a combination of (ER.69) and (ER.70) yields that

$$\begin{aligned}|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_{1,k}| \leq\; & c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}\sum_{i=1}^{k-1}d_i^*d_{k+1}^*d_k^{*-3}n^{-1/2}\\ &+c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\{n^{-1}\log(pq)\}d_k^{*-2}d_{k+1}^*.\end{aligned}$$

Along with (ER.67), it follows that

$$\begin{aligned}|\boldsymbol{a}^T\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\delta}}_k| \leq\; & c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)^{1/2}\eta_n^2\{\log(pq)\}^{1/2}d_{k+1}^*d_k^{*-3}(\sum_{i=1}^{k-1}d_i^*)/n\\ &+c\|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2(r^*+s_u+s_v)\eta_n^4\{n^{-1}\log(pq)\}d_k^{*-2}d_{k+1}^*.\end{aligned}$$

This completes the proof of Lemma EC.13.

ER.1.14. Proof of Lemma EC.14

Under the orthogonality constraints $\boldsymbol{v}_i^T\boldsymbol{v}_j = 0$ for each $i, j \in \{1, \cdots, r^*\}$ with $i \neq j$, we have that

$$L = (2n)^{-1}\left\{\|\mathbf{Y}\|_F^2 + 2\langle \mathbf{Y}, -\mathbf{X}\mathbf{C}_{-k}\rangle + \boldsymbol{u}_k^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}_k\boldsymbol{v}_k^T\boldsymbol{v}_k + \|\mathbf{X}\mathbf{C}_{-k}\|_F^2 - 2\boldsymbol{u}_k^T\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_k\right\}.$$

After some calculations with $\|\boldsymbol{v}_k\|_2 = 1$, we can obtain that

$$\frac{\partial L}{\partial \boldsymbol{u}_k} = \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k - n^{-1}\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_k, \tag{ER.71}$$

$$\frac{\partial L}{\partial \boldsymbol{v}_k} = \boldsymbol{v}_k\boldsymbol{u}_k^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k - n^{-1}\mathbf{Y}^T\mathbf{X}\boldsymbol{u}_k. \tag{ER.72}$$

For each $j \neq k$, similarly we also have that

$$\frac{\partial L}{\partial \boldsymbol{u}_j} = \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j - n^{-1}\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_j, \tag{ER.73}$$

$$\frac{\partial L}{\partial \boldsymbol{v}_j} = \boldsymbol{v}_j\boldsymbol{u}_j^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j - n^{-1}\mathbf{Y}^T\mathbf{X}\boldsymbol{u}_j. \tag{ER.74}$$

It follows from the derivatives (ER.72)–(ER.74) that

$$\begin{aligned}
\mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_k}\Big|_{\boldsymbol{\eta}_k^*} &= \mathbf{M}_k^v\left\{\boldsymbol{v}_k^*\boldsymbol{u}_k^T\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) - \mathbf{C}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k - n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k\right\} \\
&\quad + \sum_{j\neq k}\mathbf{M}_j^v\left\{-\mathbf{C}_{-j}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* - n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_j^*\right\} - \sum_{j\neq k}\mathbf{M}_j^u\mathbf{X}^T\mathbf{E}\boldsymbol{v}_j^* \\
&= \mathbf{M}_k^v(\boldsymbol{v}_k^*\boldsymbol{u}_k^T - \mathbf{C}_{-k}^{*T})\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) - \mathbf{M}_k^v\mathbf{C}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^* \\
&\quad - \sum_{j\neq k}\mathbf{M}_j^v\mathbf{C}_{-j}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* - n^{-1}\mathbf{M}_k^v\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k - n^{-1}\sum_{j\neq k}\mathbf{M}_j^v\mathbf{E}^T\mathbf{X}\boldsymbol{u}_j^* - n^{-1}\sum_{j\neq k}\mathbf{M}_j^u\mathbf{X}^T\mathbf{E}\boldsymbol{v}_j^* \\
&= \mathbf{M}_k^v(\boldsymbol{v}_k\boldsymbol{u}_k^T\widehat{\boldsymbol{\Sigma}} - \mathbf{C}_{-k}^T\widehat{\boldsymbol{\Sigma}})(\boldsymbol{u}_k - \boldsymbol{u}_k^*) - \mathbf{M}_k^v\mathbf{C}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^* - \sum_{j\neq k}\mathbf{M}_j^v\mathbf{C}_{-j}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* + \boldsymbol{\delta}_1',
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\delta}_1' &= -\mathbf{M}_k^v\left\{(\boldsymbol{v}_k - \boldsymbol{v}_k^*)\boldsymbol{u}_k^T - (\mathbf{C}_{-k}^T - \mathbf{C}_{-k}^{*T})\right\}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) \\
&\quad - n^{-1}\mathbf{M}_k^v\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k - n^{-1}\sum_{j\neq k}\mathbf{M}_j^u\mathbf{X}^T\mathbf{E}\boldsymbol{v}_j^* - \sum_{j\neq k}\mathbf{M}_j^v n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_j^*.
\end{aligned} \tag{ER.75}$$

Along with the derivative (ER.71), it holds that

$$\begin{aligned}
\widetilde{\psi}(\boldsymbol{u}_k, \boldsymbol{\eta}_k^*) &= \frac{\partial L}{\partial \boldsymbol{u}_k}\Big|_{\boldsymbol{\eta}_k^*} - \mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_k}\Big|_{\boldsymbol{\eta}_k^*} \\
&= (\mathbf{I}_p - \mathbf{M}_k^v\boldsymbol{v}_k\boldsymbol{u}_k^T + \mathbf{M}_k^v\mathbf{C}_{-k}^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) + \mathbf{M}_k^v\mathbf{C}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^* \\
&\quad + \sum_{j\neq k}\mathbf{M}_j^v\mathbf{C}_{-j}^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* + \boldsymbol{\delta}_1,
\end{aligned} \tag{ER.76}$$

where $\boldsymbol{\delta}_1 = -\boldsymbol{\delta}_1' - n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_k^*$.

Hence, combining (ER.75) and (ER.76) yields that

$$
\begin{aligned}
\widetilde{\psi}(\boldsymbol{u}_k, \boldsymbol{\eta}_k^*) = &(\mathbf{I}_p - \mathbf{M}_k^v \boldsymbol{v}_k \boldsymbol{u}_k^T \widehat{\boldsymbol{\Sigma}} + \mathbf{M}_k^v \mathbf{C}_{-k}^T \widehat{\boldsymbol{\Sigma}})(\boldsymbol{u}_k - \boldsymbol{u}_k^*) \\
&+ \mathbf{M}_k^v \mathbf{C}_{-k}^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^* + \sum_{j \neq k} \mathbf{M}_j^v \mathbf{C}_{-j}^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_j^* + \boldsymbol{\delta}_k + \boldsymbol{\epsilon}_k,
\end{aligned}
$$

where $\boldsymbol{\delta}_k = \left\{\mathbf{M}_k^v(\boldsymbol{v}_k - \boldsymbol{v}_k^*)\boldsymbol{u}_k^T - \mathbf{M}_k^v(\mathbf{C}_{-k}^T - \mathbf{C}_{-k}^{*T})\right\} \widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*)$ and

$$
\boldsymbol{\epsilon}_k = n^{-1} \sum_{j \neq k} \mathbf{M}_j^u \mathbf{X}^T \mathbf{E} \boldsymbol{v}_j^* + n^{-1} \mathbf{M}_k^v \mathbf{E}^T \mathbf{X} \boldsymbol{u}_k + \sum_{j \neq k} \mathbf{M}_j^v n^{-1} \mathbf{E}^T \mathbf{X} \boldsymbol{u}_j^* - n^{-1} \mathbf{X}^T \mathbf{E} \boldsymbol{v}_k^*.
$$

This concludes the proof of Lemma EC.14.

ER.1.15. Proof of Lemma EC.15

Observe that the loss function (9) is equivalent to

$$
\begin{aligned}
L = (2n)^{-1} \Big\{ &\|\mathbf{Y}\|_F^2 - 2\boldsymbol{u}_k^T \mathbf{Y}^T \mathbf{X} \boldsymbol{v}_k + \boldsymbol{u}_k^T \mathbf{X}^T \mathbf{X} \boldsymbol{u}_k \boldsymbol{v}_k^T \boldsymbol{v}_k \\
&+ 2\langle \mathbf{Y}, -\mathbf{X}\widehat{\mathbf{C}}^{(1)} \rangle + 2\langle \mathbf{Y}, -\mathbf{X}\mathbf{C}^{(2)} \rangle + 2\boldsymbol{u}_k^T \mathbf{X}^T \mathbf{X} \widehat{\mathbf{C}}^{(1)} \boldsymbol{v}_k + 2\boldsymbol{u}_k^T \mathbf{X}^T \mathbf{X} \mathbf{C}^{(2)} \boldsymbol{v}_k \\
&+ 2\langle \mathbf{X}\widehat{\mathbf{C}}^{(1)}, \mathbf{X}\mathbf{C}^{(2)} \rangle + \|\mathbf{X}\widehat{\mathbf{C}}^{(1)}\|_F^2 + \|\mathbf{X}\mathbf{C}^{(2)}\|_F^2 \Big\}.
\end{aligned}
$$

For each $j, j' \in \{k+1, \cdots, r^*\}$ with $j \neq j'$, we have $\boldsymbol{v}_k^T \boldsymbol{v}_j = 0$ and $\boldsymbol{v}_j^T \boldsymbol{v}_{j'} = 0$. Then the loss function can be simplified further as

$$
\begin{aligned}
L = (2n)^{-1} \Big\{ &\|\mathbf{Y}\|_F^2 - 2\boldsymbol{u}_k^T \mathbf{Y}^T \mathbf{X} \boldsymbol{v}_k + \boldsymbol{u}_k^T \mathbf{X}^T \mathbf{X} \boldsymbol{u}_k \boldsymbol{v}_k^T \boldsymbol{v}_k \\
&+ 2\langle \mathbf{Y}, -\mathbf{X}\widehat{\mathbf{C}}^{(1)} \rangle + 2\langle \mathbf{Y}, -\mathbf{X}\mathbf{C}^{(2)} \rangle + 2\boldsymbol{u}_k^T \mathbf{X}^T \mathbf{X} \widehat{\mathbf{C}}^{(1)} \boldsymbol{v}_k \\
&+ 2\langle \mathbf{X}\widehat{\mathbf{C}}^{(1)}, \mathbf{X}\mathbf{C}^{(2)} \rangle + \|\mathbf{X}\widehat{\mathbf{C}}^{(1)}\|_F^2 + \sum_{j=k+1}^{r^*} \boldsymbol{u}_j^T \mathbf{X}^T \mathbf{X} \boldsymbol{u}_j \boldsymbol{v}_j^T \boldsymbol{v}_j \Big\}.
\end{aligned}
$$

After some calculations with $\|\boldsymbol{v}_k\|_2 = 1$, we can show that

$$
\frac{\partial L}{\partial \boldsymbol{u}_k} = \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k - n^{-1} \mathbf{X}^T \mathbf{Y} \boldsymbol{v}_k + \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{C}}^{(1)} \boldsymbol{v}_k, \tag{ER.77}
$$

$$
\frac{\partial L}{\partial \boldsymbol{v}_k} = \boldsymbol{v}_k \boldsymbol{u}_k^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k - n^{-1} \mathbf{Y}^T \mathbf{X} \boldsymbol{u}_k + (\widehat{\mathbf{C}}^{(1)})^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k. \tag{ER.78}
$$

For each $j \in \{k+1, \cdots, r^*\}$, similarly by $\|\boldsymbol{v}_j\|_2 = 1$ it follows that

$$
\frac{\partial L}{\partial \boldsymbol{u}_j} = \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_j - n^{-1} \mathbf{X}^T \mathbf{Y} \boldsymbol{v}_j + \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{C}}^{(1)} \boldsymbol{v}_j, \tag{ER.79}
$$

$$
\frac{\partial L}{\partial \boldsymbol{v}_j} = \boldsymbol{v}_j \boldsymbol{u}_j^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_j - n^{-1} \mathbf{Y}^T \mathbf{X} \boldsymbol{u}_j + (\widehat{\mathbf{C}}^{(1)})^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_j. \tag{ER.80}
$$

Note that $\boldsymbol{\eta}_k = \left[\boldsymbol{v}_k^T, \cdots, \boldsymbol{v}_{r^*}^T, \boldsymbol{u}_{k+1}^T, \cdots, \boldsymbol{u}_{r^*}^T\right]^T$ and $\boldsymbol{\eta}_k^* = \left[\boldsymbol{v}_k^{*T}, \cdots, \boldsymbol{v}_{r^*}^{*T}, \boldsymbol{u}_{k+1}^{*T}, \cdots, \boldsymbol{u}_{r^*}^{*T}\right]^T$. Let us simplify (ER.77)–(ER.80) using $\widehat{\mathbf{C}}^{(1)}\boldsymbol{v}_k = 0$ and $\widehat{\mathbf{C}}^{(1)}\boldsymbol{v}_j = 0$. It holds that

$$
\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{u}_k}\Big|_{\boldsymbol{\eta}_k^*} &= \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k - \sum_{l=1}^{r^*}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_l^*\boldsymbol{v}_l^{*T}\boldsymbol{v}_k^* - n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_k^* = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) - n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_k^*,\\
\frac{\partial L}{\partial \boldsymbol{u}_j}\Big|_{\boldsymbol{\eta}_k^*} &= \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* - \sum_{l=1}^{r^*}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_l^*\boldsymbol{v}_l^{*T}\boldsymbol{v}_j^* - n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_j^* = -n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_j^*.
\end{aligned}
$$

Moreover, we can deduce that

$$
\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{v}_k}\Big|_{\boldsymbol{\eta}_k^*} &= \boldsymbol{v}_k^*\boldsymbol{u}_k^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k - \sum_{l=1}^{r^*}\boldsymbol{v}_l^*\boldsymbol{u}_l^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k + \sum_{i=1}^{k-1}\widetilde{\boldsymbol{v}}_i\widehat{\boldsymbol{u}}_i^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k - n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k\\
&= \boldsymbol{v}_k^*\boldsymbol{u}_k^T\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) + \sum_{i=1}^{k-1}(\widetilde{\boldsymbol{v}}_i\widehat{\boldsymbol{u}}_i^T - \boldsymbol{v}_i^*\boldsymbol{u}_i^{*T})\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k - \sum_{j=k+1}^{r^*}\boldsymbol{v}_j^*\boldsymbol{u}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k - n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k,\\
&= (\boldsymbol{v}_k\boldsymbol{u}_k^T - \sum_{j=k+1}^{r^*}\boldsymbol{v}_j\boldsymbol{u}_j^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) - \sum_{j=k+1}^{r^*}\boldsymbol{v}_j^*\boldsymbol{u}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*\\
&\quad + \sum_{j=k+1}^{r^*}(\boldsymbol{v}_j\boldsymbol{u}_j^T - \boldsymbol{v}_j^*\boldsymbol{u}_j^{*T})\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) + \sum_{i=1}^{k-1}(\widetilde{\boldsymbol{v}}_i\widehat{\boldsymbol{u}}_i^T - \boldsymbol{v}_i^*\boldsymbol{u}_i^{*T})\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k\\
&\quad - (\boldsymbol{v}_k - \boldsymbol{v}_k^*)\boldsymbol{u}_k^T\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*) - n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k,\\
\frac{\partial L}{\partial \boldsymbol{v}_j}\Big|_{\boldsymbol{\eta}_k^*} &= \boldsymbol{v}_j^*\boldsymbol{u}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* - \sum_{l=1}^{r^*}\boldsymbol{v}_l^*\boldsymbol{u}_l^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* + \sum_{i=1}^{k-1}\widetilde{\boldsymbol{v}}_i\widehat{\boldsymbol{u}}_i^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* - n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_j^*\\
&= \sum_{i=1}^{k-1}(\widetilde{\boldsymbol{v}}_i\widehat{\boldsymbol{u}}_i^T - \boldsymbol{v}_i^*\boldsymbol{u}_i^{*T})\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* - \sum_{k\le l\le r^*, l\neq j}\boldsymbol{v}_l^*\boldsymbol{u}_l^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^* - n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_j^*.
\end{aligned}
$$

Recall that $\mathbf{M} = \left[\mathbf{M}_k^v, \cdots, \mathbf{M}_{r^*}^v, \mathbf{M}_{k+1}^u, \cdots, \mathbf{M}_{r^*}^u\right]$. Hence, combining the above results leads to

$$
\begin{aligned}
\widetilde{\psi}_k(\boldsymbol{u}_k, \boldsymbol{\eta}_k^*) &= (\mathbf{I}_p - \mathbf{M}_k^v\boldsymbol{v}_k\boldsymbol{u}_k^T + \mathbf{M}_k^v(\mathbf{C}^{(2)})^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*)\\
&\quad + \mathbf{M}_k^v(\mathbf{C}^{*(2)})^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^* + \sum_{i=k+1}^{r^*}\mathbf{M}_i^v(\boldsymbol{v}_k^*\boldsymbol{u}_k^{*T} + (\mathbf{C}_{-i}^{*(2)})^T)\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_i^* + \boldsymbol{\epsilon}_k + \boldsymbol{\delta}_k,
\end{aligned}
$$

where $\boldsymbol{\delta}_k = \mathbf{M}_k^v((\boldsymbol{v}_k - \boldsymbol{v}_k^*)\boldsymbol{u}_k^T - (\mathbf{C}^{(2)} - \mathbf{C}^{*(2)})^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_k - \boldsymbol{u}_k^*)$ and

$$
\begin{aligned}
\boldsymbol{\epsilon}_k &= -n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_k^* + n^{-1}\mathbf{M}_k^v\mathbf{E}^T\mathbf{X}\boldsymbol{u}_k + n^{-1}\sum_{i=k+1}^{r^*}(\mathbf{M}_i^u\mathbf{X}^T\mathbf{E}\boldsymbol{v}_i^* + \mathbf{M}_i^v\mathbf{E}^T\mathbf{X}\boldsymbol{u}_i^*)\\
&\quad - \mathbf{M}_k^v(\widehat{\mathbf{C}}^{(1)} - \mathbf{C}^{*(1)})^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k.
\end{aligned}
$$

This completes the proof of Lemma EC.15.

