Supplement to "Nonparametric covariance estimation for mixed longitudinal studies, with applications in midlife women's health"

Anru R. Zhang and Kehui Chen

Abstract

In this supplement, we provide the proofs for theoretical results of the paper.

A Proof of Theorem 1.

We prove Theorem 1 by steps. Some key technical procedures are postponed to Lemmas 4, 7, and 8.

Step 1 Since we can always rescale the time domain, let $\mathcal{T} = [0,1]$ throughout the proof without loss generality. We introduce some notations and prove basic properties in this step. Recall $T = \{t_1, \ldots, t_p\}$ is a regular grid on $\mathcal{T}$. Denote

$$\begin{aligned} &\Sigma = \mathrm{Cov}((X(t_1), \ldots, X(t_p))^\top), \quad \Sigma_0 = \mathrm{Cov}((Z(t_1), \ldots, Z(t_p))^\top) \in \mathbb{R}^{p\times p}, \\ &\Sigma_l = \mathrm{Cov}(X_{I_l}), \quad \Sigma_{0l} = \mathrm{Cov}(Z_{I_l}) \in \mathbb{R}^{|I_l|\times|I_l|}, \quad l = 1, \ldots, l_{\max}. \end{aligned} \tag{14}$$

Then

$$\Sigma = \Sigma_0 + \sigma^2 I_p, \quad \Sigma_l = \Sigma_{0l} + \sigma^2 I_{|I_l|},$$

Σ_l and Σ_{0l} are submatrices of Σ and Σ_0,

$$\Sigma_l = \Sigma_{[I_l, I_l]}, \quad \Sigma_{0l} = (\Sigma_0)_{[I_l,I_l]}, \quad l = 1, \ldots, l_{\max}. \tag{15}$$

For each subject k, recall $X_k(T) = (X_k(t_1), \ldots, X_k(t_p))^\top$ is the discretization of the sample path X_k. Given $G = G^{(r)} + G^{(-r)}$, we also decompose $\Sigma_0 = \Sigma_0^{(r)} + \Sigma_0^{-(r)}$, where $(\Sigma_0^{(r)})_{ij} = G^{(r)}(t_i, t_j), (\Sigma_0^{(-r)})_{ij} = G^{(-r)}(t_i, t_j)$. Suppose the eigenvalue decomposition of $\Sigma_0^{(r)}$ and $\Sigma_{0l}^{(r)}$ are

$$\Sigma_0^{(r)} = UDU^\top, \quad U \in \mathbb{O}_{p,r}, \quad D \in \mathbb{R}^{r\times r} \text{ is diagonal;} \tag{16}$$

$$\Sigma_{0l}^{(r)} = (\Sigma_0^{(r)})_{[I_l,I_l]}, \quad \Sigma_{0l}^{(-r)} = (\Sigma_0^{(-r)})_{[I_l,I_l]}, \quad l = 1, \ldots, l_{\max}.$$

Namely, $\Sigma_{0l}^{(r)}$ and $\Sigma_{0l}^{(-r)}$ are the submatrices of $\Sigma_0^{(r)}$ and $\Sigma_0^{(-r)}$. Then $\Sigma_{0l}^{(r)} + \Sigma_{0l}^{(-r)} = \Sigma_{0l}$ and $\Sigma_{0l}^{(r)} = U_{[I_l,:]} D U_{[I_l,:]}^\top$. It is also noteworthy that $\Sigma_{0l}^{(r)}$ and $\Sigma_{0l}^{(-r)}$ are not necessarily orthogonal, and $\Sigma_{0l}^{(r)}$ is not necessarily the best rank-r approximation of Σ_{0l}. We also define

$$\begin{aligned} &A = UD^{1/2} \in \mathbb{R}^{p\times r}, \quad A_l = U_{[I_l,:]} D^{1/2} \in \mathbb{R}^{|I_l|\times r}, l = 1, \dots, l_{\max} \\ &\text{then} \quad \Sigma_0^{(r)} = AA^\top, \quad \Sigma_{0l}^{(r)} = A_l A_l^\top. \end{aligned} \tag{17}$$

Especially, A and A_l can be seen as the factors of Σ_0 and Σ_{0l}.

Since $G(s_1, s_2)$ is Liptchitz, by Weyl's inequality (Weyl, 1949),

$$|\sigma_j(\Sigma_0)/p - \sigma_j(G)| \leq O(1/p), \quad \forall j; \quad \|\Sigma_0^{(-r)}\|_F/p \leq C\|G_0^{(-r)}\|_{HS} + O(1/p), \tag{18}$$

$$\|\Sigma_0\| \leq p \cdot \sigma_1(G) + O(1) \leq Cp, \tag{19}$$

$$\|\Sigma_0^{(r)}\|_F \leq \|\Sigma_0\|_F \leq p \cdot \|G\|_{HS} + O(1) \leq Cp. \tag{20}$$

We also have

$$\begin{aligned} \mathbb{E}\|X(T)\|_2^4 =& \mathbb{E}\left(\sum_{i=1}^p X(T(i))^2\right)^2 \overset{\text{Cauchy-Schwarz}}{\leq} p\sum_{i=1}^p \mathbb{E}X(T(i))^4 \\ \leq& Cp^2 \sup_t \mathbb{E}|X(t)|^4 \leq Cp^2. \end{aligned} \tag{21}$$

Let $\mathcal{I}_l = [t_{(l-1)a+1} - 1/(2p), \{t_{(l-1)a+b} + 1/(2p)\} \wedge 1]$, then $\mathcal{I}_l$ is the time sub-domain corresponding to the grid indices subset I_l. By the construction of I_l in (2), $I_l \cap I_{l+1} = \{la+1, \dots, (l-1)a+b\}$, so $|I_l \cap I_{l+1}| = b - a$, $\mathcal{L}(\mathcal{I}_l \cap \mathcal{I}_{l+1}) \geq \frac{b-a}{p} \geq \kappa$ (introduced in Assumption 2), thus

$$\sigma_r(G^{(r)}_{[\mathcal{I}_l\cap\mathcal{I}_{l+1},\mathcal{I}_l\cap\mathcal{I}_{l+1}]}) \geq c/\gamma \cdot \frac{\mathcal{L}(\mathcal{I}_l \cap \mathcal{I}_{l+1})}{\mathcal{L}(\mathcal{T})}$$

based on the assumption. Provided that $p > C\gamma$ for large constant $C > 0$, we further have

$$\begin{aligned} &\sigma_r\left(\Sigma^{(r)}_{0,[I_l\cap I_{l+1},I_l\cap I_{l+1}]}\right) = \sigma_r\left(G^{(r)}_{[\mathcal{I}_l\cap\mathcal{I}_{l+1},\mathcal{I}_l\cap\mathcal{I}_{l+1}]}\right)\cdot p + O(1) \\ \overset{\text{Assumption 2}}{\geq} & tr(G)p/\gamma \cdot \frac{\mathcal{L}(\mathcal{I}_l \cap \mathcal{I}_{l+1})}{\mathcal{L}(\mathcal{T})} + O(1) \geq cp/\gamma. \end{aligned} \tag{22}$$

The constant c here may depend on constant κ. Provided that $\|D\|/p = \sigma_1(\Sigma_0)/p \leq \sigma_1(G) + O(1/p) \leq C$, $A_{[I_l\cap I_{l+1},:]} A^\top_{[I_l\cap I_{l+1},:]} = \Sigma^{(r)}_{0,[I_l\cap I_{l+1},I_l\cap I_{l+1}]}$, we further have

$$\sigma_r\left(A_{[I_l\cap I_{l+1},:]}\right) = \sqrt{\sigma_r\left(\Sigma^{(r)}_{0,[I_l\cap I_{l+1},I_l\cap I_{l+1}]}\right)} \geq c\sqrt{p/\gamma}, \quad l = 1, \dots, l_{\max}. \tag{23}$$

$$\|A\| \leq \sqrt{\sigma_1(\Sigma_0)} \leq C\sqrt{p}. \tag{24}$$

Step 2 Our aim in this step is to develop a perturbation bound for $\hat{A}_l$, i.e. to characterize the distance between $\hat{A}_l$ and A_l for each $l = 1, \ldots, l_{\max}$. Recall $\Delta_l = \|\hat{\Sigma}_l - \Sigma_l\|_F$, $\lambda = \|\Sigma_0^{(-r)}\|_F$. By Lemma 4 and $b \geq 2r$,

$$\left\|\hat{A}_l\hat{A}_l^\top - \Sigma_{0l}^{(r)}\right\|_F = \left\|\hat{A}_l\hat{A}_l^\top - A_lA_l^\top\right\|_F \leq C\left(\Delta_l + \|\Sigma_0^{(-r)}\|_F\right), l = 1, \ldots, l_{\max}. \tag{25}$$