ER.1.16. Proof of Lemma EC.16

Recall that $\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{\eta}_1) = \frac{\partial L}{\partial \boldsymbol{u}_1} - \mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_1}$, where $\boldsymbol{\eta}_1 = [\boldsymbol{v}_1^T, \boldsymbol{u}_2^T, \boldsymbol{v}_2^T]^T \in \mathbb{R}^{p+2q}$. We will prove the result by conducting the Taylor expansion with respect to $\boldsymbol{v}_1$, $\boldsymbol{u}_2$, and $\boldsymbol{v}_2$, respectively. In order to show the Taylor expansion clearly, let us write function $\widetilde{\psi}_1$ in the form

$$\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{\eta}_1) = \widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2).$$

We will exploit the path below to carry out the Taylor expansion of $\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2)$

$$(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2) \to (\boldsymbol{u}_1, \boldsymbol{v}_1^*, \boldsymbol{u}_2, \boldsymbol{v}_2) \to (\boldsymbol{u}_1, \boldsymbol{v}_1^*, \boldsymbol{u}_2, \boldsymbol{v}_2^*) \to (\boldsymbol{u}_1, \boldsymbol{v}_1^*, \boldsymbol{u}_2^*, \boldsymbol{v}_2^*).$$

Let us first treat $\boldsymbol{u}_1, \boldsymbol{u}_2$, and $\boldsymbol{v}_2$ as fixed and do the expansion of $\widetilde{\psi}_1$ with respect to $\boldsymbol{v}_1$. Since both $\boldsymbol{v}_1$ and $\boldsymbol{v}_1^*$ belong to set $\{\boldsymbol{v} \in \mathbb{R}^q : \boldsymbol{v}^T\boldsymbol{v} = 1\}$, we have that $\boldsymbol{v}_1, \boldsymbol{v}_1^* \in \mathrm{St}(1, q)$ by the definition of the Stiefel manifold. Then by the representation of orthonormal matrices on the Stiefel manifold given in (ER.144) in Section ER.2.2, we see that there exists some tangent vector $\boldsymbol{\xi}_1 \in T_{\boldsymbol{v}_1^*}\mathrm{St}(1, q)$ such that $\boldsymbol{v}_1$ can be represented through the exponential map as

$$\boldsymbol{v}_1 = \exp_{\boldsymbol{v}_1^*}(\boldsymbol{\xi}_1).$$

Meanwhile, the tangent vector $\boldsymbol{\xi}_1$ can be represented as $\boldsymbol{\xi}_1 = \exp_{\boldsymbol{v}_1^*}^{-1}\boldsymbol{v}_1$, where $\exp_{\boldsymbol{v}_1^*}^{-1}$ denotes the inverse of the exponential map.

Then by Lemma EC.27 in Section ER.2.1, we have the first-order Taylor expansion of $\widetilde{\psi}_1$ with respect to $\boldsymbol{v}_1$ given by

$$\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2) = \widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1^*, \boldsymbol{u}_2, \boldsymbol{v}_2) + \langle \nabla_{\boldsymbol{v}_1^*}\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2), \boldsymbol{\xi}_1 \rangle + \boldsymbol{r}_{\boldsymbol{v}_1^*}, \tag{ER.81}$$

where $\nabla_{\boldsymbol{v}_1^*}\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2)$ is the gradient of $\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2)$ with respect to $\boldsymbol{v}_1$ at $\boldsymbol{v}_1^*$ on the Stiefel manifold, $\langle \cdot, \cdot \rangle$ is the metric defined in (ER.148) in Section ER.2.2, and $\boldsymbol{r}_{\boldsymbol{v}_1^*} \in \mathbb{R}^p$ is the corresponding Taylor remainder term satisfying that

$$\|\boldsymbol{r}_{\boldsymbol{v}_1^*}\|_2 = O(\|\boldsymbol{\xi}_1\|_2^2).$$

Applying Lemma EC.28 in Section ER.2.2, the gradient on the Stiefel manifold $\mathrm{St}(1, q)$ is given by

$$\nabla_{\boldsymbol{v}_1^*}\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2) = (\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\frac{\partial \widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2)}{\partial \boldsymbol{v}_1}\Big|_{\boldsymbol{v}_1^*},$$

where $\frac{\partial \widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2)}{\partial \boldsymbol{v}_1}\Big|_{\boldsymbol{v}_1^*}$ represents the partial derivative of $\widetilde{\psi}_1(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2)$ with respect to $\boldsymbol{v}_1$ at $\boldsymbol{v}_1^*$ in the (usual) Euclidean space.

In view of (ER.148) in Section ER.2.2, we further have that

$$\begin{aligned}\langle \nabla_{\boldsymbol{v}_1^*}\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2),\boldsymbol{\xi}_1\rangle &= \mathrm{tr}([\nabla_{\boldsymbol{v}_1^*}\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2)]^T\boldsymbol{\xi}_1) = [\nabla_{\boldsymbol{v}_1^*}\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2)]^T\boldsymbol{\xi}_1\\ &= \frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2)}{\partial \boldsymbol{v}_1^T}\Big|_{\boldsymbol{v}_1^*}(\mathbf{I}_q-\boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1,\end{aligned}$$

where the second equality above holds since $\nabla_{\boldsymbol{v}_1^*}\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2)$ and $\boldsymbol{\xi}_1$ are q-dimensional vectors. Hence, combining the above results leads to

$$\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2) = \widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2) + \frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2)}{\partial \boldsymbol{v}_1^T}\Big|_{\boldsymbol{v}_1^*}(\mathbf{I}_q-\boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1 + \boldsymbol{r}_{\boldsymbol{v}_1^*}, \tag{ER.82}$$

where $\|\boldsymbol{r}_{\boldsymbol{v}_1^*}\|_2 = O(\|\boldsymbol{\xi}_1\|_2^2)$.

Moreover, similar to $\boldsymbol{v}_1$, for the Taylor expansion with respect to $\boldsymbol{v}_2$ we can deduce that

$$\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2) = \widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2^*) + \frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2^T}\Big|_{\boldsymbol{v}_2^*}(\mathbf{I}_q-\boldsymbol{v}_2^*\boldsymbol{v}_2^{*T})\boldsymbol{\xi}_2 + \boldsymbol{r}_{\boldsymbol{v}_2^*}, \tag{ER.83}$$

where $\frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2}\Big|_{\boldsymbol{v}_2^*}$ is the partial derivative of $\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2)$ with respect to $\boldsymbol{v}_2$ at $\boldsymbol{v}_2^*$, $\boldsymbol{\xi}_2 = \exp^{-1}_{\boldsymbol{v}_2^*}(\boldsymbol{v}_2)$ is the corresponding tangent vector, and $\boldsymbol{r}_{\boldsymbol{v}_2^*}\in\mathbb{R}^p$ is the Taylor remainder term satisfying that

$$\|\boldsymbol{r}_{\boldsymbol{v}_2^*}\|_2 = O(\|\boldsymbol{\xi}_2\|_2^2).$$

Since there is no unit length constraint on $\boldsymbol{u}_2$, we can take the Taylor expansion of $\widetilde{\psi}$ with respect to $\boldsymbol{u}_2$ directly on the Euclidean space $\mathbb{R}^p$. It gives that

$$\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2^*) = \widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2^*,\boldsymbol{v}_2^*) + \frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2^*)}{\partial \boldsymbol{u}_2^T}\Big|_{\boldsymbol{u}_2^*}(\boldsymbol{u}_2-\boldsymbol{u}_2^*) + \boldsymbol{r}_{\boldsymbol{u}_2^*}, \tag{ER.84}$$

where $\frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2^*)}{\partial \boldsymbol{u}_2}\Big|_{\boldsymbol{u}_2^*}$ is the partial derivative of $\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2^*)$ with respect to $\boldsymbol{u}_2$ at $\boldsymbol{u}_2^*$, and $\boldsymbol{r}_{\boldsymbol{u}_2^*}\in\mathbb{R}^p$ is the corresponding Taylor remainder term satisfying that

$$\|\boldsymbol{r}_{\boldsymbol{u}_2^*}\|_2 = O(\|\boldsymbol{u}_2-\boldsymbol{u}_2^*\|_2^2).$$

Combining (ER.82)–(ER.83), we can obtain that

$$\begin{aligned}\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2) &= \widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2^*,\boldsymbol{v}_2^*) + \frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2)}{\partial \boldsymbol{v}_1^T}\Big|_{\boldsymbol{v}_1^*}(\mathbf{I}_q-\boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1\\ &\quad + \frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2^T}\Big|_{\boldsymbol{v}_2^*}(\mathbf{I}_q-\boldsymbol{v}_2^*\boldsymbol{v}_2^{*T})\boldsymbol{\xi}_2 + \frac{\partial\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2^*)}{\partial \boldsymbol{u}_2^T}\Big|_{\boldsymbol{u}_2^*}(\boldsymbol{u}_2-\boldsymbol{u}_2^*)\\ &\quad + \boldsymbol{r}_{\boldsymbol{v}_1^*} + \boldsymbol{r}_{\boldsymbol{u}_2^*} + \boldsymbol{r}_{\boldsymbol{v}_2^*}.\end{aligned} \tag{ER.85}$$

On the other hand, it follows from the definition of $\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{\eta}_1)$ with $\boldsymbol{\eta}_1 = [\boldsymbol{v}_1^T,\boldsymbol{u}_2^T,\boldsymbol{v}_2^T]^T$ that

$$\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{\eta}_1) = \frac{\partial L}{\partial \boldsymbol{u}_1} - \mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_1} = \frac{\partial L}{\partial \boldsymbol{u}_1} - \mathbf{M}_1\frac{\partial L}{\partial \boldsymbol{v}_1} - \mathbf{M}_2\frac{\partial L}{\partial \boldsymbol{u}_2} - \mathbf{M}_3\frac{\partial L}{\partial \boldsymbol{v}_2}.$$

Through some calculations with (ER.2)–(ER.5), we can show that

$$\frac{\partial \widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2)}{\partial \boldsymbol{v}_1^T}\Big|_{\boldsymbol{v}_1^*} = -n^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\mathbf{M}_1,$$
$$\frac{\partial \widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2^T}\Big|_{\boldsymbol{v}_2^*} = n^{-1}\mathbf{M}_2\mathbf{X}^T\mathbf{Y} - \boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{M}_3,$$
$$\frac{\partial \widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{v}_1^*,\boldsymbol{u}_2,\boldsymbol{v}_2^*)}{\partial \boldsymbol{u}_2^T}\Big|_{\boldsymbol{u}_2^*} = -\mathbf{M}_2\widehat{\boldsymbol{\Sigma}} - 2\mathbf{M}_3\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}} + n^{-1}\mathbf{M}_3\mathbf{Y}^T\mathbf{X}.$$

Then plugging them into (ER.85) entails that

$$\begin{aligned}\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{\eta}_1) = {}&\widetilde{\psi}_1(\boldsymbol{u}_1,\boldsymbol{\eta}_1^*) + (-n^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\mathbf{M}_1)(\mathbf{I}_q - \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T})\boldsymbol{\xi}_1\\ &+ (n^{-1}\mathbf{M}_2\mathbf{X}^T\mathbf{Y} - \boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{M}_3)(\mathbf{I}_q - \boldsymbol{v}_2^*\boldsymbol{v}_2^{*T})\boldsymbol{\xi}_2\\ &+ (-\mathbf{M}_2\widehat{\boldsymbol{\Sigma}} - 2\mathbf{M}_3\boldsymbol{v}_2^*\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}} + n^{-1}\mathbf{M}_3\mathbf{Y}^T\mathbf{X})(\boldsymbol{u}_2 - \boldsymbol{u}_2^*) + \boldsymbol{r}_{\boldsymbol{v}_1^*} + \boldsymbol{r}_{\boldsymbol{u}_2^*} + \boldsymbol{r}_{\boldsymbol{v}_2^*}.\end{aligned}$$

Since $\boldsymbol{\xi}_1 = \exp^{-1}_{\boldsymbol{v}_1^*}(\boldsymbol{v}_1)$ and $\boldsymbol{\xi}_2 = \exp^{-1}_{\boldsymbol{v}_2^*}(\boldsymbol{v}_2)$, this concludes the proof of Lemma EC.16.

ER.1.17. Proof of Lemma EC.17

We first prove part (a). In view of Definition 2, it holds that

$$\|\widetilde{d}_k\widetilde{\boldsymbol{v}}_k - d_k^*\boldsymbol{v}_k^*\|_2 \le c\gamma_n \ \text{ and } \ |d_k^* - \widetilde{d}_k| \le c\gamma_n,$$

where $\gamma_n = (r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}$. Observe that $d_k^*(\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*) = (\widetilde{d}_k\widetilde{\boldsymbol{v}}_k - d_k^*\boldsymbol{v}_k^*) + (d_k^* - \widetilde{d}_k)\widetilde{\boldsymbol{v}}_k$. Since $\|\widetilde{\boldsymbol{v}}_k\|_2 = 1$, we have that

$$\|d_k^*(\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*)\|_2 \le \|\widetilde{d}_k\widetilde{\boldsymbol{v}}_k - d_k^*\boldsymbol{v}_k^*\|_2 + |d_k^* - \widetilde{d}_k|\|\widetilde{\boldsymbol{v}}_k\|_2 \le c\gamma_n.$$

Since the true singular values $d_k^* \neq 0$ for each $k = 1, \cdots, r^*$, it follows that

$$\|\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*\|_2 \le c\gamma_n/d_k^*.$$

Also, by Definition 2 it holds that

$$\begin{aligned}\sum_{k=1}^{r^*}\|\widetilde{d}_k(\widetilde{\boldsymbol{v}}_k - \boldsymbol{v}_k^*)\|_0 &\le \sum_{k=1}^{r^*}\|\widetilde{d}_k\widetilde{\boldsymbol{v}}_k - d_k^*\boldsymbol{v}_k^*\|_0 + \sum_{k=1}^{r^*}\|(d_k^* - \widetilde{d}_k)\boldsymbol{v}_k^*\|_0\\ &\le \sum_{k=1}^{r^*}\|\widetilde{d}_k\widetilde{\boldsymbol{v}}_k - d_k^*\boldsymbol{v}_k^*\|_0 + \sum_{k=1}^{r^*}\|\boldsymbol{v}_k^*\|_0\\ &\le (r^* + s_u + s_v)[1 + o(1)] + s_v\\ &\le 3(r^* + s_u + s_v).\end{aligned}$$

For part (b), let us recall that $\boldsymbol{u}_k^* = d_k^*\boldsymbol{l}_k^*$. Since $\|\boldsymbol{l}_k^*\|_0 \le s_u$ and $\|\boldsymbol{l}_k^*\|_2 = 1$, it follows from Condition 2 that

$$\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_k^*\|_2 = d_k^*\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_k^*\|_2 \le \rho_u d_k^*\|\boldsymbol{l}_k^*\|_2 \le cd_k^*.$$

From Definition 2, we can show that

$$\|\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*\|_0 \le (r^* + s_u + s_v)[1 + o(1)] \text{ and } \|\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*\|_2 \le c\gamma_n.$$

Moreover, by $\|\boldsymbol{u}_k^*\|_0 \le s_u$ and $\|\boldsymbol{u}_k^*\|_2 \le d_k^*$ for sufficiently large n, and Condition 3 that $r^*\gamma_n = o(d_{r^*}^*)$, it holds for $\widetilde{\boldsymbol{u}}_k$ that

$$\begin{aligned}
&\|\widetilde{\boldsymbol{u}}_k\|_0 \le \|\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*\|_0 + \|\boldsymbol{u}_k^*\|_0 \le (r^* + s_u + s_v)[1 + o(1)] + s_u \le 3(r^* + s_u + s_v),\\
&\|\widetilde{\boldsymbol{u}}_k\|_2 \le \|\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*\|_2 + \|\boldsymbol{u}_k^*\|_2 \le c d_k^*.
\end{aligned}$$

Then it follows from Condition 2 that

$$\|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_k\|_2 \le \rho_u \|\widetilde{\boldsymbol{u}}_k\|_2 \le c d_k^* \text{ and } \|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)\|_2 \le \rho_u \|\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*\|_2 \le c\gamma_n.$$

For part (c), using part (b) of this lemma and Definition 2, we can deduce that

$$\begin{aligned}
|\widetilde{z}_{kk} - z_{kk}^*| &= |\widetilde{\boldsymbol{u}}_k^T \widehat{\boldsymbol{\Sigma}} \widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^*|\\
&\le |\widetilde{\boldsymbol{u}}_k^T \widehat{\boldsymbol{\Sigma}} (\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)| + |(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^*|\\
&\le \|\widetilde{\boldsymbol{u}}_k\|_2 \|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)\|_2 + \|\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*\|_2 \|\widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^*\|_2\\
&\le c\gamma_n d_k^*.
\end{aligned} \tag{ER.86}$$

Note that

$$|\widetilde{z}_{kk}^{-1} - z_{kk}^{*-1}| = \left| \frac{\widetilde{z}_{kk} - z_{kk}^*}{\widetilde{z}_{kk} \cdot z_{kk}^*} \right| = \left| \frac{\widetilde{\boldsymbol{u}}_k^T \widehat{\boldsymbol{\Sigma}} \widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^*}{\widetilde{\boldsymbol{u}}_k^T \widehat{\boldsymbol{\Sigma}} \widetilde{\boldsymbol{u}}_k \cdot \boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^*} \right|.$$

By Condition 2, we have that $d_k^{*2}\rho_l \le \boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^* \le d_k^{*2}\rho_u$. Then it follows that

$$|z_{kk}^{*-1}| = |\boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^*|^{-1} \le |d_k^{*-2}\rho_l| \le c d_k^{*-2}.$$

Together with (ER.86), it yields that

$$|\widetilde{z}_{kk}^{-1} - z_{kk}^{*-1}| \le \left| \frac{\widetilde{\boldsymbol{u}}_k^T \widehat{\boldsymbol{\Sigma}} \widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^*}{\boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^* (\boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_k^* + o(1))} \right| \le c\gamma_n d_k^{*-3}.$$

Furthermore, by $r^*\gamma_n = o(d_{r^*}^*)$ in Condition 3, we have for sufficiently large n it holds that

$$|\widetilde{z}_{kk}^{-1}| \le |z_{kk}^{*-1}| + |\widetilde{z}_{kk}^{-1} - z_{kk}^{*-1}| \le c d_k^{*-2}.$$

This concludes the proof of Lemma EC.17.

ER.1.18. Proof of Lemma EC.18

We will first show that $|z_{ii}^* - z_{jj}^*| \geq c$. By Condition 2 and the sparsity of $\boldsymbol{u}_i^* = d_i^* \boldsymbol{l}_i^*$, we see that $d_i^{*2}\rho_l \leq z_{ii}^* \leq d_i^{*2}\rho_u$ and $d_j^{*2}\rho_l \leq z_{jj}^* \leq d_j^{*2}\rho_u$, which lead to

$$d_i^{*2}\rho_l - d_j^{*2}\rho_u \leq z_{ii}^* - z_{jj}^* \leq d_i^{*2}\rho_u - d_j^{*2}\rho_l. \tag{ER.87}$$

In light of Condition 3, we have $d_i^{*2} - d_{i+1}^{*2} \geq \delta_1 d_i^{*2}$ for some positive constant $\delta_1 > 1 - (\rho_l/\rho_u)$ with $1 \leq i \leq r^*$. Since ρ_l, ρ_u are positive constants, there exists some positive constant c_0 such that $\delta_1 = 1 - (\rho_l/\rho_u) + c_0$, which further entails that

$$d_i^{*2}\rho_l - d_{i+1}^{*2}\rho_u \geq c_0 \rho_u d_i^{*2} \geq c, \tag{ER.88}$$

where the last inequality above is due to Condition 4 that d_i^* is at a constant level.

If $i < j$, we have $i + 1 \leq j$ so that $d_{i+1}^{*2} \geq d_j^{*2}$. This together with (ER.87) and (ER.88) shows that

$$z_{ii}^* - z_{jj}^* \geq d_i^{*2}\rho_l - d_j^{*2}\rho_u \geq d_i^{*2}\rho_l - d_{i+1}^{*2}\rho_u \geq c.$$

If $i > j$, using similar arguments we can obtain that $z_{jj}^* - z_{ii}^* \geq c$. Thus, for $i \neq j$ it holds that

$$|z_{ii}^* - z_{jj}^*| \geq c. \tag{ER.89}$$

We next bound term $\widetilde{z}_{ii} - \widetilde{z}_{jj}$ above. By part (c) of Lemma EC.17 and Condition 4 that d_i^* is at a constant level, we can deduce that

$$|(\widetilde{z}_{ii} - \widetilde{z}_{jj}) - (z_{ii}^* - z_{jj}^*)| \leq |\widetilde{z}_{ii} - z_{ii}^*| + |\widetilde{z}_{jj} - z_{jj}^*| \leq c\gamma_n,$$

where $\gamma_n = (r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}$. From the assumption of Theorem 1 that $m^{1/2}\kappa_n = o(1)$, we have $\gamma_n = o(1)$. Together with (ER.89), for all sufficiently large n it follows that

$$|\widetilde{z}_{ii} - \widetilde{z}_{jj}| \geq |z_{ii}^* - z_{jj}^*| - |(\widetilde{z}_{ii} - \widetilde{z}_{jj}) - (z_{ii}^* - z_{jj}^*)| \geq c. \tag{ER.90}$$

Now we analyze terms $\sum_{1 \leq l \leq r^*,\, l \neq k} |\widetilde{z}_{kl}|$ and $\sum_{1 \leq l \leq r^*,\, l \neq k} |z_{kl}^*|$. From Condition 4, we have that

$$\sum_{1 \leq l \leq r^*,\, l \neq k} |z_{kl}^*| = o(n^{-1/2}). \tag{ER.91}$$

Moreover, it follows that

$$\begin{aligned} |\widetilde{z}_{kl} - z_{kl}^*| &= |\widetilde{\boldsymbol{u}}_k^T \widehat{\boldsymbol{\Sigma}} \widetilde{\boldsymbol{u}}_l - \boldsymbol{u}_k^{*T} \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_l^*| \leq |\widetilde{\boldsymbol{u}}_k^T \widehat{\boldsymbol{\Sigma}} (\widetilde{\boldsymbol{u}}_l - \boldsymbol{u}_l^*)| + |(\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*)^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_l^*| \\ &\leq \|\widetilde{\boldsymbol{u}}_k\|_2 \|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_l - \boldsymbol{u}_l^*)\|_2 + \|\widetilde{\boldsymbol{u}}_k - \boldsymbol{u}_k^*\|_2 \|\widehat{\boldsymbol{\Sigma}} \boldsymbol{u}_l^*\|_2 \\ &\leq c(r^* + s_u + s_v)^{1/2}\{n^{-1}\log(pq)\}^{1/2}, \end{aligned} \tag{ER.92}$$

where the last inequality above is due to part (a) of Definition 2, part (b) of Lemma EC.17, and Condition 4 that d_k^* is at a constant level. Then for sufficiently large n, it holds that

$$|\widetilde{z}_{kl}| \leq |z_{kl}^*| + |\widetilde{z}_{kl} - z_{kl}^*| \leq c(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2},$$

which further yields $\sum_{1\leq l\leq r^*,\ l\neq k} |\widetilde{z}_{kl}| = O(r^*\gamma_n)$.

Let us recall the assumption that $m^{1/2}\kappa_n = o(1)$ with

$$\kappa_n = \max\{s_{\max}^{1/2}, (r^* + s_u + s_v)^{1/2}, \eta_n^2\}(r^* + s_u + s_v)\eta_n^2\log(pq)/\sqrt{n}.$$

It follows from $\gamma_n = (r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}$ and $(r^* + s_u + s_v)^{1/2} \leq \max\{s_{\max}^{1/2}, (r^* + s_u + s_v)^{1/2}, \eta_n^2\}$ that

$$m^{1/2}(r^* + s_u + s_v)\sqrt{\log(pq)}\gamma_n = o(1), \tag{ER.93}$$

which further leads to

$$\sum_{1\leq l\leq r^*,\ l\neq k} |\widetilde{z}_{kl}| = O(r^*\gamma_n) = o(1).$$

Therefore, along with (ER.89), (ER.90), and (ER.91), it yields that

$$\sum_{1\leq l\leq r^*,\ l\neq k} |z_{kl}^*| = o(|z_{ii}^* - z_{jj}^*|) \text{ and } \sum_{1\leq l\leq r^*,\ l\neq k} |\widetilde{z}_{kl}| = o(|\widetilde{z}_{ii} - \widetilde{z}_{jj}|),$$

which completes the proof of Lemma EC.18.

ER.1.19. Proof of Lemma EC.19

It is easy to see that

$$\begin{aligned}\|\boldsymbol{a}^T\mathbf{W}_1^*\|_2 &= \|\sum_{i=1}^p a_i\boldsymbol{w}_i^{*T}\|_2 \leq \sum_{i=1}^p |a_i|\cdot\|\boldsymbol{w}_i^*\|_2 \\ &\leq \|\boldsymbol{a}\|_1 \max_{1\leq i\leq p}\|\boldsymbol{w}_i^*\|_2 \leq \|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 \max_{1\leq i\leq p}\|\boldsymbol{w}_i^*\|_2.\end{aligned} \tag{ER.94}$$

Similarly, we also have that

$$\|\boldsymbol{a}^T\widetilde{\mathbf{W}}_1\|_2 \leq \|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 \max_{1\leq i\leq p}\|\widetilde{\boldsymbol{w}}_i\|_2, \tag{ER.95}$$

$$\|\boldsymbol{a}^T(\widetilde{\mathbf{W}}_1 - \mathbf{W}_1^*)\|_2 \leq \|\boldsymbol{a}\|_0^{1/2}\|\boldsymbol{a}\|_2 \max_{1\leq i\leq p}\|\widetilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^*\|_2. \tag{ER.96}$$

Then it can be seen that once parts (a) and (b) of this lemma are established, the results in part (c) can be obtained immediately with the aid of (ER.94)–(ER.96). Thus, it remains to prove parts (a) and (b).

We begin with proving part (a). Since $\boldsymbol{w}_i^{*T}$ is the ith row of $\mathbf{W}_1^*$, we have $\boldsymbol{w}_i^{*T} = \widehat{\boldsymbol{\theta}}_i^T\{\mathbf{I}_p + (z_{11}^* - z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\}$, where $\widehat{\boldsymbol{\theta}}_i^T$ is the ith row of $\widehat{\boldsymbol{\Theta}}$. Noting that $\widehat{\boldsymbol{\theta}}_i^T(z_{11}^* - z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*$ is a scalar and $\|\boldsymbol{u}_2^*\|_0 \le s_u$, we can deduce that

$$\max_{1\le i\le p}\|(\widehat{\boldsymbol{\theta}}_i^T(z_{11}^* - z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*)\cdot\boldsymbol{u}_2^{*T}\|_0 \le \|\boldsymbol{u}_2^*\|_0 \le s_u \le r^* + s_u + s_v.$$

From Definition 1, we see that $\max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_0 \le s_{max}$. Hence, it follows that

$$\max_{1\le i\le p}\|\boldsymbol{w}_i^*\|_0 \le \max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_0 + \|\boldsymbol{u}_2^*\|_0 \le 2\max\{s_{\max}, r^* + s_u + s_v\}. \tag{ER.97}$$

Observe that $\widetilde{\boldsymbol{w}}_i^T = \widehat{\boldsymbol{\theta}}_i^T\{\mathbf{I}_p + (\widetilde{z}_{11} - \widetilde{z}_{22})^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{u}}_2^T\}$. By Definition 2, we have that

$$\|\widetilde{\boldsymbol{u}}_2\|_0 \le \|\boldsymbol{u}_2^*\|_0 + \|\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*\|_0 \le 3(r^* + s_u + s_v).$$

Then an application of similar arguments as for (ER.97) leads to

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i\|_0 \le \max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_0 + \|\widetilde{\boldsymbol{u}}_2\|_0 \le 2\max\{s_{\max}, 3(r^* + s_u + s_v)\}.$$

Further, we can show that

$$\begin{aligned}\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^*\|_0 &\le \max_{1\le i\le p}\|(\widehat{\boldsymbol{\theta}}_i^T(\widetilde{z}_{11} - \widetilde{z}_{22})^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2)\cdot\widetilde{\boldsymbol{u}}_2^T - (\widehat{\boldsymbol{\theta}}_i^T(z_{11}^* - z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*)\cdot\boldsymbol{u}_2^{*T}\|_0\\ &\le \max_{1\le i\le p}\|(\widehat{\boldsymbol{\theta}}_i^T(\widetilde{z}_{11} - \widetilde{z}_{22})^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2)\cdot(\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*)^T\|_0 + \|\boldsymbol{u}_2^{*T}\|_0\\ &\le \|\widetilde{\boldsymbol{u}}_2 - \boldsymbol{u}_2^*\|_0 + \|\boldsymbol{u}_2^*\|_0 \le 3(r^* + s_u + s_v),\end{aligned}$$

where the last step above is due to Definition 2. This completes the proof for part (a).

We next show part (b), which consists of two main steps. Since Condition 4 is satisfied, the proof below will exploit the fact that the nonzero eigenvalues d_i^{*2} are at the constant level.

(1). The upper bound on $\max_{1\le i\le p}\|\boldsymbol{w}_i^*\|_2$. Let us recall that

$$\boldsymbol{w}_i^{*T} = \widehat{\boldsymbol{\theta}}_i^T\{\mathbf{I}_p + (z_{11}^* - z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\}.$$

Under Condition 2, it follows from part (b) of Lemma EC.17 and $\|\boldsymbol{u}_2^*\|_2 = d_2^* \le c$ that

$$\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2 \le \|\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\|_2\|\boldsymbol{u}_2^{*T}\|_2 \le c.$$

Also, under Conditions 2–4, Lemma EC.18 gives that $|z_{11}^* - z_{22}^*| \ge c$. Then we can obtain that

$$\|(z_{11}^* - z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2 \le |z_{11}^* - z_{22}^*|^{-1}\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2 \le c.$$

Together with Definition 1 that $\max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_2 \le c$, it yields that

$$\max_{1\le i\le p}\|\boldsymbol{w}_i^*\|_2 \le \max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_2 + \max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_2\|(z_{11}^* - z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2 \le c. \tag{ER.98}$$

(2). The upper bounds on $\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i-\boldsymbol{w}_i^*\|_2$ **and** $\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i\|_2$. From Definition 1 that $\max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_2\le c$, we have that

$$\begin{aligned}\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i-\boldsymbol{w}_i^*\|_2 &\le \max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_2\|(\widetilde{z}_{11}-\widetilde{z}_{22})^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{u}}_2^T-(z_{11}^*-z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2\\ &\le c\|(\widetilde{z}_{11}-\widetilde{z}_{22})^{-1}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{u}}_2^T-(z_{11}^*-z_{22}^*)^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2\\ &\le c\|(\widetilde{z}_{11}-\widetilde{z}_{22})^{-1}(\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{u}}_2^T-\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T})\|_2\\ &\quad+c\|[(\widetilde{z}_{11}-\widetilde{z}_{22})^{-1}-(z_{11}^*-z_{22}^*)^{-1}]\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2.\end{aligned}$$

We will bound the two terms introduced above separately. It follows from part (b) of Lemma EC.17 and part (a) of Definition 1 that

$$\begin{aligned}\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2&\le\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\|_2\|\boldsymbol{u}_2^{*T}\|_2\le c,\\ \|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2\widetilde{\boldsymbol{u}}_2^T-\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2^*\boldsymbol{u}_2^{*T}\|_2&\le\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{u}}_2-\boldsymbol{u}_2^*)\boldsymbol{u}_2^{*T}\|_2+\|\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_2(\widetilde{\boldsymbol{u}}_2-\boldsymbol{u}_2^*)^T\|_2\\ &\le c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned}$$

Lemma EC.18 implies that $|z_{11}^*-z_{22}^*|\ge c$ and $|\widetilde{z}_{11}-\widetilde{z}_{22}|\ge c$. Further, it holds that

$$\begin{aligned}&|(\widetilde{z}_{11}-\widetilde{z}_{22})^{-1}-(z_{11}^*-z_{22}^*)^{-1}|\\ &=\left|\frac{1}{z_{11}^*-z_{22}^*}\cdot\frac{(\widetilde{z}_{11}-\widetilde{z}_{22})-(z_{11}^*-z_{22}^*)}{(z_{11}^*-z_{22}^*)+(\widetilde{z}_{11}-\widetilde{z}_{22})-(z_{11}^*-z_{22}^*)}\right|.\end{aligned}$$

In view of part (c) of Lemma EC.17, we have that

$$\begin{aligned}|(\widetilde{z}_{11}-\widetilde{z}_{22})-(z_{11}^*-z_{22}^*)|&\le|\widetilde{z}_{11}-z_{11}^*|+|\widetilde{z}_{22}-z_{22}^*|\\ &\le c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned}$$

Together with $|z_{11}^*-z_{22}^*|\ge c$, for sufficiently large n it holds that

$$\begin{aligned}|(\widetilde{z}_{11}-\widetilde{z}_{22})^{-1}-(z_{11}^*-z_{22}^*)^{-1}|&=\left|\frac{(\widetilde{z}_{11}-\widetilde{z}_{22})-(z_{11}^*-z_{22}^*)}{(z_{11}^*-z_{22}^*)^2+o((z_{11}^*-z_{22}^*)^2)}\right|\\ &\le c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned}$$

Combining the above results gives that

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i-\boldsymbol{w}_i^*\|_2\le c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}. \tag{ER.99}$$

Therefore, using (ER.98), (ER.99), and the triangle inequality, we can obtain that for all sufficiently large n,

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i\|_2\le\max_{1\le i\le p}\|\boldsymbol{w}_i^*\|_2+\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i-\boldsymbol{w}_i^*\|_2\le c,$$

which concludes the proof of Lemma EC.19.