By Lemma 7, there exists $Q_l \in \mathbb{O}_r$ such that

$$\left\|\hat{A}_l - A_lQ_l\right\|_F \leq \frac{\|\hat{A}_l\hat{A}_l^\top - A_lA_l^\top\|_F}{\sqrt{\sigma_r(A_l)\sigma_r(\hat{A}_l)}} \leq \frac{C\left(\Delta_l + \lambda\right)}{\sqrt{\sigma_r(A_l)\sigma_r(\hat{A}_l)}} \wedge \sqrt{\|\hat{A}\|_F^2 + \|A\|_F^2}. \tag{26}$$

We analyze each term in (26) as follows. By (23),

$$\begin{aligned} &\sigma_r(A_l) \geq c\sqrt{p/\gamma}; \\ &\sigma_r^2(\hat{A}_l) = \sigma_r(\hat{A}_l\hat{A}_l^\top) \overset{(25)}{\geq} \sigma_r(\Sigma_{0l}^{(r)}) - C\left(\Delta_l + \lambda\right) \\ &\overset{(22)}{\geq} cp/\gamma - C\left(\Delta_l + \lambda\right). \end{aligned} \tag{27}$$

$$\begin{aligned} \|A_l\|_F^2 =& \mathrm{tr}(A_lA_l^\top) = \sum_{j=1}^r \sigma_j(\Sigma_{0l}^{(r)}) \leq \sqrt{r}\|\Sigma_{0l}^{(r)}\|_F \leq \sqrt{r}\|\Sigma_0^{(r)}\|_F \\ &\overset{(20)}{\leq} \sqrt{r}(p \cdot \|G\|_{HS} + O(1)) \leq Cp\sqrt{r}, \\ \|\hat{A}_l\|_F^2 =& \mathrm{tr}(\hat{A}_l\hat{A}_l^\top) \leq \sqrt{r}\|\hat{A}_l\hat{A}_l^\top\|_F \overset{(25)}{\leq} \sqrt{r}\|\Sigma_{0l}\|_F + C\sqrt{r}\left(p + \Delta_l + \lambda\right) \\ \leq& C\sqrt{r}\left(p + \Delta_l + \lambda\right). \end{aligned}$$

By combining the previous inequalities, we conclude that

$$\begin{aligned} \left\|\hat{A}_l - A_lQ_l\right\|_F \leq& \frac{C(\Delta_l + \lambda)}{\left(\frac{p}{\gamma}\left(\frac{p}{\gamma} - C(\Delta_l + \lambda)\right)_+\right)^{1/4}} \wedge C\left\{r^{1/4}(p^{1/2} + \lambda^{1/2} + \Delta_l^{1/2})\right\} \\ :=& \tilde{\Delta}_l, \end{aligned} \tag{28}$$

for $l = 1, \ldots, l_{\max}$ and some uniform constant $C > 0$. Here $(x)_+ = \max\{x, 0\}$ for any $x \in \mathbb{R}$.

Step 3 In this step, we assume (28) hold. Recall $\hat{O}_l$ is calculated sequentially. In this step, we study how the statistical error of $\hat{O}_l$ is accumulated based on (28) in this step. Ideally speaking, $\hat{O}_{l+1}$ can be seen as an estimation of $(Q_{l+1}^\top Q_1\hat{O}_1)$. Specifically, we aim to show that there exists a uniform constant $C > 0$ such that

$$\left\|Q_{l+1}\hat{O}_{l+1} - Q_1\hat{O}_1\right\|_F \leq \frac{C\tilde{\Delta}}{\sqrt{p/\gamma}}, \quad l = 0, \ldots, l_{\max} - 1. \tag{29}$$

and

$$\left\|\hat{A}_l\hat{O}_l - A_l Q_1 \hat{O}_1\right\|_F \le C\sqrt{\gamma}\tilde{\Delta}, \quad l = 1, \ldots, l_{\max}. \tag{30}$$

Here, $\tilde{\Delta} = \sum_{l=1}^{l_{\max}} \tilde{\Delta}_l$. First, for each $l = 1, \ldots, l_{\max} - 1$, we introduce

$$B_l^{(2)} := (\hat{A}_l)_{[(a+1):b,:]} \cdot \hat{O}_l \in \mathbb{R}^{(b-a)\times r}, \quad B_{l+1}^{(1)} := (\hat{A}_{l+1})_{[1:(b-a),:]} \in \mathbb{R}^{(b-a)\times r}.$$

Essentially, $B_l^{(2)}$ contains the last $(b-a)$ rows of $\hat{A}_l$ after rotation and $B_{l+1}^{(1)}$ contains the first $(b-a)$ rows of $\hat{A}_{l+1}$ before rotation. According to the proposed procedure (7),

$$\hat{O}_{l+1} = \underset{O\in\mathbb{O}_r}{\arg\min} \left\|B_l^{(2)} - B_{l+1}^{(1)} \cdot O\right\|_F. \tag{31}$$

Since $B_l^{(2)}$ and $B_{l+1}^{(1)}$ are submatrices of $\hat{A}_l$ and $\hat{A}_{l+1}$ respectively, they also satisfy

$$\begin{aligned} &\left\|B_l^{(2)} - A_{l,[(a+1):b,:]}Q_l\hat{O}_l\right\|_F = \left\|\hat{A}_{l,[(a+1):b,:]}\hat{O}_l - A_{l,[(a+1):b,:]}Q_l\hat{O}_l\right\|_F \\ \le &\left\|\hat{A}_l\hat{O}_l - A_l Q_l \hat{O}_l\right\|_F = \left\|\hat{A}_l - A_l Q_l\right\|_F \overset{(28)}{\le} \tilde{\Delta}_l. \end{aligned} \tag{32}$$

$$\begin{aligned} &\left\|B_{l+1}^{(1)} - A_{l+1,[1:(b-a),:]}Q_{l+1}\right\|_F = \left\|\hat{A}_{l+1,[1:(b-a),:]} - A_{l+1,[1:(b-a),:]}Q_{l+1}\right\|_F \\ \le &\left\|\hat{A}_{l+1} - A_{l+1}Q_{l+1}\right\|_F \overset{(28)}{\le} \tilde{\Delta}_{l+1}. \end{aligned} \tag{33}$$

More importantly, $A_{l,[(a+1):b,:]} = A_{l+1,[1:(b-a),:]} = A_{[I_l\cap I_{l+1},:]}$, as they actually represent the same submatrix of A. Then (31)–(33) and Lemma 1 yield

$$\begin{aligned} &\left\|\hat{O}_{l+1} - Q_{l+1}^\top Q_l\hat{O}_l\right\|_F = \left\|Q_{l+1}\hat{O}_{l+1} - Q_l\hat{O}_l\right\|_F \\ \le &\frac{2\left(\left\|B_l^{(2)} - A_{l,[(a+1):b,:]}Q_l\hat{O}_l\right\|_F + \left\|B_{l+1}^{(1)} - A_{l+1,[1:(b-a),:]}Q_{l+1}\right\|_F\right)}{\sigma_r(A_{I_l\cap I_{l+1}})} \\ \le &\frac{2(\tilde{\Delta}_l + \tilde{\Delta}_{l+1})}{\sigma_r(A_{[I_l\cap I_{l+1},:]})} \overset{(23)}{\le} \frac{C(\tilde{\Delta}_l + \tilde{\Delta}_{l+1})}{\sqrt{p/\gamma}}. \end{aligned} \tag{34}$$