ER.1.20. Proof of Lemma EC.20

We will first analyze matrix $\mathbf{I}_{r^*-1} - \widetilde{z}_{kk}^{-1}\widetilde{\mathbf{U}}_{-k}^T\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}$, which is equivalent to analyzing the nonsingularity of matrix $\mathbf{A} =: \widetilde{z}_{kk}\mathbf{I}_{r^*-1} - \widetilde{\mathbf{U}}_{-k}^T\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}$. For simplicity, denote by $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{(r^*-1)\times(r^*-1)}$ with $i,j \in \mathcal{A} = \{1 \le \ell \le r^* : \ell \ne k\}$. It can be seen that for each $i,j \in \mathcal{A}$,

$$a_{ij} = \begin{cases} \widetilde{z}_{kk} - \widetilde{z}_{ii} & \text{if } i = j, \\ -\widetilde{z}_{ij} & \text{if } i \ne j. \end{cases} \tag{ER.100}$$

From Lemma EC.18, we have that $\sum_{j\in\mathcal{A},\, j\ne i} |a_{ij}| = o(|a_{\ell\ell}|)$ for any $i,\ell \in \mathcal{A}$. Then it holds that

$$|a_{ii}| > \sum_{j\in\mathcal{A},\, j\ne i} |a_{ij}| \ \text{ for all } i \in \mathcal{A},$$

which shows that $\mathbf{A}$ is strictly diagonally dominant. Using the Levy–Desplanques Theorem in Horn and Johnson (2012), we see that matrix $\mathbf{A}$ is nonsingular, which entails that $\mathbf{I}_{r^*-1} - \widetilde{z}_{kk}^{-1}\widetilde{\mathbf{U}}_{-k}^T\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}$ is nonsingular. Moreover, with similar arguments we can also show that $\mathbf{I}_{r^*-1} - z_{kk}^{*-1}\mathbf{U}_{-k}^{*T}\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*$ is strictly diagonally dominant and thus is nonsingular. Therefore, we see that both $\widetilde{\mathbf{W}}_k$ and $\mathbf{W}_k^*$ are well-defined and satisfy the property in Proposition 3, which concludes the proof of Lemma EC.20.

ER.1.21. Proof of Lemma EC.21

Let us first bound terms $\|\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\|_2$ and $\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\|_2$. By definition, it holds that $\|\mathbf{U}_{-k}^*\|_0 \le \|\mathbf{U}^*\|_0 = s_u$. For any vector $\mathbf{x} \in \mathbb{R}^{r^*-1}$, we see that $\|\mathbf{U}_{-k}^*\mathbf{x}\|_0 \le s_u$. It follows from the definition of the induced 2-norm and Condition 2 that

$$\|\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\|_2 = \sup_{\mathbf{x}^T\mathbf{x}=1} \|\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\mathbf{x}\|_2 \le c \sup_{\mathbf{x}^T\mathbf{x}=1} \|\mathbf{U}_{-k}^*\mathbf{x}\|_2 \le c\|\mathbf{U}_{-k}^*\|_2. \tag{ER.101}$$

Since $\mathbf{U}_{-k}^{*T}\mathbf{U}_{-k}^* = \mathbf{D}_{-k}^{*2}$ with $\mathbf{D}_{-k}^{*2} = \mathrm{diag}\{d_1^{*2}, \cdots, d_{k-1}^{*2}, d_{k+1}^{*2}, \cdots, d_{r^*}^{*2}\}$, we can show that

$$\sup_{\mathbf{x}^T\mathbf{x}=1} \|\mathbf{U}_{-k}^*\mathbf{x}\|_2^2 = \sup_{\mathbf{x}^T\mathbf{x}=1} \mathbf{x}^T\mathbf{U}_{-k}^{*T}\mathbf{U}_{-k}^*\mathbf{x} = \sup_{\mathbf{x}^T\mathbf{x}=1} \mathbf{x}^T\mathbf{D}_{-k}^{*2}\mathbf{x} \le d_1^{*2}, \tag{ER.102}$$

which leads to $\|\mathbf{U}_{-k}^*\|_2 \le cd_1^*$. It also implies that $\|\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\|_2 \le cd_1^*$.

In view of Definition 2, we have that

$$\|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*\|_0 \le \|\widetilde{\mathbf{U}} - \mathbf{U}^*\|_0 \le 2(r^* + s_u + s_v),$$

$$\|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*\|_2 \le \|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*\|_F \le \|\widetilde{\mathbf{U}} - \mathbf{U}^*\|_F \le c\gamma_n.$$

Then using similar arguments as for (ER.101), we can deduce that

$$\|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*)\|_2 \le c\|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*\|_2 \le c\gamma_n.$$

Hence, for sufficiently large n it holds that

$$\|\widetilde{\mathbf{U}}_{-k}\|_2 \leq \|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}^*_{-k}\|_2 + \|\mathbf{U}^*_{-k}\|_2 \leq c d_1^*,$$
$$\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\|_2 \leq \|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{U}}_{-k} - \mathbf{U}^*_{-k})\|_2 + \|\widehat{\mathbf{\Sigma}}\mathbf{U}^*_{-k}\|_2 \leq c d_1^*.$$

For term $\|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}^*_{-k})\|_2$ above, it follows that

$$\begin{aligned}
&\|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}^*_{-k})\|_2 = \|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{U}}_{-k}\widetilde{\mathbf{V}}^T_{-k} - \mathbf{U}^*_{-k}\mathbf{V}^{*T}_{-k})\|_2 \\
&\leq \|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}(\widetilde{\mathbf{V}}_{-k} - \mathbf{V}^*_{-k})^T\|_2 + \|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{U}}_{-k} - \mathbf{U}^*_{-k})\mathbf{V}^{*T}_{-k}\|_2 \\
&\leq \|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{L}}_{-k}\|_2\|\widetilde{\mathbf{D}}_{-k}(\widetilde{\mathbf{V}}_{-k} - \mathbf{V}^*_{-k})^T\|_2 + \|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{U}}_{-k} - \mathbf{U}^*_{-k})\|_2\|\mathbf{V}^{*T}_{-k}\|_2.
\end{aligned} \tag{ER.103}$$

Note that $\widetilde{\mathbf{L}}^T_{-k}\widetilde{\mathbf{L}}_{-k} = \mathbf{I}$. An application of similar arguments as for (ER.101) and (ER.102) leads to

$$\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{L}}_{-k}\|_2 \leq c.$$

For term $\|(\widetilde{\mathbf{V}}_{-k} - \mathbf{V}^*_{-k})\widetilde{\mathbf{D}}_{-k}\|_2$ above, we can deduce that

$$\begin{aligned}
\|(\widetilde{\mathbf{V}}_{-k} - \mathbf{V}^*_{-k})\widetilde{\mathbf{D}}_{-k}\|_2 &\leq \|\widetilde{\mathbf{V}}_{-k}\widetilde{\mathbf{D}}_{-k} - \mathbf{V}^*_{-k}\mathbf{D}^*_{-k}\|_2 + \|\widetilde{\mathbf{V}}_{-k}\|_2\|\widetilde{\mathbf{D}}_{-k} - \mathbf{D}^*_{-k}\|_2 \\
&\leq \|\widetilde{\mathbf{V}}_{-k}\widetilde{\mathbf{D}}_{-k} - \mathbf{V}^*_{-k}\mathbf{D}^*_{-k}\|_F + \|\widetilde{\mathbf{D}}_{-k} - \mathbf{D}^*_{-k}\|_F \\
&\leq \|\widetilde{\mathbf{V}}\widetilde{\mathbf{D}} - \mathbf{V}^*\mathbf{D}^*\|_F + \|\widetilde{\mathbf{D}} - \mathbf{D}^*\|_F \leq c\gamma_n,
\end{aligned}$$

where we have used Definition 2 and $\|\widetilde{\mathbf{V}}_{-k}\|_2 = 1$. Along with $\|\mathbf{V}^*_{-k}\|_2 = 1$ and $\|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{U}}_{-k} - \mathbf{U}^*_{-k})\|_2 \leq c\gamma_n$, it yields that

$$\|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}^*_{-k})\|_2 \leq c\gamma_n. \tag{ER.104}$$

Further, using similar arguments we can obtain that

$$\begin{aligned}
\|\widetilde{\mathbf{C}}_{-k} - \mathbf{C}^*_{-k}\|_2 &= \|\widetilde{\mathbf{U}}_{-k}\widetilde{\mathbf{V}}^T_{-k} - \mathbf{U}^*_{-k}\mathbf{V}^{*T}_{-k}\|_2 \\
&\leq \|\widetilde{\mathbf{U}}_{-k}(\widetilde{\mathbf{V}}_{-k} - \mathbf{V}^*_{-k})^T\|_2 + \|(\widetilde{\mathbf{U}}_{-k} - \mathbf{U}^*_{-k})\mathbf{V}^{*T}_{-k}\|_2 \\
&\leq \|\widetilde{\mathbf{L}}_{-k}\|_2\|\widetilde{\mathbf{D}}_{-k}(\widetilde{\mathbf{V}}_{-k} - \mathbf{V}^*_{-k})^T\|_2 + \|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}^*_{-k}\|_2\|\mathbf{V}^{*T}_{-k}\|_2. \\
&\leq c\gamma_n.
\end{aligned} \tag{ER.105}$$

Observe that $\widetilde{\mathbf{M}}_k = -\widetilde{z}^{-1}_{kk}\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\widetilde{\mathbf{V}}^T_{-k}$ and $\mathbf{M}^*_k = -z^{*-1}_{kk}\widehat{\mathbf{\Sigma}}\mathbf{U}^*_{-k}\mathbf{V}^{*T}_{-k}$. For $\widetilde{\mathbf{M}}_k = -\widetilde{z}^{-1}_{kk}\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{C}}_{-k}$, it holds that

$$\begin{aligned}
\|\widetilde{\mathbf{M}}_k\|_2 &\leq |\widetilde{z}^{-1}_{kk}|\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\widetilde{\mathbf{V}}^T_{-k}\|_2 \leq |\widetilde{z}^{-1}_{kk}|\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\|_2\|\widetilde{\mathbf{V}}^T_{-k}\|_2 \\
&\leq c d_k^{*-2} d_1^*,
\end{aligned}$$

where we have used part (c) of Lemma EC.17, $\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\|_2 \leq cd_1^*$, and $\|\widetilde{\mathbf{V}}_k\|_2 = 1$. With the aid of similar arguments, we can show that

$$\|\mathbf{M}_k^*\|_2 \leq cd_k^{*-2} d_1^*.$$

For term $\|\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*\|_2$, it follows from part (c) of Lemma EC.17, (ER.104), $\|\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\|_2 \leq cd_{k+1}^*$, and $\|\mathbf{V}_{-k}^*\|_2 = 1$ that

$$\begin{aligned}
\|\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*\|_2 &\leq |\widetilde{z}_{kk}^{-1} - z_{kk}^{*-1}| \|\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\|_2 \|(\mathbf{V}_{-k}^*)^T\|_2 + |z_{kk}^{*-1}| \|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{C}}_{-k} - \mathbf{C}_{-k}^*)\|_2 \\
&\leq c\gamma_n d_1^* d_k^{*-3}.
\end{aligned}$$

This completes the proof of Lemma EC.21.

ER.1.22. Proof of Lemma EC.22

Similar to the proof of Lemma EC.19, with the aid of (ER.94)–(ER.96) we can obtain immediately the results in parts (c)–(e) once the results in parts (a) and (b) are shown. Hence, it remains to establish parts (a) and (b). We start with proving part (a). Let us recall that

$$\begin{aligned}
\widetilde{\mathbf{W}}_k &= \widehat{\mathbf{\Theta}} \left\{ \mathbf{I}_p + \widetilde{z}_{kk}^{-1} \widehat{\mathbf{\Sigma}} \widetilde{\mathbf{U}}_{-k} (\mathbf{I}_{r^*-1} - \widetilde{z}_{kk}^{-1} \widetilde{\mathbf{U}}_{-k}^T \widehat{\mathbf{\Sigma}} \widetilde{\mathbf{U}}_{-k})^{-1} \widetilde{\mathbf{U}}_{-k}^T \right\}, \\
\mathbf{W}_k^* &= \widehat{\mathbf{\Theta}} \left\{ \mathbf{I}_p + z_{kk}^{*-1} \widehat{\mathbf{\Sigma}} \mathbf{U}_{-k}^* (\mathbf{I}_{r^*-1} - z_{kk}^{*-1} \mathbf{U}_{-k}^{*T} \widehat{\mathbf{\Sigma}} \mathbf{U}_{-k}^*)^{-1} \mathbf{U}_{-k}^{*T} \right\}.
\end{aligned}$$

Denote by $\widetilde{\mathbf{A}} = (\widetilde{z}_{kk} \mathbf{I}_{r^*-1} - \widetilde{\mathbf{U}}_{-k}^T \widehat{\mathbf{\Sigma}} \widetilde{\mathbf{U}}_{-k})^{-1}$ and $\mathbf{A}^* = (z_{kk}^* \mathbf{I}_{r^*-1} - \mathbf{U}_{-k}^{*T} \widehat{\mathbf{\Sigma}} \mathbf{U}_{-k}^*)^{-1}$. Then it holds that

$$\boldsymbol{w}_i^T = \widehat{\boldsymbol{\theta}}_i^T (\mathbf{I}_p + \widehat{\mathbf{\Sigma}} \widetilde{\mathbf{U}}_{-k} \widetilde{\mathbf{A}} \widetilde{\mathbf{U}}_{-k}^T) \text{ and } \boldsymbol{w}_i^{*T} = \widehat{\boldsymbol{\theta}}_i^T (\mathbf{I}_p + \widehat{\mathbf{\Sigma}} \mathbf{U}_{-k}^* \mathbf{A}^* \mathbf{U}_{-k}^{*T}),$$

where $\widehat{\boldsymbol{\theta}}_i^T$ represents the ith row of $\widehat{\mathbf{\Theta}}$. Let $\boldsymbol{\delta}_i^T = \widehat{\boldsymbol{\theta}}_i^T \widehat{\mathbf{\Sigma}} \widetilde{\mathbf{U}}_{-k} \widetilde{\mathbf{A}}$ with $\boldsymbol{\delta}_i = (\delta_{ij})$. It follows from $\|\mathbf{U}^*\|_0 = s_u$ and part (b) of Definition 2 that

$$\begin{aligned}
&\max_{1 \leq i \leq p} \|\widehat{\boldsymbol{\theta}}_i^T \widehat{\mathbf{\Sigma}} \widetilde{\mathbf{U}}_{-k} \widetilde{\mathbf{A}} \widetilde{\mathbf{U}}_{-k}^T\|_0 = \max_{1 \leq i \leq p} \|\boldsymbol{\delta}_i^T \widetilde{\mathbf{U}}_{-k}^T\|_0 = \max_{1 \leq i \leq p} \| \sum_{\substack{1 \leq j \leq r^* \\ j \neq k}} \delta_{ij} \widetilde{\boldsymbol{u}}_j^T \|_0 \\
&\leq \sum_{\substack{1 \leq j \leq r^* \\ j \neq k}} \|\widetilde{\boldsymbol{u}}_j\|_0 \leq \|\widetilde{\mathbf{U}}\|_0 \leq \|\widetilde{\mathbf{U}} - \mathbf{U}^*\|_0 + \|\mathbf{U}^*\|_0 \\
&\leq 2(r^* + s_u + s_v) + s_u \leq 3(r^* + s_u + s_v). \qquad \text{(ER.106)}
\end{aligned}$$

Also, by Definition 1 we see that $\max_{1 \leq i \leq p} \|\widehat{\boldsymbol{\theta}}_i\|_0 \leq s_{max}$. Thus, it holds that

$$\begin{aligned}
\max_{1 \leq i \leq p} \|\widetilde{\boldsymbol{w}}_i\|_0 &\leq \max_{1 \leq i \leq p} \|\widehat{\boldsymbol{\theta}}_i\|_0 + \max_{1 \leq i \leq p} \|\widehat{\boldsymbol{\theta}}_i^T \widehat{\mathbf{\Sigma}} \widetilde{\mathbf{U}}_{-k} \widetilde{\mathbf{A}} \widetilde{\mathbf{U}}_{-k}^T\|_0 \\
&\leq 2 \max\{s_{\max}, 3(r^* + s_u + s_v)\}.
\end{aligned}$$

Similar to (ER.106), we can show that

$$\max_{1\le i\le p} \|\widehat{\boldsymbol\theta}_i^T \widehat{\boldsymbol\Sigma}\mathbf{U}_{-k}^*\mathbf{A}^*\mathbf{U}_{-k}^{*T}\|_0 \le \|\mathbf{U}^*\|_0 \le s_u \le r^* + s_u + s_v. \tag{ER.107}$$

It follows that

$$\max_{1\le i\le p} \|\boldsymbol w_i^*\|_0 \le \max_{1\le i\le p} \|\widehat{\boldsymbol\theta}_i\|_0 + \|\mathbf{U}^*\|_0 \le 2\max\{s_{\max}, (r^* + s_u + s_v)\}.$$

Let us further denote by $\boldsymbol\delta_i^{*T} = \widehat{\boldsymbol\theta}_i^T \widehat{\boldsymbol\Sigma}\mathbf{U}_{-k}^*\mathbf{A}^*$. Then using similar arguments as for (ER.106) and (ER.107), we can deduce that

$$\begin{aligned}\max_{1\le i\le p} \|\widetilde{\boldsymbol w}_i - \boldsymbol w_i^*\|_0 &\le \max_{1\le i\le p} \|\widehat{\boldsymbol\theta}_i^T \widehat{\boldsymbol\Sigma}\widetilde{\mathbf{U}}_{-k}\widetilde{\mathbf{A}}\widetilde{\mathbf{U}}_{-k}^T - \widehat{\boldsymbol\theta}_i^T \widehat{\boldsymbol\Sigma}\mathbf{U}_{-k}^*\mathbf{A}^*\mathbf{U}_{-k}^{*T}\|_0 \\ &\le \max_{1\le i\le p} \|\boldsymbol\delta_i^T (\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*)^T + (\boldsymbol\delta_i - \boldsymbol\delta_i^*)^T \mathbf{U}_{-k}^{*T}\|_0 \\ &\le \max_{1\le i\le p} \|\boldsymbol\delta_i^T (\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*)^T\|_0 + \max_{1\le i\le p} \|(\boldsymbol\delta_i - \boldsymbol\delta_i^*)^T \mathbf{U}_{-k}^{*T}\|_0 \\ &\le \|\widetilde{\mathbf{U}} - \mathbf{U}^*\|_0 + \|\mathbf{U}^*\|_0 \le 3(r^* + s_u + s_v),\end{aligned}$$

which completes the proof for part (a).

We next proceed with proving part (b), which will consist of two parts.

(1). The upper bound on $\max_{1\le i\le p}\|\boldsymbol w_i^*\|_2$. In light of Definition 1 that $\max_{1\le i\le p}\|\widehat{\boldsymbol\theta}_i\|_2 \le c$, it holds that

$$\begin{aligned}\max_{1\le i\le p} \|\boldsymbol w_i^*\|_2 &\le \max_{1\le i\le p} \|\widehat{\boldsymbol\theta}_i\|_2 + \max_{1\le i\le p} \|\widehat{\boldsymbol\theta}_i \widehat{\boldsymbol\Sigma}\mathbf{U}_{-k}^*\mathbf{A}^*\mathbf{U}_{-k}^{*T}\|_2 \\ &\le c(1 + \|\widehat{\boldsymbol\Sigma}\mathbf{U}_{-k}^*\|_2 \|\mathbf{A}^*\|_2 \|\mathbf{U}_{-k}^{*T}\|_2).\end{aligned} \tag{ER.108}$$

We will bound term $\|\mathbf{A}^*\|_2$. Denote by $\mathbf{A}_0 = (\mathbf{A}^*)^{-1} = z_{kk}^*\mathbf{I}_{r^*-1} - \mathbf{U}_{-k}^{*T}\widehat{\boldsymbol\Sigma}\mathbf{U}_{-k}^*$. Then using the technical arguments in the proof of Lemma EC.20 in Section ER.1.20, we can see that $\mathbf{A}_0 = z_{kk}^*\mathbf{I}_{r^*-1} - \mathbf{U}_{-k}^{*T}\widehat{\boldsymbol\Sigma}\mathbf{U}_{-k}^* = (a_{ij})$ is symmetric and strictly diagonally dominant, and $a_{ii} = z_{kk}^* - z_{ii}^*$, $a_{ij} = -z_{ij}^*$ for each $i \ne j$. Moreover, we have that $\sum_{j\ne i}|a_{ij}| = o(|a_{ii}|)$. Let us define

$$\alpha_1 = \min_i(|a_{ii}| - \sum_{j\ne i}|a_{ij}|) \text{ and } \alpha_2 = \min_i(|a_{ii}| - \sum_{j\ne i}|a_{ji}|).$$

Then it holds that $\alpha_1 = \alpha_2 \asymp \min_i |a_{ii}| = \min_{i\ne k}|z_{kk}^* - z_{ii}^*|$. It follows from (ER.87) and (ER.88) in the proof of Lemma EC.18 that for each $i \in \{1, \cdots, r^*\}$,

$$z_{ii}^* - z_{i+1,i+1}^* \ge d_i^{*2}\rho_l - d_{i+1}^{*2}\rho_u \ge c_0\rho_u d_i^{*2} \ge c,$$

which entails that $z_{ii}^* > z_{i+1,i+1}^*$. We can further see that

$$\min_{i\ne k}|z_{kk}^* - z_{ii}^*| = \min\{|z_{k-1,k-1}^* - z_{kk}^*|, |z_{kk}^* - z_{k+1,k+1}^*|\} \ge c.$$

Since $(\mathbf{A}^*)^{-1} = \mathbf{A}_0$ is symmetric and strictly diagonally dominant, an application of Corollary 2 in Varah (1975) leads to

$$\|\mathbf{A}^*\|_2 = \|\mathbf{A}_0^{-1}\|_2 \leq \frac{1}{\sqrt{\alpha_1\alpha_2}} \leq c(\min_{i\neq k}|z_{kk}^* - z_{ii}^*|)^{-1} \leq c. \tag{ER.109}$$

Also, in light of Lemma EC.21 we have that

$$\|\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\|_2 \leq cd_1^* \text{ and } \|\mathbf{U}_{-k}^*\|_2 \leq cd_1^*. \tag{ER.110}$$

Hence, under Condition 4 that the nonzero eigenvalues d_i^{*2} are at the constant level and Definition 1 that $\max_{1\leq i\leq p}\|\widehat{\boldsymbol{\theta}}_i\|_2 \leq c$, combining (ER.109)–(ER.110) yields that

$$\max_{1\leq i\leq p}\|\widehat{\boldsymbol{\theta}}_i\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\mathbf{A}^*\mathbf{U}_{-k}^{*T}\|_2 \leq \max_{1\leq i\leq p}\|\widehat{\boldsymbol{\theta}}_i\|_2\|\widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\|_2\|\mathbf{A}^*\|_2\|\mathbf{U}_{-k}^{*T}\|_2 \leq c, \tag{ER.111}$$

which further results in

$$\max_{1\leq i\leq p}\|\boldsymbol{w}_i^*\|_2 \leq c.$$

(2). The upper bounds on $\max_{1\leq i\leq p}\|\widetilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^*\|_2$ **and** $\max_{1\leq i\leq p}\|\widetilde{\boldsymbol{w}}_i\|_2$. For term $\max_{1\leq i\leq p}\|\widetilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^*\|_2$ above, in view of Definition 1 that $\max_{1\leq i\leq p}\|\widehat{\boldsymbol{\theta}}_i\|_2 \leq c$, it holds that

$$\begin{aligned}\max_{1\leq i\leq p}\|\widetilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^*\|_2 &\leq \max_{1\leq i\leq p}\|\widehat{\boldsymbol{\theta}}_i^T\|_2\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\widetilde{\mathbf{A}}\widetilde{\mathbf{U}}_{-k}^T - \widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\mathbf{A}^*\mathbf{U}_{-k}^{*T}\|_2 \\ &\leq c\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\widetilde{\mathbf{A}}\widetilde{\mathbf{U}}_{-k}^T - \widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\mathbf{A}^*\mathbf{U}_{-k}^{*T}\|_2.\end{aligned} \tag{ER.112}$$

Some simple calculations give that

$$\begin{aligned}&\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}\widetilde{\mathbf{A}}\widetilde{\mathbf{U}}_{-k}^T - \widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\mathbf{A}^*\mathbf{U}_{-k}^{*T} \\ &= (\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k} - \widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*)\widetilde{\mathbf{A}}\widetilde{\mathbf{U}}_{-k}^T + \widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*(\widetilde{\mathbf{A}} - \mathbf{A}^*)\widetilde{\mathbf{U}}_{-k}^T \\ &\quad + \widehat{\mathbf{\Sigma}}\mathbf{U}_{-k}^*\mathbf{A}^*(\widetilde{\mathbf{U}}_{-k}^T - \mathbf{U}_{-k}^{*T}).\end{aligned} \tag{ER.113}$$

We aim to bound the three terms on the right-hand side of (ER.113) above.

Recall that $\widetilde{\mathbf{A}} = (\widetilde{z}_{kk}\mathbf{I}_{r^*-1} - \widetilde{\mathbf{U}}_{-k}^T\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k})^{-1}$. Observe that $\widetilde{\mathbf{A}}_0 = \widetilde{\mathbf{A}}^{-1} = \widetilde{z}_{kk}\mathbf{I}_{r^*-1} - \widetilde{\mathbf{U}}_{-k}^T\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}_{-k}$ is also symmetric and strictly diagonally dominant from the proof of Lemma EC.20. Using similar arguments as for (ER.109), we can deduce that

$$\|\widetilde{\mathbf{A}}\|_2 = \|\widetilde{\mathbf{A}}_0^{-1}\|_2 \leq c. \tag{ER.114}$$

By Definition 2, we have that

$$\|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*\|_2 \leq \|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*\|_F \leq \|\widetilde{\mathbf{U}} - \mathbf{U}^*\|_F \leq c\gamma_n. \tag{ER.115}$$

An application of Lemma EC.21 leads to

$$\|\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}_{-k}\|_2 \le c d_1^*, \ \ \|\widetilde{\mathbf{U}}_{-k}\|_2 \le c d_1^*, \tag{ER.116}$$

$$\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*)\|_2 \le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}. \tag{ER.117}$$

Notice that Condition 4 implies that the nonzero eigenvalues d_i^{*2} are at the constant level. A combination of (ER.114)–(ER.117) yields that

$$\begin{aligned}\|(\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}_{-k} - \widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*)\widetilde{\mathbf{A}}\widetilde{\mathbf{U}}_{-k}^T\|_2 &\le \|\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*)\|_2\|\widetilde{\mathbf{A}}\|_2\|\widetilde{\mathbf{U}}_{-k}^T\|_2 \\ &\le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned} \tag{ER.118}$$

Moreover, it follows from (ER.109), (ER.110), and (ER.115) that

$$\begin{aligned}\|\widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*\mathbf{A}^*(\widetilde{\mathbf{U}}_{-k}^T - \mathbf{U}_{-k}^{*T})\|_2 &\le \|\widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*\|_2\|\mathbf{A}^*\|_2\|\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^*\|_2 \\ &\le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned} \tag{ER.119}$$

We proceed with bounding term $\|\widetilde{\mathbf{A}} - \mathbf{A}^*\|_2$ above. From (ER.109) and (ER.114), we see that $\|\mathbf{A}_0^{-1}\|_2 = \|\mathbf{A}^*\|_2 \le c$ and $\|\widetilde{\mathbf{A}}_0^{-1}\|_2 = \|\widetilde{\mathbf{A}}\|_2 \le c$. Then it holds that

$$\begin{aligned}\|\widetilde{\mathbf{A}} - \mathbf{A}^*\|_2 &= \|\widetilde{\mathbf{A}}_0^{-1} - \mathbf{A}_0^{-1}\|_2 = \|\widetilde{\mathbf{A}}_0^{-1}(\mathbf{A}_0 - \widetilde{\mathbf{A}}_0)\mathbf{A}_0^{-1}\|_2 \\ &\le \|\widetilde{\mathbf{A}}_0^{-1}\|_2\|\mathbf{A}_0 - \widetilde{\mathbf{A}}_0\|_2\|\mathbf{A}_0^{-1}\|_2 \le c\|\mathbf{A}_0 - \widetilde{\mathbf{A}}_0\|_2.\end{aligned} \tag{ER.120}$$

It remains to bound term $\|\mathbf{A}_0 - \widetilde{\mathbf{A}}_0\|_2$ above. Note that

$$\begin{aligned}\|\widetilde{\mathbf{A}}_0 - \mathbf{A}_0\|_2 &= \|(\widetilde{z}_{kk}\mathbf{I}_{r^*-1} - \widetilde{\mathbf{U}}_{-k}^T\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}_{-k}) - (z_{kk}^*\mathbf{I}_{r^*-1} - \mathbf{U}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*)\|_2 \\ &\le |\widetilde{z}_{kk} - z_{kk}^*| + \|\widetilde{\mathbf{U}}_{-k}^T\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*\|_2.\end{aligned}$$

In light of (ER.110), (ER.115), and (ER.116), we have that

$$\begin{aligned}&\|\widetilde{\mathbf{U}}_{-k}^T\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}_{-k} - \mathbf{U}_{-k}^{*T}\widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*\|_2 \le \|\widetilde{\mathbf{U}}_{-k}^T(\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}_{-k} - \widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*)\|_2 \\ &\quad + \|(\widetilde{\mathbf{U}}_{-k}^T - \mathbf{U}_{-k}^{*T})\widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*\|_2 \\ &\le \|\widetilde{\mathbf{U}}_{-k}^T\|_2\|\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}_{-k} - \widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*\|_2 + \|\widetilde{\mathbf{U}}_{-k}^T - \mathbf{U}_{-k}^{*T}\|_2\|\widehat{\boldsymbol{\Sigma}}\mathbf{U}_{-k}^*\|_2 \\ &\le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned}$$

Together with the upper bound of $|\widetilde{z}_{kk} - z_{kk}^*|$ in Lemma EC.17, it yields that

$$\|\widetilde{\mathbf{A}}_0 - \mathbf{A}_0\|_2 \le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2},$$

which further entails that

$$\|\widetilde{\mathbf{A}} - \mathbf{A}^*\|_2 \le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}. \tag{ER.121}$$

For term $\widehat{\boldsymbol{\Sigma}}\mathbf{U}^*_{-k}(\widetilde{\mathbf{A}}-\mathbf{A}^*)\widetilde{\mathbf{U}}^T_{-k}$ above, it follows from (ER.110), (ER.116), and (ER.121) that

$$\begin{aligned}\|\widehat{\boldsymbol{\Sigma}}\mathbf{U}^*_{-k}(\widetilde{\mathbf{A}}-\mathbf{A}^*)\widetilde{\mathbf{U}}^T_{-k}\|_2 &\le \|\widehat{\boldsymbol{\Sigma}}\mathbf{U}^*_{-k}\|_2\|\widetilde{\mathbf{A}}-\mathbf{A}^*\|_2\|\widetilde{\mathbf{U}}^T_{-k}\|_2\\ &\le c(r^*+s_u+s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}.\end{aligned} \tag{ER.122}$$

Therefore, combining (ER.112), (ER.113), (ER.118), (ER.119), and (ER.122) gives that

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i-\boldsymbol{w}_i^*\|_2\le c(r^*+s_u+s_v)\{n^{-1}\log(pq)\}^{1/2}.$$

Moreover, from the triangle inequality we have that for sufficiently large n,

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i\|_2\le \max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i-\boldsymbol{w}_i^*\|_2+\max_{1\le i\le p}\|\boldsymbol{w}_i^*\|_2\le c,$$

which completes the proof of part (b). This concludes the proof of Lemma EC.22.

ER.1.23. Proof of Lemma EC.23

Let us first analyze terms $|z^*_{kk}-z^*_{ii}|$ and $|\widetilde{z}_{kk}-\widetilde{z}_{ii}|$. For each $i\in\{k+1,\cdots,r^*\}$, in view of Condition 2 we have that $d_k^{*2}\rho_l\le z^*_{kk}\le d_k^{*2}\rho_u$ and $d_i^{*2}\rho_l\le z^*_{ii}\le d_i^{*2}\rho_u$, which lead to

$$d_k^{*2}\rho_l-d_i^{*2}\rho_u\le z^*_{kk}-z^*_{ii}\le d_k^{*2}\rho_u-d_i^{*2}\rho_l.$$

Since $\max_i d_i^*=d^*_{k+1}$, using similar arguments as for (ER.87) and (ER.88), it holds that

$$\min_i(z^*_{kk}-z^*_{ii})\ge\min_i(d_k^{*2}\rho_l-d_i^{*2}\rho_u)=d_k^{*2}\rho_l-d_{k+1}^{*2}\rho_u\ge cd_k^{*2}. \tag{ER.123}$$

Further, from part (c) of Lemma EC.17 and $d_i^*<d_k^*$, we can obtain that

$$\begin{aligned}|(\widetilde{z}_{kk}-\widetilde{z}_{ii})-(z^*_{kk}-z^*_{ii})| &\le |\widetilde{z}_{kk}-z^*_{kk}|+|\widetilde{z}_{ii}-z^*_{ii}|\\ &\le c(r^*+s_u+s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_k^*.\end{aligned}$$

By the assumption in Theorem 4 that $m^{1/2}\kappa_n^{(k)}=o(1)$, we have $\gamma_n=o(1)$. Then for sufficiently large n, it follows that

$$|\widetilde{z}_{kk}-\widetilde{z}_{ii}|\ge|z^*_{kk}-z^*_{ii}|-|(\widetilde{z}_{kk}-\widetilde{z}_{ii})-(z^*_{kk}-z^*_{ii})|\ge cd_k^{*2}.$$

We next prove that $\sum_{k+1\le j\le r^*, j\ne i}|z^*_{ij}|=o(|z^*_{kk}-z^*_{ii}|)$. Observe that $|z^*_{ij}|=d_i^*d_j^*|\boldsymbol{l}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_j^*|$. For each $i,j\in\{k+1,\cdots,r^*\}$, when $j<i$, from Condition 5 we have that

$$(d_i^{*2}/d_j^*)|\boldsymbol{l}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_i^*|=o(n^{-1/2}).$$

When $j > i$, similarly we can obtain that $(d_j^{*2}/d_i^*)|\boldsymbol{l}_j^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_i^*| = o(n^{-1/2})$. Then for each $i \in \{k+1, \cdots, r^*\}$, we can deduce that

$$\begin{aligned}
\sum_{k+1 \le j \le r^*,\, j \ne i} \frac{|z_{ij}^*|}{|z_{kk}^* - z_{ii}^*|} &\le c \sum_{k+1 \le j \le r^*,\, j \ne i} \frac{d_i^* d_j^* |\boldsymbol{l}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{l}_j^*|}{d_k^{*2}} \\
&\le c \sum_{k+1 \le j \le r^*,\, j < i} \frac{d_j^{*2}}{d_i^* d_k^{*2}\sqrt{n}} + c \sum_{k+1 \le j \le r^*,\, j > i} \frac{d_i^{*2}}{d_j^* d_k^{*2}\sqrt{n}} \\
&\le \frac{cr^*}{d_{r^*}\sqrt{n}}.
\end{aligned}$$