Recall $\tilde{\Delta} = \sum_{l=1}^{l_{\max}} \tilde{\Delta}_l$. Thus,

$$\begin{aligned} &\left\|Q_l\hat{O}_l - Q_1\hat{O}_1\right\|_F \le \sum_{k=1}^{l-1}\left\|Q_{k+1}\hat{O}_{k+1} - Q_k\hat{O}_k\right\|_F \\ \le &\sum_{k=1}^{l-1}\frac{C(\tilde{\Delta}_l + \tilde{\Delta}_{l+1})}{\sqrt{p/\gamma}} = \frac{C\tilde{\Delta}}{\sqrt{p/\gamma}}, \end{aligned}$$

which has finished the proof for (29). Then

$$\begin{aligned}\left\|\hat{A}_l\hat{O}_l - A_lQ_1\hat{O}_1\right\|_F &\le \left\|\hat{A}_l\hat{O}_l - A_lQ_l\hat{O}_l\right\|_F + \left\|A_lQ_l\hat{O}_l - A_lQ_1\hat{O}_1\right\|_F \\ &\le \left\|\hat{A}_l - A_lQ_l\right\|_F + \|A_l\| \cdot \left\|Q_l\hat{O}_l - Q_1\hat{O}_1\right\|_F \\ &\overset{(28)(29)}{\le} \tilde{\Delta}_l + \|\Sigma_{0l}\|^{1/2} \cdot \frac{C\tilde{\Delta}}{\sqrt{p/\gamma}} \overset{(19)}{\le} C\sqrt{\gamma}\tilde{\Delta}\end{aligned} \tag{35}$$

for $l = 1, \ldots, l_{\max}$, which has finished the proof for (30).

Step 4 In this step, we develop the error bound from sequential aggregation based on (30). Recall

$$\tilde{A} \in \mathbb{R}^{p\times r}, \quad \tilde{A}_{[i,:]} = \frac{\sum_{l:i\in I_l} \hat{A}^*_{l,[i,:]}}{|\{l : i \in I_l\}|}. \tag{36}$$

The direct way to analyze $\tilde{A}$ is complicated. We instead consider the following half integers between $1/2$ and $p + 1/2$,

$$\begin{aligned}\mathcal{B} = &\{.5, a + .5, \ldots, (l_{\max} - 1)a + .5\} \\ &\cup \{b + .5, b + a + .5, \ldots, b + (l_{\max} - 1)a + .5\},\end{aligned} \tag{37}$$

and divide the whole index set $\{1, \ldots, p\}$ into pieces, say $K_1, \ldots, K_m$, by inserting "bars" with the half integers in (37). For example, when $p = 10, b = 5, a = 3$, then $\mathcal{B} = \{.5, 3.5, 6.5\} \cup \{5.5, 8.5, 11.5\}$, and $\{1, \ldots, 10\}$ is divided as the following subsets

$$K_1, \ldots, K_5 \quad = \quad \{1, 2, 3\}, \quad \{4, 5\}, \quad \{6\}, \quad \{7, 8\}, \quad \{9, 10\}.$$

Such a division has two important properties,

- Given $|\mathcal{B}| \le 2l_{\max}$ and $0.5 \in \mathbb{B}$, $\{1, \ldots, p\}$ is divided into at most $2l_{\max}$ intervals, so $m \le 2l_{\max}$.
- For any piece K_s and two indices $i, j \in K_s$, we must have

  $$\{l : i \in I_l\} = \{l : j \in I_l\},$$

  namely Indices i and j belong to the same set of sub-intervals $\{I_l\}$. Thus, we can further denote $J_{K_s} = \{l : i \in I_l, \forall i \in K_s\}$ as the sub-intervals that covers K_s. Then the following equality holds,

  $$\tilde{A}_{[K_s,:]} = \frac{1}{|J_{K_s}|} \sum_{l\in J_{K_s}} \hat{A}^*_{l,[K_s,:]}. \tag{38}$$

  Based on the definition of J_{K_s}, we also know

  $$\forall l \in J_{K_s}, \quad K_s \subseteq I_l. \tag{39}$$

Based on these two points, we analyze $\tilde{A}$ on each piece K_s and then aggregate as follows,

$$
\begin{aligned}
\left\|\tilde{A} - AQ_1\hat{O}_1\right\|_F =& \sqrt{\sum_{s=1}^{m} \left\|\tilde{A}_{[K_s,:]} - A_{[K_s,:]}Q_1\hat{O}_1\right\|_F^2} \\
\overset{(38)}{=}& \sqrt{\sum_{s=1}^{m} \left\|\frac{1}{|J_{K_s}|}\sum_{l\in J_{K_s}} \left(\hat{A}^*_{l,[K_s,:]} - A_{l,[K_s,:]}Q_1\hat{O}_1\right)\right\|_F^2}.
\end{aligned}
\tag{40}
$$

Now for each $s = 1, \dots, m$,

$$
\begin{aligned}
&\left\|\frac{1}{|J_{K_s}|}\sum_{l\in J_{K_s}} \left(\hat{A}^*_{l,[K_s,:]} - A_{[K_s,:]}Q_1\hat{O}_1\right)\right\|_F \\
\leq& \frac{1}{|J_{K_s}|}\sum_{l\in J_{K_s}} \left\|\left(\hat{A}^*_{l,[K_s,:]} - A_{[K_s,:]}Q_1\hat{O}_1\right)\right\|_F \\
\overset{(39)}{\leq}& \frac{1}{|J_{K_s}|}\sum_{l\in J_{K_s}} \left\|\left(\hat{A}_l\hat{O}_l - A_lQ_1\hat{O}_1\right)\right\|_F \overset{(30)}{\leq} C\gamma^{1/2}\tilde{\Delta}.
\end{aligned}
\tag{41}
$$

Combining (40) and (41), we obtain

$$
\left\|\tilde{A} - AQ_1\hat{O}_1\right\|_F \leq C\sqrt{m}\gamma^{1/2}\tilde{\Delta} \leq Cl_{\max}\gamma^{1/2}\tilde{\Delta}. \tag{42}
$$

By definition of $l_{\max}$, $b/p \leq Cl_{\max}^{-1}$, thus $l_{\max} \leq C$. Provided that $\sigma_{\max}(A) = \|\Sigma_0\|$ and $n^* \geq C\gamma^2$ for large constant C,

$$
\sigma_{\max}(\tilde{A}) \leq \sigma_{\max}(AQ_1\hat{O}_1) + \|\tilde{A} - AQ_1\hat{O}_1\|_F \overset{(42)}{\leq} Cp^{1/2} + C\gamma^{1/2}\tilde{\Delta}.
$$

Then the following inequality holds,

$$
\begin{aligned}
&\left\|\hat{\Sigma}_0 - \Sigma_0^{(r)}\right\|_F = \left\|\tilde{A}\tilde{A}^\top - AA^\top\right\|_F \\
\leq& \left\|\tilde{A}\tilde{A}^\top - AQ_1\hat{O}_1\tilde{A}^\top\right\|_F + \left\|AQ_1\hat{O}_1\tilde{A}^\top - AQ_1\hat{O}_1\hat{O}_1^\top Q_1^\top A^\top\right\|_F \\
\overset{(42)}{\leq}& \sigma_{\max}(\tilde{A}^\top)\cdot C\gamma^{1/2}\tilde{\Delta} + \sigma_{\max}(A)\cdot C\gamma^{1/2}\tilde{\Delta} \\
\leq& C\gamma^{1/2}\tilde{\Delta}\left(Cp^{1/2} + C\gamma^{1/2}\tilde{\Delta}\right).
\end{aligned}
\tag{43}
$$

Given $\Sigma_0 = \Sigma_0^{(r)} + \Sigma_0^{(-r)}$ and $\|\Sigma_0^{(-r)}\|_F \leq \lambda$, in summary, we have proved the upper bound

$$
\left\|\hat{\Sigma}_0 - \Sigma_0\right\|_F \leq C\gamma^{1/2}\tilde{\Delta}\left(Cp^{1/2} + C\gamma^{1/2}\tilde{\Delta}\right) + \lambda = C(\gamma p)^{1/2}\tilde{\Delta} + C\gamma\tilde{\Delta}^2 + \lambda.
$$

Step 5 It remains to develop the expected error upper bound for $\hat{\Sigma}_0$ and $\hat{G}$. Recall $\mathrm{Cov}(X_{I_l}) = \Sigma_l$. If $\hat{\Sigma}_l$ is calculated from complete samples by (3) in Step 2, we have

$$\begin{aligned}
\mathbb{E}\Delta_l^2 =& \mathbb{E}\left\|\frac{1}{n_l^*}\sum_{k\in J_l}((X_k)_{I_l} - \bar{X}_{I_l})((X_k)_{I_l} - \bar{X}_{I_l})^\top - \Sigma_l\right\|_F^2 \\
\leq& \frac{C}{n_l^*}\mathbb{E}\left\|((X_k)_{I_l} - \mu(I_l))((X_k)_{I_l} - \mu(I_l))^\top - \Sigma_l\right\|_F^2 \\
\leq& \frac{C}{n_l^*}\mathbb{E}\|(X_k)_{I_l}(X_k)_{I_l}^\top\|_F^2 \leq \frac{1}{n_l^*}\mathbb{E}\|X(T)\|_2^4 \overset{(21)}{\leq} \frac{Cp^2}{n_l^*}.
\end{aligned} \tag{44}$$

Under the incomplete observation scenario (Step 2'), we have

$$\begin{aligned}
&\mathbb{E}\Delta_l^2 \\
=& \sum_{i,j=1}^{|I_l|}\mathbb{E}\left\{\frac{\sum_{k:I_l(i),I_l(j)\in T_k}\left(X_k(I_l(i)) - \bar{X}(I_l(i))\right)\left(X_k(I_l(j)) - \bar{X}(I_l(j))\right)}{(n^*)_{i,j,l}}\right. \\
&\qquad \left. - \Sigma_{l,ij}\right\} \\
=& \sum_{i,j=1}^{|I_l|}\frac{\mathbb{E}\left\{\left(X(I_l(i)) - \mu(I_l(i))\right)\left(X(I_l(j)) - \mu(I_l(j))\right) - \Sigma_{l,ij}\right\}^2}{(n^*)_{i,j,l}} \\
\leq& C\sum_{i,j=1}^{|I_l|}\frac{\mathbb{E}\left\{X(I_l(i))\cdot X(I_l(j))\right\}^2}{(n^*)_{i,j,l}} \leq C\sum_{i,j=1}^{|I_l|}\frac{\mathbb{E}X(I_l(i))^4 + \mathbb{E}X(I_l(j))^4}{2n_{i,j,l}^*} \\
\leq& \frac{C|I_l|^2}{n^*}\sup_t \mathbb{E}X(t)^4 \leq \frac{Cp^2}{n^*}.
\end{aligned} \tag{45}$$

Now we analyze $\|\hat{\Sigma}_0 - \Sigma_0\|_F$ in two scenarios under the complete sample case (Step 2). The incomplete sample case (Step 2') similarly follows. Recall the definitions of $\tilde{\Delta}_l$ and $\tilde{\Delta}$,

$$\tilde{\Delta}_l = \frac{C(\Delta_l + \lambda)}{\left(\frac{p}{\gamma}\left(\frac{p}{\gamma} - C(\Delta_l+\lambda)\right)_+\right)^{1/4}} \wedge C\left\{r^{1/4}(p^{1/2} + \lambda^{1/2} + \Delta_l^{1/2})\right\}, \tilde{\Delta} = \sum_l \tilde{\Delta}_l.$$

Let

$$\begin{aligned}
B =& \{p/\gamma - C(\Delta_l + \lambda) \geq p/(2\gamma), \forall l = 1,\ldots,l_{\max}\} \\
=& \{C(\Delta_l + \lambda) \leq p/(2\gamma), \forall l = 1,\ldots,l_{\max}\}
\end{aligned} \tag{46}$$

be a "good" event. By Markov's inequality,

$$\mathbb{P}(B^c) \leq \sum_{l=1}^{l_{\max}}\frac{\mathbb{E}\{C(\Delta_l+\lambda)\}^2}{p^2/(2\gamma)^2} \overset{(44)(45)}{\leq} \frac{Cl_{\max}p^2/n^*}{p^2/\gamma^2} \leq C\gamma^2/n^*. \tag{47}$$

When B holds, note that $n^* \geq Cp$, we have

$$\begin{aligned}
&\mathbb{E}\left\|\hat{\Sigma}_0 - \Sigma_0\right\|_F 1_B \leq \mathbb{E}\left\{C(\gamma p)^{1/2}\tilde{\Delta} + Cr\tilde{\Delta}^2 + \lambda\right\} 1_B \\
\leq & C(\gamma p)^{1/2} \sum_{l=1}^{l_{\max}} \mathbb{E}\frac{(\Delta_l + \lambda)}{(p/\gamma)^{1/2}} + C\gamma \cdot \mathbb{E}\frac{\{\sum_{l=1}^{l_{\max}}(\Delta_l + \lambda)\}^2}{(p/\gamma)} + \lambda \\
\leq & C\gamma\sqrt{p^2/n^*} + C\gamma p/n^* \leq Cp\sqrt{\gamma^2/n^*}.
\end{aligned}$$

When B^c holds, given $n^* \geq \gamma^2, p \geq \gamma$, we have

$$\begin{aligned}
&\mathbb{E}\left\|\hat{\Sigma}_0 - \Sigma_0\right\|_F 1_{B^c} \leq \mathbb{E}C\left\{\gamma^{1/4}(p^{1/2} + \lambda^{1/2} + \sum_{l=1}^{l_{\max}} \Delta_l^{1/2})\right\} 1_{B^c} \\
\leq & CP(B^c)\gamma^{1/4}p^{1/2} + \sum_{l=1}^{l_{\max}} C\gamma^{1/4}\left(\mathbb{E}(\Delta_l^{1/2})^2\right)^{1/2} \cdot \left(\mathbb{E}1_B^2\right)^{1/2} \\
\leq & \frac{C\gamma^{9/4}p^{1/2}}{n^*} + Cl_{\max}\gamma^{1/4}\left(p/(n^*)^{1/2} \cdot \gamma^2/n^*\right)^{1/2} \\
\leq & \frac{C\gamma^{9/4}p^{1/2}}{n^*} + \frac{C\gamma^{5/4}p^{1/2}}{(n^*)^{3/4}} \leq Cp\sqrt{\gamma^2/n^*}.
\end{aligned}$$

In summary,

$$\mathbb{E}\left\|\hat{\Sigma}_0 - \Sigma_0\right\|_F \leq Cp\sqrt{\gamma^2/n^*}.$$

Finally, since Σ_0 is a p-by-p linear interpolation for G, we finally have

$$\left\|\hat{G} - G\right\|_{HS} \leq \frac{1}{p}\|\hat{\Sigma}_0 - \Sigma_0\|_F + O(p^{-1}) = O(\sqrt{\gamma^2/n^*} + p^{-1}).$$

□

B Proof of Proposition 1

The key of developing a sharper rate for $\hat{\Sigma}$ is on a better estimation bound for $\|\hat{A}_l - A_l O\|_F$, where $\hat{A}_l$ is the estimated factor computed in Step 3 of the proposed procedure. The essence of the sharper bound relies on the following lemma.

Lemma 2. *Suppose all conditions in Theorem 1 and Proposition 1 hold. Recall $\hat{A}_l$ is the estimation of the factor of each piece calculated in Step 3 in the proposed procedure. Then there exists a "good event" B_* (defined later in Equation 65) that happens with probability at least $1 - C\gamma r/n^*$, such that*

$$\mathbb{E}\min_{O \in \mathbb{O}_r} \|\hat{A}_l - A_l O\|_F^2 \cdot 1_{\{B^* \text{ holds}\}} \leq Cpr/n_l^*, \quad \forall l = 1, \ldots, l_{\max}.$$

Proof of Lemma 2. We assume $\mu(t) = 0$ without changing the covariance estimators essentially. Note that the sample covariance $\hat{\Sigma}_l$ is calculated in Step 2 as

$$\hat{\Sigma}_l = \frac{1}{n_l^*} \sum_{k \in J_l} \left((X_k)_{I_l} - (\bar{X})_{I_l} \right) \left((X_k)_{I_l} - \bar{X}_{I_l} \right)^\top .$$

The proof of this lemma is divided into steps.