From Condition 3 that $r^*\gamma_n = o(d_{r^*}^*)$ and $\gamma_n = (r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2} \ge n^{-1/2}$, it holds that $r^* n^{-1/2} = o(d_{r^*}^*)$, which further leads to

$$\sum_{k+1 \le j \le r^*,\, j \ne i} \frac{|z_{ij}^*|}{|z_{kk}^* - z_{ii}^*|} = o(1). \tag{ER.124}$$

It remains to show that $\sum_{k+1 \le j \le r^*, j \ne i} |\widetilde{z}_{ij}| = o(|\widetilde{z}_{kk} - \widetilde{z}_{ii}|)$. For term $|\widetilde{z}_{ij}|$, an application of similar arguments as for (ER.92) gives that

$$|\widetilde{z}_{ij} - z_{ij}^*| = |\widetilde{\boldsymbol{u}}_i^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{u}}_j - \boldsymbol{u}_i^{*T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_j^*| \le c\gamma_n \max\{d_i^*, d_j^*\}.$$

It follows that

$$|\widetilde{z}_{ij}| \le |z_{ij}^*| + |\widetilde{z}_{ij} - z_{ij}^*| = |z_{ij}^*| + O(\gamma_n \max\{d_i^*, d_j^*\}).$$

Since $|z_{kk}^* - z_{ii}^*| \ge cd_k^{*2}$ and $|\widetilde{z}_{kk} - \widetilde{z}_{ii}| \ge cd_k^{*2}$, in light of (ER.124) it suffices to show that $\sum_{k+1 \le j \le r^*, j \ne i} \gamma_n \max\{d_i^*, d_j^*\} = o(|\widetilde{z}_{kk} - \widetilde{z}_{ii}|)$. Since $i, j \ge k+1$ such that $\max\{d_i^*, d_j^*\} \le d_{k+1}^*$, we can show that

$$\sum_{k+1 \le j \le r^*,\, j \ne i} \frac{\max\{d_i^*, d_j^*\}\gamma_n}{|\widetilde{z}_{kk} - \widetilde{z}_{ii}|} \le \frac{cd_{k+1}^* r^*\gamma_n}{d_k^{*2}} \le c\frac{r^*\gamma_n}{d_{k^*}} \le c\frac{r^*\gamma_n}{d_{r^*}} = o(1),$$

where we have used Condition 3 that $r^*\gamma_n = o(d_{r^*}^*)$. Therefore, it holds that

$$\sum_{k+1 \le j \le r^*,\, j \ne i} \frac{|\widetilde{z}_{ij}|}{|\widetilde{z}_{kk} - \widetilde{z}_{ii}|} = o(1),$$

which concludes the proof of Lemma EC.23.

ER.1.24. Proof of Lemma EC.24

Let us first analyze term $\mathbf{I}_{r^*-k} - \widetilde{z}_{kk}^{-1}(\widetilde{\mathbf{U}}^{(2)})^T\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)}$, which is equivalent to analyzing the nonsingularity of matrix $\mathbf{A} =: \widetilde{z}_{kk}\mathbf{I}_{r^*-k} - (\widetilde{\mathbf{U}}^{(2)})^T\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)} \in \mathbb{R}^{(r^*-k)\times(r^*-k)}$. Denote by $\mathbf{A} = (a_{ij})$. Then we can see that for each $i, j \in \{k+1, \cdots, r^*\}$,

$$a_{ij} = \begin{cases} \widetilde{z}_{kk} - \widetilde{z}_{ii} & \text{if } i = j, \\ -\widetilde{z}_{ij} & \text{if } i \ne j. \end{cases} \tag{ER.125}$$

From Lemma EC.23, it holds that $\sum_{j\neq i}|a_{ij}|=o(|a_{ii}|)$ for each $i\in\{k+1,\cdots,r^*\}$. Hence, it follows that

$$|a_{ii}|>\sum_{j\neq i}|a_{ij}|$$

for all $i\in\{k+1,\cdots,r^*\}$, which entails that matrix $\mathbf{A}$ is strictly diagonally dominant. With the aid of the Levy–Desplanques Theorem in Horn and Johnson (2012), we see that matrix $\mathbf{A}$ is nonsingular and thus matrix $\mathbf{I}_{r^*-k}-\widetilde{z}_{kk}^{-1}(\widetilde{\mathbf{U}}^{(2)})^T\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}^{(2)}$ is also nonsingular. Moreover, using similar arguments we can also show that matrix $\mathbf{I}_{r^*-k}-z_{kk}^{*-1}(\mathbf{U}^{*(2)})^T\widehat{\mathbf{\Sigma}}\mathbf{U}^{*(2)}$ is nonsingular. Therefore, both $\widetilde{\mathbf{W}}_k$ and $\mathbf{W}_k^*$ are well-defined and satisfy the property in Proposition 5, which completes the proof of Lemma EC.24.

ER.1.25. Proof of Lemma EC.25

The proof of Lemma EC.25 is similar to that of Lemma EC.21 in Section ER.1.21. Notice that $\|\mathbf{U}^{*(2)}\|_0\leq\|\mathbf{U}^*\|_0=s_u$ and $(\mathbf{U}^{*(2)})^T\mathbf{U}^{*(2)}=\mathrm{diag}\{d_{k+1}^{*2},\cdots,d_{r^*}^{*2}\}$. Using similar arguments as for (ER.101) and (ER.102), we can obtain that

$$\|\mathbf{U}^{*(2)}\|_2\leq cd_{k+1}^*\ \text{ and }\ \|\widehat{\mathbf{\Sigma}}\mathbf{U}^{*(2)}\|_2\leq cd_{k+1}^*. \tag{ER.126}$$

From Definition 2, we have that

$$\begin{aligned}&\|\widetilde{\mathbf{U}}^{(2)}-\mathbf{U}^{*(2)}\|_0\leq\|\widetilde{\mathbf{U}}-\mathbf{U}^*\|_0\leq 2(r^*+s_u+s_v),\\&\|\widetilde{\mathbf{U}}^{(2)}-\mathbf{U}^{*(2)}\|_2\leq\|\widetilde{\mathbf{U}}^{(2)}-\mathbf{U}^{*(2)}\|_F\leq\|\widetilde{\mathbf{U}}-\mathbf{U}^*\|_F\leq c\gamma_n.\end{aligned} \tag{ER.127}$$

An application of similar arguments as for (ER.101) leads to

$$\|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{U}}^{(2)}-\mathbf{U}^{*(2)})\|_2\leq c\|\widetilde{\mathbf{U}}^{(2)}-\mathbf{U}^{*(2)}\|_2\leq c\gamma_n. \tag{ER.128}$$

Then for sufficiently large n, it follows that

$$\|\widetilde{\mathbf{U}}^{(2)}\|_2\leq\|\widetilde{\mathbf{U}}^{(2)}-\mathbf{U}^{*(2)}\|_2+\|\mathbf{U}^{*(2)}\|_2\leq cd_{k+1}^*, \tag{ER.129}$$

$$\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}^{(2)}\|_2\leq\|\widehat{\mathbf{\Sigma}}(\widetilde{\mathbf{U}}^{(2)}-\mathbf{U}^{*(2)})\|_2+\|\widehat{\mathbf{\Sigma}}\mathbf{U}^{*(2)}\|_2\leq cd_{k+1}^*. \tag{ER.130}$$

For $\widetilde{\mathbf{M}}_k=-\widetilde{z}_{kk}^{-1}\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{C}}^{(2)}$, we can deduce that

$$\begin{aligned}\|\widetilde{\mathbf{M}}_k\|_2&\leq|\widetilde{z}_{kk}^{-1}|\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}^{(2)}(\widetilde{\mathbf{V}}^{(2)})^T\|_2\leq|\widetilde{z}_{kk}^{-1}|\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}^{(2)}\|_2\|(\widetilde{\mathbf{V}}^{(2)})^T\|_2\\&\leq cd_k^{*-2}d_{k+1}^*,\end{aligned}$$

where we have used part (c) of Lemma EC.17, $\|\widehat{\mathbf{\Sigma}}\widetilde{\mathbf{U}}^{(2)}\|_2\leq cd_{k+1}^*$, and $\|\widetilde{\mathbf{V}}_k^{(2)}\|_2=1$. With similar arguments, we can show that

$$\|\mathbf{M}_k^*\|_2\leq cd_k^{*-2}d_{k+1}^*.$$

For term $\|\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*\|_2$, with the aid of similar arguments as for (ER.103)–(ER.104), it holds that

$$\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}^{(2)} - \mathbf{C}^{*(2)})\|_2 \le c(r^* + s_u + s_v)^{1/2}\eta_n^2 \left\{n^{-1}\log(pq)\right\}^{1/2}.$$

Together with part (c) of Lemma EC.17, $\|\widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\|_2 \le cd_{k+1}^*$, and $\|(\mathbf{V}^{*(2)})^T\|_2 = 1$, we can obtain that

$$\begin{aligned}
\|\widetilde{\mathbf{M}}_k - \mathbf{M}_k^*\|_2 &\le |\widetilde{z}_{kk}^{-1} - z_{kk}^{*-1}|\|\widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\|_2\|(\mathbf{V}^{*(2)})^T\|_2 + |z_{kk}^{*-1}|\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{C}}^{(2)} - \mathbf{C}^{*(2)})\|_2 \\
&\le c(r^* + s_u + s_v)^{1/2}\eta_n^2\{n^{-1}\log(pq)\}^{1/2} d_k^{*-2}.
\end{aligned}$$

This concludes the proof of Lemma EC.25.

ER.1.26. Proof of Lemma EC.26

Similar to the proof of Lemma EC.19, an application of similar arguments as for (ER.94)–(ER.96) yields that once the results in parts (a) and (b) are established, we can obtain immediately the results in parts (c)–(e). Hence, we need only to prove parts (a) and (b), which will be based on the proof of Lemma EC.22. Compared to Lemma EC.22, we see that the only difference is that matrices $\widetilde{\mathbf{U}}_{-k}$ and $\mathbf{U}_{-k}^*$ in $\widetilde{\mathbf{W}}_k$ and $\mathbf{W}_k^*$ of Lemma EC.22 are now replaced with their submatrices $\widetilde{\mathbf{U}}^{(2)}$ and $\mathbf{U}^{*(2)}$ in this lemma. As a result, $\widetilde{\mathbf{U}}^{(2)}$ and $\mathbf{U}^{*(2)}$ will only be more sparse than $\widetilde{\mathbf{U}}_{-k}$ and $\mathbf{U}_{-k}^*$, respectively. Then applying similar arguments as in the proof of part (a) of Lemma EC.22, we can obtain the results in part (a) of the current lemma.

It remains to show part (b), which also follows similar technical arguments as in the proof of part (b) of Lemma EC.22. In view of (ER.126), (ER.129), and (ER.130), we can deduce that

$$\{\|\widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\|_2, \|\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)}\|_2,\ \|\widetilde{\mathbf{U}}^{(2)}\|_2, \|\mathbf{U}^{*(2)}\|_2\} \le cd_{k+1}^*. \tag{ER.131}$$

Denote by $\mathbf{A}^* = (z_{kk}^*\mathbf{I}_{r^*-k} - (\mathbf{U}^{(2)*})^T\widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)})^{-1}$, $\mathbf{A}_0^{-1} = \mathbf{A}^*$, $\widetilde{\mathbf{A}} = (\widetilde{z}_{kk}\mathbf{I}_{r^*-k} - (\widetilde{\mathbf{U}}^{(2)})^T\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)})^{-1}$, and $\widetilde{\mathbf{A}}_0^{-1} = \widetilde{\mathbf{A}}$. Using the technical arguments in the proof of Lemma EC.24, we see that both $\mathbf{A}_0$ and $\widetilde{\mathbf{A}}_0$ are strictly diagonally dominant. Similar to (ER.109), it holds that

$$\begin{aligned}
\|\mathbf{A}^*\|_2 = \|\mathbf{A}_0^{-1}\|_2 &\le c(\min_{k<i\le r^*}|z_{kk}^* - z_{ii}^*|)^{-1} \\
&\le c|z_{kk}^* - z_{k+1,k+1}^*|^{-1} \le cd_k^{*-2},
\end{aligned} \tag{ER.132}$$

where the last inequality above is due to (ER.123). Moreover, with similar arguments we have that

$$\|\widetilde{\mathbf{A}}\|_2 = \|\mathbf{A}_0^{-1}\|_2 \le cd_k^{*-2}. \tag{ER.133}$$

Observe that $\boldsymbol{w}_i^{*T} = \widehat{\boldsymbol{\theta}}_i^T(\mathbf{I}_p + \widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\mathbf{A}^*(\mathbf{U}^{*(2)})^T)$, where $\widehat{\boldsymbol{\theta}}_i^T$ is the ith row of $\widehat{\boldsymbol{\Theta}}$. It follows from Definition 1 that $\max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_2 \le c$, (ER.131), and (ER.132) that

$$\begin{aligned}
\max_{1\le i\le p}\|\boldsymbol{w}_i^*\|_2 &\le \max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_2\|\mathbf{I}_p + \widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\mathbf{A}^*(\mathbf{U}^{*(2)})^T\|_2 \\
&\le \max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i\|_2(1 + \|\widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\|_2\|\mathbf{A}^*\|_2\|(\mathbf{U}^{*(2)})^T\|_2) \\
&\le c\max\{1, d_{k+1}^{*2}/d_k^{*2}\} \le c.
\end{aligned}$$

Since $\widetilde{\boldsymbol{w}}_i^T = \widehat{\boldsymbol{\theta}}_i^T(\mathbf{I}_p + \widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)}\widetilde{\mathbf{A}}(\widetilde{\mathbf{U}}^{(2)})^T)$, we can show that

$$\begin{aligned}\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^*\|_2 &\le \max_{1\le i\le p}\|\widehat{\boldsymbol{\theta}}_i^T\|_2\|\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)}\widetilde{\mathbf{A}}(\widetilde{\mathbf{U}}^{(2)})^T - \widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\mathbf{A}^*(\mathbf{U}^{*(2)})^T\|_2 \\ &\le c\|\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)}\widetilde{\mathbf{A}}(\widetilde{\mathbf{U}}^{(2)})^T - \widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\mathbf{A}^*(\mathbf{U}^{*(2)})^T\|_2. \end{aligned} \tag{ER.134}$$

By some simple calculations, the term above can be decomposed as

$$\begin{aligned}&\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)}\widetilde{\mathbf{A}}(\widetilde{\mathbf{U}}^{(2)})^T - \widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\mathbf{A}^*(\mathbf{U}^{*(2)})^T = (\widehat{\boldsymbol{\Sigma}}\widetilde{\mathbf{U}}^{(2)} - \widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)})\widetilde{\mathbf{A}}(\widetilde{\mathbf{U}}^{(2)})^T \\ &\quad + \widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}(\widetilde{\mathbf{A}} - \mathbf{A}^*)(\widetilde{\mathbf{U}}^{(2)})^T + \widehat{\boldsymbol{\Sigma}}\mathbf{U}^{*(2)}\mathbf{A}^*(\widetilde{\mathbf{U}}^{(2)} - \mathbf{U}^{*(2)})^T. \end{aligned} \tag{ER.135}$$

From (ER.127) and (ER.128), we see that

$$\|\widehat{\boldsymbol{\Sigma}}(\widetilde{\mathbf{U}}^{(2)} - \mathbf{U}^{*(2)})\|_2 \le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}, \tag{ER.136}$$

$$\|\widetilde{\mathbf{U}}^{(2)} - \mathbf{U}^{*(2)}\|_2 \le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}. \tag{ER.137}$$

Also, applying similar arguments as for (ER.120)–(ER.122), it holds that

$$\|\widetilde{\mathbf{A}} - \mathbf{A}^*\|_2 \le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_k^{*-3}. \tag{ER.138}$$

Combining (ER.131) and (ER.134)–(ER.138) yields that

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^*\|_2 \le c(r^* + s_u + s_v)\eta_n^2\{n^{-1}\log(pq)\}^{1/2}d_{k+1}^*d_k^{*-2}.$$

Further, by the triangle inequality we have that

$$\max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i\|_2 \le \max_{1\le i\le p}\|\widetilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^*\|_2 + \max_{1\le i\le p}\|\boldsymbol{w}_i^*\|_2 \le c.$$

For $k = r^*$, it holds that $\widetilde{\mathbf{W}}_k = \mathbf{W}_k^* = \widehat{\boldsymbol{\Theta}}$. Therefore, based on Definition 1, all conclusions of this lemma still hold using similar analysis as above. This completes the proof of Lemma EC.26.