Step 1 We introduce a series of notation in addition to the symbols in the proof of Theorem 1 here. Based on Karhunen-Loève decomposition, the continuous sample trajectory $X_k(t) = Z_k(t) + \epsilon_k(t)$ can be decomposed into three parts: the leading part of signal, the non-leading part of signal, and the noise:

$$X_k(t) = \sum_{j=1}^{r} \xi_{jk}\phi_j(t) + \sum_{j \geq r+1} \xi_{jk}\phi_j(t) + \epsilon_k(t), \quad k \in J_l. \tag{48}$$

Then, $\lambda_j(G) = \mathrm{Var}(\xi_{kj})$. Let $\bar{\xi}_{jk} = \xi_{jk}/\lambda_j^{1/2}(G)$ be the normalized score. We further define

$$S = \begin{bmatrix} \bar{\xi}_{11} & \cdots & \bar{\xi}_{1n} \\ \vdots & & \vdots \\ \bar{\xi}_{r1} & \cdots & \bar{\xi}_{rn} \end{bmatrix} \in \mathbb{R}^{r \times n}, \tag{49}$$

$$\Phi_l = \begin{bmatrix} (\phi_1)_{I_l(1)} \cdot \lambda_1^{1/2}(G) & \cdots & (\phi_1)_{I_l(1)} \cdot \lambda_r^{1/2}(G) \\ \vdots & & \vdots \\ (\phi_1)_{I_l(|I_l|)} \cdot \lambda_1^{1/2}(G) & \cdots & (\phi_1)_{I_l(|I_l|)} \cdot \lambda_r^{1/2}(G) \end{bmatrix} \in \mathbb{R}^{|I_l| \times r}$$

as the matrix of leading scores and the discretized loadings, respectively. Then, Φ_l matches the definition (15) in Theorem 1 as

$$\Phi_l \Phi_l^\top = \Sigma_{0l}^{(r)} \in \mathbb{R}^{|I_l| \times |I_l|}. \tag{50}$$

We further let $Z_k^{(-r)}(t) = \sum_{j \geq r+1} \xi_{jk}\phi_j(t)$ be the tail part of sample. By restricting (48) onto the index set I_l, one has

$$\begin{aligned} (X_k)_{I_l} =& \sum_{j=1}^{r} \xi_{jk}(\phi_j)_{I_l} + \sum_{j \geq r+1} \xi_{jk}(\phi_j)_{I_l} + (\epsilon_k)_{I_l} \\ =& \Phi_l S_{[:,k]} + (Z_k^{(-r)})_{I_l} + (\epsilon_k)_{I_l}, \quad k \in J_l. \end{aligned}$$

Based on the proof of Theorem 1, $\mathrm{rank}(\Sigma_{0l}^{(r)}) = r$ and

$$\sigma_j^2(\Phi) \overset{(50)}{=} \sigma_j(\Sigma_{01}^{(r)}) \overset{(22)}{=} p\sigma_j(G^{(r)}_{[\mathcal{I}_l, \mathcal{I}_l]}) + O(1), \quad j = 1, \ldots, r,$$

$$\|\Phi\|_F = (\mathrm{tr}(\Phi\Phi^\top))^{1/2} = \left(\mathrm{tr}(\Sigma_{0l}^{(r)})\right)^{1/2} \le \left(p \cdot \mathrm{tr}(G^{(r)}_{[\mathcal{I}_l,\mathcal{I}_l]}) + O(r)\right)^{1/2} \overset{\text{Assumption 2}}{\le} Cp^{1/2}. \tag{51}$$

In particularly,

$$\sigma_r(\Phi) = \sigma_r(A_l) = \sqrt{\sigma_r(\Sigma^{(r)}_{0,[I_l,I_l]})} \overset{(22)}{\ge} c\sqrt{p/\gamma}. \tag{52}$$

Recall the central goal of this proposition is to provide an upper bound for $\min_{O\in\mathbb{O}_r}\|\hat{A}_l - A_lO\|_F$. One can only show $\min_O \|\hat{A}_l - A_lO\|_F^2 \le Cp\gamma/n_l^*$ by directly applying Lemma 7 on $\hat{\Sigma}_{0l}$ and $\Sigma_{0l}^{(r)}$. Instead, we introduce a "bridge" covariance in this proof

$$\bar{\Sigma}_{0l} = \frac{1}{n_l^*}\sum_{k\in J_l}(\Phi_l S_{[:,k]})(\Phi_l S_{[:,k]})^\top = \frac{1}{n_l^*}\Phi_l S_{J_l} S_{J_l}^\top \Phi_l^\top \in \mathbb{R}^{|I_l|\times|I_l|}. \tag{53}$$

Let $\bar{A}_l = \Phi_l S_{J_l}/\sqrt{n_l^*}$. Then for all $Q \in \mathbb{O}_r$, we have

$$\begin{aligned}\min_{O\in\mathbb{O}_r}\|\hat{A}_l - A_lO\|_F \le& \min_{O\in\mathbb{O}_r}\left\{\|\hat{A}_l - \bar{A}_lQ\|_F + \|\bar{A}_lQ - A_lO\|_F\right\}\\ =&\|\hat{A}_l - \bar{A}_lQ\|_F + \min_{O\in\mathbb{O}_r}\|\bar{A}_l - A_lOQ^\top\|_F.\end{aligned}$$

By taking the infimum over $Q \in \mathbb{O}_r$, we obtain the following triangle inequality,

$$\min_{O\in\mathbb{O}_r}\|\hat{A}_l - A_lO\|_F \le \min_{O\in\mathbb{O}_r}\|\hat{A}_l - \bar{A}_lO\|_F + \min_{O\in\mathbb{O}_r}\|\bar{A}_l - A_lO\|_F. \tag{54}$$

In the next two steps, we give upper bounds for $\min_{O\in\mathbb{O}_r}\|\bar{A}_l - A_lO\|_F$ and $\min_{O\in\mathbb{O}_r}\|\hat{A}_l - \bar{A}_lO\|_F$, respectively.

Step 2 Since $\mathrm{rank}(S_{J_l}S_{J_l}^\top/n_l^*) = r$, we can further factorize

$$S_{J_l}S_{J_l}^\top/n_l^* = F_lF_l^\top$$

for some $F_l \in \mathbb{R}^{r\times r}$. Then,

$$\begin{aligned}\sigma_{\min}\left(S_{J_l}/\sqrt{n_l^*}\right) &= \sqrt{\sigma_{\min}\left(S_{J_l}S_{J_l}^\top/n_l^*\right)} = \sigma_{\min}(F_l),\\ \sigma_{\max}\left(S_{J_l}/\sqrt{n_l^*}\right) &= \sqrt{\sigma_{\max}\left(S_{J_l}S_{J_l}^\top/n_l^*\right)} = \sigma_{\max}(F_l).\end{aligned}$$

Suppose

$$F = U_F\Sigma_FV_F^\top, \quad U_F, V_F \in \mathbb{O}_r, \quad \Sigma_F \in \mathbb{R}^{r\times r}$$

is the singular value decomposition. Since Σ_F is diagonal, we have

$$\begin{aligned}\|\Sigma_F - I_{r\times r}\| \le& \max\left\{\sigma_{\max}(F_l) - 1, 1 - \sigma_{\min}(F_l)\right\}\\ =& \max\left\{\sigma_{\max}(S_{J_l}/\sqrt{n_l^*}) - 1, 1 - \sigma_{\min}(S_{J_l}/\sqrt{n_l^*})\right\}.\end{aligned} \tag{55}$$

We set $\bar{A}_l = \Phi_l F_l \in \mathbb{R}^{|I_l|\times r}$, then

$$\bar{\Sigma}_{0l} = \Phi_l F_l F_l^\top \Phi_l^\top = \bar{A}_l \bar{A}_l^\top.$$