ER.2. Additional technical details

ER.2.1. The Taylor expansion on the Riemannian manifold

To facilitate our technical analysis, let us first introduce briefly some necessary background on the Riemannian manifold. For more detailed and rigorous introduction to the Riemannian manifold, see, e.g., Do Carmo and Flaherty (1992). Let $\mathcal{M}$ be a p-dimensional compact Riemannian manifold. For a given $\mathbf{X} \in \mathcal{M}$, the tangent space to $\mathcal{M}$ at $\mathbf{X}$ is a p-dimensional linear space and will be denoted as $T_X\mathcal{M}$. A Riemannian metric g_X is defined at each point $\mathbf{X} \in \mathcal{M}$ by the map $g_X : T_X\mathcal{M} \times T_X\mathcal{M} \to \mathbb{R}$ and is an inner product on the tangent space $T_X\mathcal{M}$. For $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in T_X\mathcal{M}$, denote the inner product as $\langle\boldsymbol{\xi}_1, \boldsymbol{\xi}_2\rangle = g_X(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$. Then the inner product induces a norm $\|\cdot\|$, which is denoted as $\|\boldsymbol{\xi}\| = \sqrt{\langle\boldsymbol{\xi}, \boldsymbol{\xi}\rangle}$ for each $\boldsymbol{\xi} \in T_X\mathcal{M}$. Given $\mathbf{X} \in \mathcal{M}$ and its tangent vector $\boldsymbol{\xi} \in T_X\mathcal{M}$, let $\gamma(t;\mathbf{X},\boldsymbol{\xi})$ be the geodesic (the locally length-minimizing curve) satisfying that $\gamma(0;\mathbf{X},\boldsymbol{\xi}) = \mathbf{X}$ and $\dot{\gamma}(0;\mathbf{X},\boldsymbol{\xi}) = \boldsymbol{\xi}$. The exponential map is defined through the geodesic. Specifically, the exponential map at a point $\mathbf{X}$ is defined as

$$\exp_X : T_X\mathcal{M} \to \mathcal{M},\, \boldsymbol{\xi} \mapsto \exp_X \boldsymbol{\xi} = \gamma(1;\mathbf{X},\boldsymbol{\xi}). \tag{ER.139}$$

By Theorem 3.7 and Remark 3.8 in Do Carmo and Flaherty (1992), for each point $\mathbf{X} \in \mathcal{M}$ a normal neighborhood S of $\mathbf{X} \in \mathcal{M}$ is the one that satisfies (1) each point $\mathbf{Y}$ in S can be joined to $\mathbf{X}$ by a unique geodesic $\gamma(t;\mathbf{X},\boldsymbol{\xi})$, $0 \le t \le 1$, with $\gamma(0;\mathbf{X},\boldsymbol{\xi}) = \mathbf{X}$ and $\gamma(1;\mathbf{X},\boldsymbol{\xi}) = \mathbf{Y}$; and (2) the exponential map $\exp_X$ is a local diffeomorphism between a neighborhood of $\mathbf{0} \in T_X\mathcal{M}$ and a neighborhood S of $\mathbf{X} \in \mathcal{M}$. Since $\exp_X$ is a local diffeomorphism, the exponential map is defined only locally in that it maps a small neighborhood of $\mathbf{0} \in T_X\mathcal{M}$ to a neighborhood S of $\mathbf{X} \in \mathcal{M}$. Denote by $\exp_X^{-1}$ the inverse of the exponential map. For each point $\mathbf{Y}$ in S, we can connect two points $\mathbf{X}$ and $\mathbf{Y}$ by the exponential map that $\exp_X \boldsymbol{\xi} = \mathbf{Y}$, or equivalently, $\boldsymbol{\xi} = \exp_X^{-1} \mathbf{Y}$ for $\boldsymbol{\xi} \in T_X\mathcal{M}$.

LEMMA EC.27. *(Mukherjee et al. 2010, Lemma A.3) Let $\mathcal{M}$ be a compact Riemannian manifold. Assume that f is a twice differentiable function on $\mathcal{M}$. Denote by $\nabla_{\mathcal{M}} f(\mathbf{X})$ the gradient of f. Then there exists a constant $C > 0$ such that for all $\mathbf{X} \in \mathcal{M}$ and $\boldsymbol{\xi} \in T_X\mathcal{M}$, $\|\boldsymbol{\xi}\| \le \epsilon_0$ with some $\epsilon_0 > 0$, the first-order Taylor expansion below satisfies that*

$$\|f(\exp_X(\boldsymbol{\xi})) - f(\mathbf{X}) - \langle \nabla_{\mathcal{M}} f(\mathbf{X}), \boldsymbol{\xi}\rangle\| \le C\|\boldsymbol{\xi}\|^2. \tag{ER.140}$$

From (ER.140) in Lemma EC.27 above, we can write the first-order Taylor expansion on the Riemannian manifold as

$$f(\exp_X(\boldsymbol{\xi})) = f(\mathbf{X}) + \langle \nabla_{\mathcal{M}} f(\mathbf{X}), \boldsymbol{\xi}\rangle + O(\|\boldsymbol{\xi}\|^2) \tag{ER.141}$$

or

$$f(\mathbf{Y}) = f(\mathbf{X}) + \left\langle \nabla_{\mathcal{M}} f(\mathbf{X}), \exp_X^{-1} \mathbf{Y}\right\rangle + O(\|\exp_X^{-1} \mathbf{Y}\|^2) \tag{ER.142}$$

for $\boldsymbol{\xi} = \exp_X^{-1} \mathbf{Y}$ with $\boldsymbol{\xi} \in T_X\mathcal{M}$.

ER.2.2. The geometry of the Stiefel manifold

We now focus on a special manifold, the so-called Stiefel manifold. We will briefly introduce some necessary background on the Stiefel manifold. The Stiefel manifold $\mathrm{St}(p,n)$ denotes the set of all orthonormal p-frames in the Euclidean space $\mathbb{R}^n$, where the p-frame is a set of p orthonormal vectors in $\mathbb{R}^n$. Specifically, the Stiefel manifold is given by $\mathrm{St}(p,n) = \{\mathbf{X} \in \mathbb{R}^{n\times p} : \mathbf{X}^T\mathbf{X} = \mathbf{I}_p\}$. For $p = 1$, the Stiefel manifold $\mathrm{St}(p,n)$ reduces to the unit sphere $\mathcal{S}^{n-1}$ in $\mathbb{R}^n$, where $\mathcal{S}^{n-1} := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T\mathbf{x} = 1\}$. For $p = n$, the Stiefel manifold

St(p,n) becomes the orthogonal group $O(n)$, where $O(j) := \{\mathbf{O}_j \in \mathbb{R}^{j\times j} : \mathbf{O}_j^T \mathbf{O}_j = \mathbf{I}_j\}$ for each positive integer j. We can also represent the Stiefel manifold as St$(p,n) = O(n)/O(n-p)$. See, e.g., Edelman et al. (1998), Lv (2013) for more details on these representations.

Denote by $T_X \operatorname{St}(p,n)$ the tangent space of the Stiefel manifold. The tangent space $T_X \operatorname{St}(p,n)$ admits the form

$$T_X \operatorname{St}(p,n) = \left\{\mathbf{X}\mathbf{A} + \mathbf{B} : \mathbf{A} \in \mathbb{R}^{p\times p}, \mathbf{A} = -\mathbf{A}^T, \mathbf{B} \in \mathbb{R}^{n\times p}, \mathbf{X}^T\mathbf{B} = \mathbf{0}\right\}. \tag{ER.143}$$

According to the Stiefel manifold representation of orthonormal matrices in Edelman et al. (1998) and Chen and Huang (2012), for each given $\mathbf{X}^* \in \operatorname{St}(p,n)$, matrices on the Stiefel manifold $\operatorname{St}(p,n)$ can be represented as

$$\{\mathbf{X} = \exp_{X^*} \boldsymbol{\xi}^* : \boldsymbol{\xi}^* = \mathbf{X}^*\mathbf{A} + \mathbf{B} \in T_{X^*} \operatorname{St}(p,n)\}, \tag{ER.144}$$

where $\exp_{X^*}$ is the exponential map defined in (ER.139) above. For the Stiefel manifold, it is common to use the canonical metric as suggested in Edelman et al. (1998). For the tangent vector $\boldsymbol{\xi} = \mathbf{X}\mathbf{A} + \mathbf{B} \in T_X \operatorname{St}(p,n)$, the canonical metric $\langle \cdot, \cdot \rangle_c$ is given by

$$\begin{aligned} \langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle_c &= \operatorname{tr}(\boldsymbol{\xi}^T (\mathbf{I}_n - \frac{1}{2}\mathbf{X}\mathbf{X}^T)\boldsymbol{\xi}) \\ &= \frac{1}{2}\operatorname{tr}(\mathbf{A}^T\mathbf{A}) + \operatorname{tr}(\mathbf{B}^T\mathbf{B}), \end{aligned} \tag{ER.145}$$

where the second equality above can be derived easily using $\mathbf{X}^T\mathbf{X} = \mathbf{I}_p$ and $\mathbf{X}^T\mathbf{B} = 0$. With such canonical metric, we can obtain the gradient of a function on the Stiefel manifold below.

LEMMA EC.28. *((Edelman et al. 1998, Section 2.4.4)) For a real-valued function f defined on the Stiefel manifold* $\operatorname{St}(p,n)$*, let $\nabla_X f$ be the gradient of f at $\mathbf{X} \in \operatorname{St}(p,n)$. Then it holds that*

$$\nabla_X f = \frac{\partial f}{\partial \mathbf{X}} - \mathbf{X}\frac{\partial f}{\partial \mathbf{X}^T}\mathbf{X}.$$

Let us now consider a special case that $p = 1$ for the Stiefel manifold $\operatorname{St}(p,n)$. For such case, the Stiefel manifold St(p,n) reduces to the unit sphere $\mathcal{S}^{n-1}$. For $\mathbf{x} \in \operatorname{St}(1,n)$ and its tangent space $T_{\mathrm{x}} \operatorname{St}(1,n)$, from (ER.143) we see that $\mathbf{A} = 0$. Then we can write $T_{\mathrm{x}} \operatorname{St}(1,n)$ as

$$T_{\mathrm{x}} \operatorname{St}(1,n) = \{\mathbf{B} : \mathbf{B} \in \mathbb{R}^n, \mathbf{x}^T\mathbf{B} = \mathbf{0}\}. \tag{ER.146}$$

In addition, similar to (ER.144), for given $\mathbf{x}^* \in \operatorname{St}(1,n)$, vectors on the Stiefel manifold $\operatorname{St}(1,n)$ can be represented as

$$\{\mathbf{x} = \exp_{\mathrm{x}^*} \boldsymbol{\xi}^* : \mathbf{x}^{*T}\boldsymbol{\xi}^* = \mathbf{0}\}. \tag{ER.147}$$

Table 6 The list of 20 selected responses for the real data application in Section 5.

Variable	Description
RPI	Real personal income
INDPRO	Total industrial production
CUMFNS	Capacity utilization: manufacturing
UNRATE	Civilian unemployment rate
PAYEMS	Total number of employees on non-agricultural payrolls
CES0600000007	Average weekly hours: goods-producing
HOUST	Total housing starts
DPCERA3M086SBEA	Real personal consumption expenditures
NAPMNOI	ISM manufacturing: new orders index
CMRMTSPLx	Real manufacturing and trade industries sales
FEDFUNDS	Effective federal funds rate
T1YFFM	1-Year treasury constant maturity minus FEDFUNDS
T10YFFM	10-Year treasury constant maturity minus FEDFUNDS
BAAFFM	Moody's baa corporate bond minus FEDFUNDS
EXUSUKx	U.S.-U.K. exchange rate
WPSFD49207	Producer price index for finished goods
PPICMM	Producer price index for commodities
CPIAUCSL	Consumer price index for all items
PCEPI	Personal consumption expenditure implicit price deflator
S&P 500	S&P's common stock price index: composite

Then for the tangent vector $\boldsymbol{\xi} = \mathbf{B} \in T_{\mathrm{x}}\,\mathrm{St}(1,n)$, the canonical metric is given by

$$\begin{aligned}
\langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle_c &= \mathrm{tr}(\boldsymbol{\xi}^T(\mathbf{I}_n - \frac{1}{2}\mathbf{x}\mathbf{x}^T)\boldsymbol{\xi}) \\
&= \mathrm{tr}(\mathbf{B}^T\mathbf{B}) = \mathrm{tr}(\boldsymbol{\xi}^T\boldsymbol{\xi}) \\
&:= \langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle_e\,,
\end{aligned}$$

where $\langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle_e = \mathrm{tr}(\boldsymbol{\xi}^T\boldsymbol{\xi})$ represents the (usual) Euclidean metric. Thus, for the case of $p = 1$ the canonical metric is in fact equivalent to the Euclidean metric. To simplify the notation, denote the metric $\langle \cdot, \cdot \rangle$ on $\mathrm{St}(1,n)$ as

$$\langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle = \mathrm{tr}(\boldsymbol{\xi}^T\boldsymbol{\xi}). \tag{ER.148}$$

Moreover, since $\boldsymbol{\xi} \in T_{\mathrm{x}}\,\mathrm{St}(1,n)$ is an n-dimensional vector and $\langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle = \mathrm{tr}(\boldsymbol{\xi}^T\boldsymbol{\xi}) = \boldsymbol{\xi}^T\boldsymbol{\xi}$, such metric induces the norm $\|\boldsymbol{\xi}\|_2^2 = \boldsymbol{\xi}^T\boldsymbol{\xi}$.

For the Stiefel manifold $\mathrm{St}(1,n)$, the gradient given in Lemma EC.28 above can be written as

$$\begin{aligned}
\nabla_{\mathrm{x}} f &= \frac{\partial f}{\partial \mathbf{x}} - \mathbf{x}\frac{\partial f}{\partial \mathbf{x}^T}\mathbf{x} = \frac{\partial f}{\partial \mathbf{x}} - \mathbf{x}\mathbf{x}^T\frac{\partial f}{\partial \mathbf{x}} \\
&= (\mathbf{I}_n - \mathbf{x}\mathbf{x}^T)\frac{\partial f}{\partial \mathbf{x}}.
\end{aligned} \tag{ER.149}$$

Furthermore, we can characterize the geodesic on $\mathrm{St}(1,n)$, i.e., on the unit sphere $\mathcal{S}^{n-1}$ below.

LEMMA EC.29. *(Absil et al. 2008, Example 5.4.1) For the unit sphere $\mathcal{S}^{n-1}$ with metric* (ER.148)*, let $\mathbf{x}\in\mathcal{S}^{n-1}$ and $\boldsymbol{\xi}$ be the tangent vector in the tangent space of $\mathcal{S}^{n-1}$ at $\mathbf{x}$. Let $t\mapsto\gamma(t;\mathbf{x},\boldsymbol{\xi})$ be the geodesic on $\mathcal{S}^{n-1}$ with $\gamma(0;\mathbf{x},\boldsymbol{\xi})=\mathbf{x}$ and $\dot\gamma(0;\mathbf{x},\boldsymbol{\xi})=\boldsymbol{\xi}$. Then the geodesic admits the representation*

$$\gamma(t;\mathbf{x},\boldsymbol{\xi})=\mathbf{x}\cos(\|\boldsymbol{\xi}\|_2 t)+\frac{\boldsymbol{\xi}}{\|\boldsymbol{\xi}\|_2}\sin(\|\boldsymbol{\xi}\|_2 t).$$

ER.3. Additional real data results

We provide in Table 6 above the list of 20 selected responses along with their descriptions for the real data application in Section 5.

ER.4. Additional insights into constructions of M and W

Here, we show that if one directly considers the SVD constraints in $\frac{\partial\widetilde{\psi}_k}{\partial\boldsymbol{\eta}_k}$ and calculates the regular derivatives in the Euclidean space, it will result in deficiency in the degrees of freedom and there would not exist a valid $\mathbf{W}$ matrix.