On the other hand, we also recall that the true factor A_l satisfies

$$A_l A_l^\top = \Sigma_{0l}^{(r)} = \Phi_l \Phi_l^\top. \tag{56}$$

Since $\mathrm{rank}(\Sigma_{0l}^{(r)}) = r$ and both $A_l, \Phi_l \in \mathbb{R}^{|I_l|\times r}$, there exists an orthogonal matrix $V_l \in \mathbb{O}_r$ such that $\Phi_l = A_l V_l$. Therefore,

$$\begin{aligned}
&\min_{O\in\mathbb{O}_r} \|\bar{A}_l - A_l O\|_F^2 = \min_{O\in\mathbb{O}_r} \|\Phi_l F_l - \Phi_l V_l^\top O\|_F^2 \\
=& \min_{O\in\mathbb{O}_r} \|\Phi_l U_F \Sigma_F V_F^\top - \Phi_l V_l^\top O\|_F^2 \\
\leq& \|\Phi_l U_F \Sigma_F V_F^\top - \Phi_l V_l^\top V_l U_F V_F^\top\|_F^2 = \|\Phi_l U_F (\Sigma_F - I_{r\times r}) V_F^\top\|_F^2 \\
\leq& \|U_F(\Sigma_F - I)V_F^\top\|^2 \cdot \|\Phi_l\|_F^2 \overset{(51)}{\leq} Cp\|\Sigma_F - I\|^2 \\
\overset{(55)}{\leq}& Cp \max\left\{\sigma_{\max}(S_{J_l})/\sqrt{n_l^*} - 1, 1 - \sigma_{\min}(S_{J_l})/\sqrt{n_l^*}\right\}^2.
\end{aligned} \tag{57}$$

Let $T = \max\left\{\sigma_{\max}(S_{J_l})/\sqrt{n_l^*} - 1, 1 - \sigma_{\min}(S_{J_l})/\sqrt{n_l^*}\right\}$. Since $(\bar{\xi}_{1k}, \ldots, \bar{\xi}_{rk})$ is a sub-Gaussian vector, by random matrix theory (c.f. Theorem 5.39 in Vershynin (2010)),

$$\begin{aligned}
&\mathbb{P}\left(T \geq C\sqrt{r/n_l^*} + t\right) \\
=&\mathbb{P}\left(\max\left\{\sigma_{\max}(S_{J_l})/\sqrt{n_l^*} - 1, 1 - \sigma_{\min}(S_{J_l})/\sqrt{n_l^*}\right\} \geq C\sqrt{r/n_l^*} + t\right) \\
\leq& 1 - \mathbb{P}\Bigg(1 - C\sqrt{r/n_l^*} - t \leq \sigma_{\min}\left(S_{J_l}/\sqrt{n_l^*}\right) \leq \sigma_{\max}\left(S_{J_l}/\sqrt{n_l^*}\right) \\
&\qquad\qquad \leq 1 + C\sqrt{r/n_l^*} + t\Bigg) \\
\leq& C\exp(-cn_l^* t^2), \quad \forall t \geq 0.
\end{aligned} \tag{58}$$

Then,

$$
\begin{aligned}
\mathbb{E}T^2 &= \int_0^\infty 2t\mathbb{P}(T \geq t)dt \leq \int_0^{C\sqrt{r/n_l^*}} 2t \cdot 1 \cdot dt + \int_{C\sqrt{r/n_l^*}}^\infty 2t\mathbb{P}\left(T \geq t\right)dt \\
&\leq Cr/n_l^* + \int_0^\infty 2\left(t + C\sqrt{r/n_l^*}\right)\mathbb{P}\left(T \geq C\sqrt{r/n_l^*} + t\right) \\
&\leq \int_0^\infty C\left(C\sqrt{r/n_l^*} + t\right)\exp(-cn_l^* t^2)dt \\
&= Cr/n_l^* + C\sqrt{r/n_l^*} \cdot \sqrt{\frac{1}{n_l^*}} \cdot \int_0^\infty \exp(-ct^2)dt + \int_0^\infty \frac{C}{n_l^*} t\exp(-ct^2)dt \\
&\leq Cr/n_l^*.
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\min_{O\in\mathbb{O}_r}\|\bar{A}_l - A_l O\|_F^2 \\
&\overset{(57)}{\leq} Cp\mathbb{E}\max\left\{\sigma_{\max}(S_{J_l})/\sqrt{n_l^*} - 1, 1 - \sigma_{\min}(S_{J_l})/\sqrt{n_l^*}\right\}^2 \\
&\leq Cp\mathbb{E}T^2 \leq Cpr/n_l^*.
\end{aligned} \tag{59}
$$

Step 3 Then we consider $\min_{O\in\mathbb{O}_r}\|\hat{A}_l - \bar{A}_l O\|_F$ in this step. We apply Lemma 4 to $\hat{A}_l\hat{A}_l^\top$ and $\bar{\Sigma}_{0l} + \sigma^2 I_{|I_l|}$. Then,

$$
\begin{aligned}
\left\|\hat{A}_l\hat{A}_l^\top - \bar{\Sigma}_{0l}\right\|_F^2 &\leq C|I_l|/(|I_l| - r)\left(\|\hat{\Sigma}_l - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F^2\right) \\
&\leq C\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F^2.
\end{aligned} \tag{60}
$$

By setting $M = \bar{\Sigma}_{0l} = \bar{A}_l\bar{A}_l^\top, \hat{M} = \hat{A}_l\hat{A}_l^\top$ in Lemma 7, we have

$$
\begin{aligned}
\min_{O\in\mathbb{O}_r}\|\hat{A}_l - \bar{A}_l O\|_F^2 &\leq \frac{\|\hat{A}_l\hat{A}_l^\top - \bar{\Sigma}_{0l}\|_F^2}{\sigma_r(\bar{A}_l)\sigma_r(\hat{A}_l)} \wedge \left(\|\hat{A}_l\|_F^2 + \|\bar{A}_l\|_F^2\right) \\
&\leq \frac{C\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F^2}{\sigma_r(\bar{A}_l)\sigma_r(\hat{A}_l)} \wedge \left(\|\hat{A}_l\|_F^2 + \|\bar{A}_l\|_F^2\right).
\end{aligned} \tag{61}
$$

Step 4 In this step, we give an upper bound for $\mathbb{E}\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F^2$. First,

$$
\begin{aligned}
&\left\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\right\|_F \\
\leq &\left\| \frac{1}{n_l^*} \sum_{k \in J_l} \left((X_k)_{I_l} - \bar{X}_{I_l}\right)\left((X_k)_{I_l} - \bar{X}_{I_l}\right)^\top - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|} \right\|_F \\
= &\left\| \frac{1}{n_l^*} \sum_{k \in J_l} (X_k)_{I_l}(X_k)_{I_l}^\top - \bar{X}_{I_l}\bar{X}_{I_l}^\top - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|} \right\|_F \\
\leq &\Bigg\| \frac{1}{n_l^*} \sum_{k \in J_l} \left(\Phi_l S_{[:,k]} + (Z_k^{(-r)})_{I_l} + (\epsilon_k)_{I_l}\right)\left(\Phi_l S_{[:,k]} + (Z_k^{(-r)})_{I_l} + (\epsilon_k)_{I_l}\right)^\top \\
&- \frac{1}{n_l^*}(\Phi_l S_{[:,k]})(\Phi_l S_{[:,k]})^\top - \sigma^2 I_{|I_l|} \Bigg\|_F + \left\|\bar{X}_{I_l}\bar{X}_{I_l}^\top\right\|_F \\
\leq &2\left\| \frac{1}{n_l^*} \sum_{k \in J_l} \Phi_l S_{[:,k]} (Z_k^{(-r)})_{I_l}^\top \right\|_F + \frac{1}{n_l^*}\left\| \sum_{k \in J_l} (Z_k^{(-r)})_{I_l}(Z_k^{(-r)})_{I_l}^\top \right\|_F \\
&+ \left\| \frac{1}{n_l^*} \sum_{k \in J_l} (\epsilon_k)_{I_l}(\epsilon_k)_{I_l}^\top - \sigma^2 I_{|I_l|} \right\|_F \\
&+ \frac{2}{n_l^*}\left\| \sum_{k \in J_l} \left(\Phi_l S_{[:,k]} + (Z_k^{(-r)})_{I_l}\right)(\epsilon_k)_{I_l}^\top \right\|_F + \left\|\bar{X}_{I_l}\bar{X}_{I_l}^\top\right\|_F.
\end{aligned}
$$

We analyze each term separately.

- Since $S_{[:,k]}$ and $Z_k^{(-r)}(t)$ correspond to different scores in the Karhunen-Loève decomposition, they must be with mean zero and uncorrelated, which implies that $\mathbb{E}\Phi_l S_{[:,k]} Z_k^{(-r)}(t) = 0$. In addition, $\left\{\Phi_l S_{[:,k]} Z_k^{(-r)}(t)\right\}$ are i.i.d. for different k. Thus,

$$
\begin{aligned}
&\frac{1}{(n_l^*)^2}\mathbb{E}\left\| \sum_{k \in J_l} \Phi_l S_{[k,:]} (Z_k^{(-r)})_{I_l}^\top \right\|_F^2 \\
= &\frac{1}{(n_l^*)^2} \sum_{k \in J_l} \mathbb{E}\left\|\Phi_l S_{[k,:]} (Z_k^{(-r)})_{I_l}^\top\right\|_F^2 = \frac{1}{n_l^*}\mathbb{E}\|\Phi_l S_{[:,1]}(Z_1^{(-r)})_{I_l}^\top\|_F^2 \\
= &\frac{1}{n_l^*}\mathbb{E}\left\{\|\Phi_l S_{[:,1]}\|_2^2 \cdot \|(Z_1^{(-r)})_{I_l}\|_2^2\right\} \leq \frac{1}{n_l^*}\left(\mathbb{E}\|\Phi_l S_{[:,1]}\|_2^4 \cdot \mathbb{E}\|(Z_1^{(-r)})_{I_l}\|_2^4\right)^{1/2}.
\end{aligned}
$$

Here,

$$
\begin{aligned}
&\mathbb{E}\|(Z_1^{(-r)})_{I_l}\|_2^4 \\
=&\mathbb{E}\left(\sum_{i\in I_l} Z_1^{(-r)}(T(I_l(i)))^2\right)^2 \overset{\text{Cauchy-Schwarz}}{\leq} |I_l|\sum_{i\in I_l} Z_1^{(-r)}(T(I_l(i)))^4 \\
\leq&|I_l|^2\cdot\sup_t \mathbb{E}(Z_1^{(-r)}(t))^4 \leq Cp^2r/(n^*\gamma),
\end{aligned}
$$

where the last inequality is due to the assumption of this proposition.

$$
\begin{aligned}
&\mathbb{E}\|\Phi_l S_{[:,1]}\|_2^4 \\
=&\mathbb{E}\|X_{I_l} - (Z_k^{(-r)})_{I_l} - (\epsilon_k)_{I_l}\|_2^4 \\
\leq&C\left(\mathbb{E}\|X_{I_l}\|_2^4 + \|(Z_k^{(-r)})_{I_l}\|_2^4 + (\epsilon_k)_{I_l}\|_2^4\right) \\
\leq&C|I_l|\left(\sum_{i\in I_l}\mathbb{E}|X(T(I_l(i)))|^4 + \sum_{i\in I_l}\mathbb{E}|Z^{(-r)}(T(I_l(i)))|^4 + \sum_{i\in I_l}\mathbb{E}|\epsilon(T(I_l(i)))|^4\right) \\
\leq&C|I_l|^2\left(\sup_t \mathbb{E}X(t)^4 + \sup_t \mathbb{E}Z^{(-r)}(t)^4 + \mathbb{E}\epsilon^4\right) \leq Cp^2.
\end{aligned}
$$

Provided that $n^* \geq C\gamma^2 \geq Cr\gamma$, we have

$$
\frac{1}{(n_l^*)^2}\mathbb{E}\left\|\sum_{k\in J_l}\Phi_l S_{[k,:]}(Z_k^{(-r)})_{I_l}^\top\right\|_F^2 \leq \frac{C}{n_l^*}\left(p^2r/(n_l^*\gamma)\cdot p^2\right)^{1/2} \leq \frac{Cp^2r}{\gamma n_l^*}.
$$

- With the assumption that $(\mathbb{E}\epsilon^4)^{1/2} \leq Cr/\gamma$, we have

$$
\begin{aligned}
\mathbb{E}\|(\epsilon_k)_{I_l}\|_2^4 =&\mathbb{E}\left(\sum_{i\in I_l}\epsilon_k(T(I_l(i)))^2\right)^2 \overset{\text{Cauchy-Schwarz}}{\leq} |I_l|\cdot\sum_{i\in I_l}\mathbb{E}\epsilon_k(T(I_l(i)))^4 \\
\leq&Cp^2(r/\gamma)^2.
\end{aligned} \tag{62}
$$

$$
\sigma^4 = \left(\mathbb{E}\epsilon^2\right)^2 \leq \mathbb{E}\epsilon^4 \leq C(r/\gamma)^2.
$$

Given $\mathbb{E}(\epsilon_k)_{I_l}(\epsilon_k)_{I_l}^\top - \sigma^2 I_{|I_l|} = 0$, we have

$$
\begin{aligned}
&\mathbb{E}\left\|\frac{1}{n_l^*}\sum_{k\in J_l}(\epsilon_k)_{I_l}(\epsilon_k)_{I_l}^\top - \sigma^2 I_{|I_l|}\right\|_F^2 = \frac{1}{(n_l^*)^2}\sum_{k\in J_l}\mathbb{E}\left\|(\epsilon_k)_{I_l}(\epsilon_k)_{I_l}^\top - \sigma^2 I_{|I_l|}\right\|_F^2 \\
=&\frac{1}{n_l^*}\mathbb{E}\left\|\epsilon_{I_l}\epsilon_{I_l}^\top - \sigma^2 I_{|I_l|}\right\|_F^2 \leq \frac{2}{n_l^*}\left(\mathbb{E}\left\|\epsilon_{I_l}\epsilon_{I_l}^\top\right\|_F^2 + \left\|\sigma^2 I_{|I_l|}\right\|_F^2\right) \\
=&\frac{2}{n_l^*}\left(\mathbb{E}\|\epsilon_{I_l}\|_2^4 + \sigma^4 p\right) \leq \frac{2}{n_l^*}\left(|I_l|\cdot\sum_{i\in I_l}\mathbb{E}\epsilon_k(T(I_l(i)))^4 + \sigma^4 p\right) \leq \frac{Cp^2r}{n_l^*\gamma}.
\end{aligned}
$$

- With the assumption that $\sup_{t\in\mathcal{T}} \mathbb{E}X(t)^4 \leq C$, and $X_k(t) = Z_k(t) + \epsilon_k(t)$, we have

$$\begin{aligned}
\mathbb{E}\|(Z_k)_{I_l}\|_2^4 \leq & C\mathbb{E}\|(X_k)_{I_l}\|_2^4 + C\mathbb{E}\|(\epsilon_k)_{I_l}\|_2^4 \\
\leq & C\mathbb{E}\left(\sum_{i\in I_l} X_k(T(I_l(i)))^2\right)^2 + C\mathbb{E}\left(\sum_{i\in I_l} \epsilon_k(T(I_l(i)))^2\right)^2 \\
\overset{\text{Cauchy-Schwarz}}{\leq} & C|I_l| \cdot \sum_{i\in I_l} \mathbb{E}X_k(T(I_l(i)))^4 + C|I_l| \cdot \sum_{i\in I_l} \mathbb{E}\epsilon_k(T(I_l(i)))^4 \\
\leq & C|I_l|^2 \leq Cp^2.
\end{aligned} \tag{63}$$

Given $\mathbb{E}\epsilon_k = 0$, ϵ_k and $(S_k, Z_k^{(-r)}(t))$ are uncorrelated, we have $\mathbb{E}\left(\Phi_l S_{[:,k]} + (Z_k^{(-r)})_{I_l}\right)(\epsilon_k)_{I_l}^\top = 0$ and

$$\begin{aligned}
& \mathbb{E}\frac{1}{(n_l^*)^2}\left\|\sum_{k\in J_l}\left(\Phi_l S_{[:,k]} + (Z_k^{(-r)})_{I_l}\right)(\epsilon_k)_{I_l}^\top\right\|_F^2 \\
= & \mathbb{E}\frac{1}{n_l^*}\mathbb{E}\left\|(Z_k)_{I_k}\cdot(\epsilon_k)_{I_k}^\top\right\|_F^2 \\
= & \mathbb{E}\frac{1}{n_l^*}\|(Z_k)_{I_k}\|_2^2 \cdot \|(\epsilon_k)_{I_k}\|_2^2 \leq \frac{1}{n_l^*}\left(\mathbb{E}\|(Z_k)_{I_k}\|_2^4 \cdot \mathbb{E}\|(\epsilon_k)_{I_k}\|_2^4\right)^{1/2} \\
\overset{(62)(63)}{\leq} & \frac{Cp^2 r}{n_l^*\gamma}.
\end{aligned}$$