For simplicity, let us consider the rank-2 case. The following derivations are similar to those in the proofs of Lemma 1, and Propositions 2 and 3, respectively. Note that we have

$$\begin{aligned}L=(2n)^{-1}\Big\{&\|\mathbf{Y}\|_F^2+\boldsymbol{u}_1^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}_1\boldsymbol{v}_1^T\boldsymbol{v}_1+\boldsymbol{u}_2^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}_2\boldsymbol{v}_2^T\boldsymbol{v}_2+2\boldsymbol{u}_1^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}_2\boldsymbol{v}_1^T\boldsymbol{v}_2\\&-2\boldsymbol{u}_1^T\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_1-2\boldsymbol{u}_2^T\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_2\Big\}.\end{aligned}$$

After some calculations, we can deduce that

$$\frac{\partial L}{\partial\boldsymbol{u}_1}=\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T\boldsymbol{v}_1-n^{-1}\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_1,\tag{ER.150}$$

$$\frac{\partial L}{\partial\boldsymbol{v}_1}=\boldsymbol{v}_1\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1+\boldsymbol{v}_2\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2-n^{-1}\mathbf{Y}^T\mathbf{X}\boldsymbol{u}_1,\tag{ER.151}$$

$$\frac{\partial L}{\partial\boldsymbol{u}_2}=\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T\boldsymbol{v}_2-n^{-1}\mathbf{X}^T\mathbf{Y}\boldsymbol{v}_2,\tag{ER.152}$$

$$\frac{\partial L}{\partial\boldsymbol{v}_2}=\boldsymbol{v}_2\boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2+\boldsymbol{v}_1\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2-n^{-1}\mathbf{Y}^T\mathbf{X}\boldsymbol{u}_2.\tag{ER.153}$$

Denote by $\boldsymbol{\eta}_1=(\boldsymbol{v}_1^T,\boldsymbol{v}_2^T,\boldsymbol{u}_2^T)^T$, and $\boldsymbol{\eta}_1^*=(\boldsymbol{v}_1^{*T},\boldsymbol{v}_2^{*T},\boldsymbol{u}_2^{*T})^T$. Utilizing the derivatives given in (ER.151)–(ER.153) and the constraints $\|\boldsymbol{v}_1\|_2=\|\boldsymbol{v}_2\|_2=1$ and $\boldsymbol{v}_1^T\boldsymbol{v}_2=0$, we can obtain that

$$\begin{aligned}\mathbf{M}\frac{\partial L}{\partial\boldsymbol{\eta}_1}\Big|_{\boldsymbol{\eta}_1^*}&=\mathbf{M}_1\left\{\boldsymbol{v}_1^*\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1-\boldsymbol{u}_1^*)-n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1\right\}-n^{-1}\mathbf{M}_3\mathbf{X}^T\mathbf{E}\boldsymbol{v}_2^*\\&\quad+\mathbf{M}_2\left\{\boldsymbol{v}_1^*\boldsymbol{u}_2^{*T}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1-\boldsymbol{u}_1^*)-n^{-1}\mathbf{E}^T\mathbf{X}\boldsymbol{u}_2^*\right\}\\&=(\mathbf{M}_1\boldsymbol{v}_1^*\boldsymbol{u}_1^T+\mathbf{M}_2\boldsymbol{v}_1^*\boldsymbol{u}_2^{*T})\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1-\boldsymbol{u}_1^*)-n^{-1}\Big\{\mathbf{M}_1\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1\\&\quad+\mathbf{M}_2\mathbf{E}^T\mathbf{X}\boldsymbol{u}_2^*+\mathbf{M}_3\mathbf{X}^T\mathbf{E}\boldsymbol{v}_2^*\Big\}\\&=(\mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^T+\mathbf{M}_2\boldsymbol{v}_1\boldsymbol{u}_2^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1-\boldsymbol{u}_1^*)+\boldsymbol{\delta}_1',\end{aligned}$$

where we denote

$$\begin{aligned}\boldsymbol{\delta}_1' = &-n^{-1}\left\{\mathbf{M}_1\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1 + \mathbf{M}_2\mathbf{E}^T\mathbf{X}\boldsymbol{u}_2^* + \mathbf{M}_3\mathbf{X}^T\mathbf{E}\boldsymbol{v}_2^*\right\} - \\ &\left\{\mathbf{M}_1(\boldsymbol{v}_1 - \boldsymbol{v}_1^*)\boldsymbol{u}_1^T - \mathbf{M}_2(\boldsymbol{v}_1\boldsymbol{u}_2^T - \boldsymbol{v}_1^*\boldsymbol{u}_2^{*T})\right\}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1 - \boldsymbol{u}_1^*).\end{aligned} \tag{ER.154}$$

Then with the derivative in (ER.150), we can represent $\widetilde{\psi}(\boldsymbol{u}_1, \boldsymbol{\eta}_1^*)$ as

$$\begin{aligned}\widetilde{\psi}(\boldsymbol{u}_1, \boldsymbol{\eta}_1^*) &= \frac{\partial L}{\partial \boldsymbol{u}_1}\Big|_{\boldsymbol{\eta}_1^*} - \mathbf{M}\frac{\partial L}{\partial \boldsymbol{\eta}_1}\Big|_{\boldsymbol{\eta}_1^*} \\ &= (\mathbf{I}_p - \mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^T - \mathbf{M}_2\boldsymbol{v}_1\boldsymbol{u}_2^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1 - \boldsymbol{u}_1^*) + \boldsymbol{\delta}_1,\end{aligned} \tag{ER.155}$$

where $\boldsymbol{\delta}_1$ is equal to $-\boldsymbol{\delta}_1' - n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*$. Thus, combining (ER.154) and (ER.155) yields that

$$\widetilde{\psi}(\boldsymbol{u}_1, \boldsymbol{\eta}_1^*) = (\mathbf{I}_p - \mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^T - \mathbf{M}_2\boldsymbol{v}_1\boldsymbol{u}_2^T)\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1 - \boldsymbol{u}_1^*) + \boldsymbol{\epsilon} + \boldsymbol{\delta}, \tag{ER.156}$$

where $\boldsymbol{\epsilon} = n^{-1}\left\{\mathbf{M}_1\mathbf{E}^T\mathbf{X}\boldsymbol{u}_1 + \mathbf{M}_2\mathbf{E}^T\mathbf{X}\boldsymbol{u}_2^* + \mathbf{M}_3\mathbf{X}^T\mathbf{E}\boldsymbol{v}_2^*\right\} - n^{-1}\mathbf{X}^T\mathbf{E}\boldsymbol{v}_1^*$ and

$$\boldsymbol{\delta} = \left\{\mathbf{M}_1(\boldsymbol{v}_1 - \boldsymbol{v}_1^*)\boldsymbol{u}_1^T - \mathbf{M}_2(\boldsymbol{v}_1\boldsymbol{u}_2^T - \boldsymbol{v}_1^*\boldsymbol{u}_2^{*T})\right\}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{u}_1 - \boldsymbol{u}_1^*).$$

Next we proceed with the construction of matrix $\mathbf{M} = [\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3]$. Utilizing the derivatives in (ER.150)–(ER.153) and after some calculations, it holds that

$$\begin{aligned}\frac{\partial^2 L}{\partial \boldsymbol{\eta}_1 \partial \boldsymbol{\eta}_1^T} &= \begin{bmatrix} \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\mathbf{I}_q & \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{I}_q & \boldsymbol{v}_2\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}} \\ \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{I}_q & \boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{I}_q & 2\boldsymbol{v}_2\boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}} + \boldsymbol{v}_1\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}} - n^{-1}\mathbf{Y}^T\mathbf{X} \\ \mathbf{0} & 2\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T - n^{-1}\mathbf{X}^T\mathbf{Y} & \widehat{\boldsymbol{\Sigma}}\boldsymbol{v}_2^T\boldsymbol{v}_2 \end{bmatrix}, \\ \frac{\partial^2 L}{\partial \boldsymbol{u}_1 \partial \boldsymbol{\eta}_1^T} &= \left[2\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T - n^{-1}\mathbf{X}^T\mathbf{Y}, \mathbf{0}, \mathbf{0}\right].\end{aligned}$$

Note that $n^{-1}\mathbf{X}^T\mathbf{Y} = \widehat{\boldsymbol{\Sigma}}\mathbf{C} + \widehat{\boldsymbol{\Sigma}}(\mathbf{C}^* - \mathbf{C}) + n^{-1}\mathbf{X}^T\mathbf{E}$, where $\mathbf{C} = \boldsymbol{u}_1\boldsymbol{v}_1^T + \boldsymbol{u}_2\boldsymbol{v}_2^T$. Plugging it into the above derivatives, we have

$$\frac{\partial^2 L}{\partial \boldsymbol{\eta}_1 \partial \boldsymbol{\eta}_1^T} = \mathbf{A} + \boldsymbol{\Delta}_a \text{ and } \frac{\partial^2 L}{\partial \boldsymbol{u}_1 \partial \boldsymbol{\eta}_1^T} = \mathbf{B} + \boldsymbol{\Delta}_b,$$

where

$$\begin{aligned}\mathbf{A} &= \begin{bmatrix} \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\mathbf{I}_q & \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{I}_q & \boldsymbol{v}_2\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}} \\ \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{I}_q & \boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{I}_q & \boldsymbol{v}_2\boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}} \\ \mathbf{0} & -\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T + \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T & \widehat{\boldsymbol{\Sigma}}\boldsymbol{v}_2^T\boldsymbol{v}_2 \end{bmatrix}, \\ \boldsymbol{\Delta}_a &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{C} - \mathbf{C}^*)^T\widehat{\boldsymbol{\Sigma}} - n^{-1}\mathbf{E}^T\mathbf{X} \\ \mathbf{0} & \widehat{\boldsymbol{\Sigma}}(\mathbf{C} - \mathbf{C}^*) - n^{-1}\mathbf{X}^T\mathbf{E} & \mathbf{0} \end{bmatrix}, \\ \mathbf{B} &= \left[\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T - \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T, \mathbf{0}, \mathbf{0}\right], \ \boldsymbol{\Delta}_b = \left[\widehat{\boldsymbol{\Sigma}}(\mathbf{C} - \mathbf{C}^*) - n^{-1}\mathbf{X}^T\mathbf{E}, \mathbf{0}, \mathbf{0}\right].\end{aligned}$$

Then we aim to find matrix $\mathbf{M} \in \mathbb{R}^{p\times(p+2q)}$ satisfying that $\mathbf{B} - \mathbf{MA} = \mathbf{0}$. It is equivalent to solving the following equations

$$
\begin{aligned}
&\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\mathbf{M}_1 + \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{M}_2 - \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T + \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T = \mathbf{0},\\
&\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{M}_1 + \boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\mathbf{M}_2 - \mathbf{M}_3\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T + \mathbf{M}_3\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T = \mathbf{0},\\
&\mathbf{M}_1\boldsymbol{v}_2\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}} + \mathbf{M}_3\widehat{\boldsymbol{\Sigma}}\boldsymbol{v}_2^T\boldsymbol{v}_2 + \mathbf{M}_2\boldsymbol{v}_2\boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}} = \mathbf{0}.
\end{aligned}
$$

Observe that the the key terms in the modified score function (ER.156) related to the construction of matrix $\mathbf{W}$ is $\mathbf{I}_p - \mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^T - \mathbf{M}_2\boldsymbol{v}_1\boldsymbol{u}_2^T$. For simplicity, here we provide the explicit expressions of matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, respectively. It is worth mentioning that matrix $\mathbf{M}_3$ can be derived following the the above equations and matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ easily.

Recall that $z_{11} = \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1$, $z_{22} = \boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2$, and $z_{12} = \boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2$. It holds that

$$
z_{11}\mathbf{M}_1 + z_{12}\mathbf{M}_2 - \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T + \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T = \mathbf{0}, \tag{ER.157}
$$

$$
z_{12}\mathbf{M}_1 + z_{22}\mathbf{M}_2 - \mathbf{M}_3\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T + \mathbf{M}_3\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T = \mathbf{0}, \tag{ER.158}
$$

$$
\mathbf{M}_1\boldsymbol{v}_2\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}} + \mathbf{M}_3\widehat{\boldsymbol{\Sigma}}\boldsymbol{v}_2^T\boldsymbol{v}_2 + \mathbf{M}_2\boldsymbol{v}_2\boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}} = \mathbf{0}. \tag{ER.159}
$$

Then we solve the above equations under the SVD constraints $\|\boldsymbol{v}_1\|_2 = \|\boldsymbol{v}_2\|_2 = 1$ and $\boldsymbol{v}_1^T\boldsymbol{v}_2 = 0$. From (ER.159), it follows that $\mathbf{M}_3\widehat{\boldsymbol{\Sigma}} = -\mathbf{M}_1\boldsymbol{v}_2\boldsymbol{u}_1^T\widehat{\boldsymbol{\Sigma}} - \mathbf{M}_2\boldsymbol{v}_2\boldsymbol{u}_2^T\widehat{\boldsymbol{\Sigma}}$. Combining it with (ER.158) leads to

$$
(z_{12}\mathbf{M}_1 + z_{22}\mathbf{M}_2)\left(\mathbf{I}_q - \boldsymbol{v}_2\boldsymbol{v}_2^T\right) + (z_{11}\mathbf{M}_1 + z_{12}\mathbf{M}_2)\,\boldsymbol{v}_2\boldsymbol{v}_1^T = \mathbf{0}.
$$

Then using $z_{11}\mathbf{M}_1 + z_{12}\mathbf{M}_2 = \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T - \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T$ in (ER.157), we can further show that

$$
(z_{12}\mathbf{M}_1 + z_{22}\mathbf{M}_2)\left(\mathbf{I}_q - \boldsymbol{v}_2\boldsymbol{v}_2^T\right) = \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_1^T. \tag{ER.160}
$$

Moreover, (ER.157) also implies that $\mathbf{M}_1 = z_{11}^{-1}(\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T - \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T - z_{12}\mathbf{M}_2)$. Combining it with (ER.160), we can obtain the solutions for matrices $\mathbf{M}_1$ and $\mathbf{M}_2$. For simplicity, let us denote by

$$
\begin{aligned}
&\alpha_1 = -z_{11}^{-1}z_{12}\left(z_{11}z_{22} - z_{12}^2\right)^{-1}, \quad \alpha_2 = z_{11}^{-1}\left[1 - z_{11}^{-1}z_{12}^2\left(z_{11}z_{22} - z_{12}^2\right)^{-1}\right],\\
&\beta_1 = \left(z_{11}z_{22} - z_{12}^2\right)^{-1}, \quad \beta_2 = -z_{11}^{-1}z_{12}\left(z_{11}z_{22} - z_{12}^2\right)^{-1}.
\end{aligned}
$$

Then matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ are given by

$$
\mathbf{M}_1 = \alpha_1\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_1^T + \alpha_2\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T - z_{11}^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_2^T, \tag{ER.161}
$$

$$
\mathbf{M}_2 = \beta_1\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{v}_1^T + \beta_2\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{v}_1^T. \tag{ER.162}
$$

Since $\mathbf{B} - \mathbf{MA} = \mathbf{0}$, it holds that

$$
\frac{\partial^2 L}{\partial \boldsymbol{u}_1 \partial \boldsymbol{\eta}_1^T} - \mathbf{M}\frac{\partial^2 L}{\partial \boldsymbol{\eta}_1 \partial \boldsymbol{\eta}_1^T} = \boldsymbol{\Delta}_b - \mathbf{M}\boldsymbol{\Delta}_a = \boldsymbol{\Delta},
$$

where $\boldsymbol{\Delta} = [\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \boldsymbol{\Delta}_3]$ with

$$\begin{aligned}
\boldsymbol{\Delta}_1 &= \widehat{\boldsymbol{\Sigma}}(\mathbf{C} - \mathbf{C}^*) - n^{-1}\mathbf{X}^T\mathbf{E}, \\
\boldsymbol{\Delta}_2 &= \mathbf{M}_3 \left\{ n^{-1}\mathbf{X}^T\mathbf{E} - \widehat{\boldsymbol{\Sigma}}(\mathbf{C} - \mathbf{C}^*) \right\}, \\
\boldsymbol{\Delta}_3 &= \mathbf{M}_2 \left\{ n^{-1}\mathbf{E}^T\mathbf{X} - (\mathbf{C} - \mathbf{C}^*)^T\widehat{\boldsymbol{\Sigma}} \right\}.
\end{aligned}$$

Next we move on to the construction of matrix $\mathbf{W}$. In view of the modified score function (ER.156), we can see that there is no intrinsic term similar to terms (6) and (10). This means that if we can find an appropriate matrix $\mathbf{W}$ to control the bias in the first term of (ER.156), we may be able to use the construction of matrices $\mathbf{M}$ and $\mathbf{W}$ to directly make inference without imposing the strong and weak orthogonality conditions (Conditions 4 and 5). However, we will show that such matrix $\mathbf{W}$ simply does not exist.

In view of matrices $\mathbf{M}_1$ and $\mathbf{M}_3$ given in (ER.161) and (ER.162), we have

$$\mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^{\mathrm{T}} = \alpha_1\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{u}_1^{\mathrm{T}} + \alpha_2\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}}, \quad \mathbf{M}_3\boldsymbol{v}_1\boldsymbol{u}_2^{\mathrm{T}} = \beta_1\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\boldsymbol{u}_2^{\mathrm{T}} + \beta_2\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\boldsymbol{u}_2^{\mathrm{T}}.$$

Then some simplifications yield that

$$\mathbf{I}_p - \mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^{\mathrm{T}} - \mathbf{M}_3\boldsymbol{v}_1\boldsymbol{u}_2^{\mathrm{T}} = \mathbf{I}_p - \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\left(\alpha_1\boldsymbol{u}_1 + \beta_1\boldsymbol{u}_2\right)^{\mathrm{T}} - \widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1\left(\alpha_2\boldsymbol{u}_1 + \beta_2\boldsymbol{u}_2\right)^{\mathrm{T}}.$$

Denote by $\mathbf{L}_2 = \left[-\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_1, -\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}_2\right] \in \mathbb{R}^{p\times 2}$ and $\mathbf{R}_2 = \left[\alpha_2\boldsymbol{u}_1 + \beta_2\boldsymbol{u}_2, \alpha_1\boldsymbol{u}_1 + \beta_1\boldsymbol{u}_2\right]^{\mathrm{T}} \in \mathbb{R}^{2\times p}$. Then we have

$$\mathbf{I}_p - \mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^{\mathrm{T}} - \mathbf{M}_3\boldsymbol{v}_1\boldsymbol{u}_2^{\mathrm{T}} = \mathbf{I}_p + \mathbf{L}_2\mathbf{R}_2.$$

By the Sherman–Morrison–Woodbury formula, $\mathbf{I}_p + \mathbf{L}_2\mathbf{R}_2$ is nonsingular if and only if $\mathbf{I}_2 + \mathbf{R}_2\mathbf{L}_2$ is nonsingular. However, it is easy to see that

$$\begin{aligned}
\mathbf{I}_2 + \mathbf{R}_2\mathbf{L}_2 &= \begin{bmatrix} 1 - (\alpha_2 z_{11} + \beta_2 z_{12}) & -(\alpha_2 z_{12} + \beta_2 z_{22}) \\ -(\alpha_1 z_{11} + \beta_1 z_{12}) & 1 - (\alpha_1 z_{12} + \beta_1 z_{22}) \end{bmatrix} \\
&= \begin{bmatrix} 1 - (\alpha_2 z_{11} + \beta_2 z_{12}) & -(\alpha_2 z_{12} + \beta_2 z_{22}) \\ 0 & 0 \end{bmatrix},
\end{aligned}$$

which shows that matrix $\mathbf{I}_p - \mathbf{M}_1\boldsymbol{v}_1\boldsymbol{u}_1^{\mathrm{T}} - \mathbf{M}_3\boldsymbol{v}_1\boldsymbol{u}_2^{\mathrm{T}}$ is in fact singular. Therefore, we see that such matrix $\mathbf{W}$ does not exist *without* incorporating the Stiefel manifold structure imposed by the SVD constraints.