- Given $\mathbb{E}X_k(t) = 0$ and $X_1(t), \ldots, X_n(t)$ are independent,

$$\begin{aligned}
\mathbb{E}\left\|\bar{X}_{I_l}\bar{X}_{I_l}^\top\right\|_F^2 = & \frac{1}{(n_l^*)^4}\mathbb{E}\left\|\sum_{k\in J_l}(X_k)_{I_l}\right\|_2^4 \leq \frac{C}{(n_l^*)^2}\mathbb{E}\|(X_k)_{I_l}\|_2^4 \\
\leq & \frac{C}{(n_l^*)^2}\cdot|I_l|\cdot\sum_{i\in I_l}\mathbb{E}X_k(T(I_l(i)))^4 \leq \frac{C|I_l|^2}{(n_l^*)^2} \leq \frac{Cp^2}{(n_l^*)^2} \leq \frac{Cp^2}{n_l^*}\cdot\frac{r}{\gamma}.
\end{aligned}$$

In summary,

$$\mathbb{E}\left\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\right\|_F^2 \leq \frac{Cp^2 r}{\gamma n_l^*}. \tag{64}$$

Step 4 In this step, we further introduce the following "good" event,

$$B_* = \left\{\sigma_r^2(\hat{A}_l) \geq \sigma_r^2(A_l)/4, \sigma_r^2(\bar{A}_l) \geq \sigma_r^2(A_l)/2, \forall 1 \leq l \leq l_{\max}\right\}. \tag{65}$$

Then we develop the upper bound under this good event to finalize the proof. First, we aim to show B^* happens with high chance. By (60), we have

$$\begin{aligned}
& \left\|\hat{A}_l\hat{A}_l^\top - \bar{A}_l\bar{A}_l^\top\right\|_F \leq C\left\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\right\|_F \\
\Rightarrow \quad & \sigma_r^2(\hat{A}_l) \geq \sigma_r^2(\bar{A}_l) - C\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F.
\end{aligned}$$

By definition,

$$\sigma_r^2(\bar{A}_l) = \sigma_r\left(\frac{1}{n_l^*}\Phi_l S_{J_l} S_{J_l}^\top \Phi_l^\top\right) \geq \sigma_{\min}^2(\Phi_l)\sigma_r^2(S_{J_l}/\sqrt{n_l}).$$

In addition,

$$\sigma_r^2(A_l) \overset{(56)}{=} \sigma_{\min}^2(\Phi_l) \overset{(52)}{\geq} cp/\gamma.$$

Thus, B_* holds if the following two conditions hold for some small constant $c > 0$:

$$\forall l, \quad \sigma_r^2(S_{J_l}/\sqrt{n_l}) \geq 1/2, \quad \text{and} \quad \|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F \leq cp/\gamma. \tag{66}$$

By Markov's inequality and the sub-Gaussian random matrix tail bound (58),

$$\begin{aligned}
&\mathbb{P}(B_* \text{ holds}) \geq \mathbb{P}((66) \text{ holds})\\
\geq& 1 - \mathbb{P}\left(\exists l, \sigma_r^2(S_{J_l}/\sqrt{n_l}) < 1/2\right) + \mathbb{P}\left(\exists l, C\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F \geq cp/\gamma\right)\\
\overset{(58)}{\geq}& 1 - l_{\max}\exp(-cn_l^*) - l_{\max}\frac{\mathbb{E}\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|\|_F^2}\|_F^2}{c(p/\gamma)^2}\\
\overset{(64)}{\geq}& 1 - C\exp(-cn^*) - Cr\gamma/n^*\\
\geq& 1 - Cr\gamma/n^*.
\end{aligned} \tag{67}$$

When B_* holds, we must have

$$\sigma_r(A_l), \sigma_r(\bar{A}_l), \sigma_r(\hat{A}_l) \geq c\sqrt{p/\gamma}.$$

By combining (61), (64), and the previous inequality, we have for all $1 \leq l \leq l_{\max}$,

$$\begin{aligned}
\min_{O\in\mathbb{O}_r} \mathbb{E}\|\hat{A}_l - \bar{A}_l O\|_F^2 1_{\{B_* \text{ holds}\}} \leq& \frac{C\mathbb{E}\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F^2}{\sigma_r(\bar{A}_l)\sigma_r(\hat{A}_l)} \cdot 1_{\{B_* \text{ holds}\}}\\
\leq& \frac{C\mathbb{E}\|\hat{\Sigma}_{0l} - \bar{\Sigma}_{0l} - \sigma^2 I_{|I_l|}\|_F^2}{p/\gamma} \leq \frac{Cpr}{n_l^*}.
\end{aligned} \tag{68}$$

Finally, (54), (59), and (68) conclude the statement of this lemma. □

Now we consider the proof of Proposition 1. Similarly to the proof of Theorem 1, we develop an upper bound on the probability of the "bad case," i.e., B_* does not hold. To this end, we define $w \in \mathbb{R}^p, w_i = |\{l : i \in I_l\}|^{-1}$ as the weight in Equation (9). Then,

$$\begin{aligned}
\|\hat{\Sigma}_0\|_F \leq & \|\tilde{A}\tilde{A}^\top\|_F = \left\| \mathrm{diag}(w) \left(\sum_{l=1}^{l_{\max}} \hat{A}_l^*\right) \left(\sum_{l=1}^{l_{\max}} \hat{A}_l^*\right)^\top \mathrm{diag}(w) \right\|_F \\
\leq & \left\| \left(\sum_{l=1}^{l_{\max}} \hat{A}_l^*\right) \left(\sum_{l=1}^{l_{\max}} \hat{A}_l^*\right)^\top \right\|_F \leq l_{\max} \cdot \sum_{l=1}^{l_{\max}} \|\hat{A}_l^*(\hat{A}_l^*)^\top\|_F \\
= & l_{\max} \sum_{l=1}^{l_{\max}} \|\hat{A}_l \hat{A}_l^\top\|_F \overset{(6)}{\leq} l_{\max} \sum_{l=1}^{l_{\max}} \|\hat{\Sigma}_l\|_F
\end{aligned}$$

Then,

$$\begin{aligned}
\mathbb{E}\|\hat{\Sigma}_l\|_F^2 = & \mathbb{E} \left\| \frac{1}{n_l^*} \sum_{k \in J_l} ((X_k)_{I_l} - \bar{X}_{I_l})((X_k)_{I_l} - \bar{X}_{I_l})^\top \right\|_F^2 \\
= & \mathbb{E} \left\| \frac{1}{n_l^*} \sum_{k \in J_l} (X_k)_{I_l} (X_k)_{I_l}^\top - \bar{X}_{I_l} \bar{X}_{I_l} \right\|_F^2 \\
\overset{\text{Cauchy-Schwarz}}{\leq} & \frac{C}{n_l^*} \sum_{k \in J_l} \mathbb{E}\|(X_k)_{I_l}(X_k)_{I_l}^\top\|_F^2 + C\mathbb{E}\|\bar{X}_{I_l}\bar{X}_{I_l}\|_F^2 \\
= & \frac{C}{n_l^*} \sum_{k \in J_l} \mathbb{E}\|(X_k)_{I_l}\|_2^4 + C\mathbb{E}\|\bar{X}_{I_l}\|_2^4 \overset{(63)}{\leq} Cp^2.
\end{aligned}$$

$$\mathbb{E}\|\hat{\Sigma}_0 - \Sigma_0\|_F^2 \leq C\mathbb{E}\|\hat{\Sigma}_0\|_F^2 + C\|\Sigma_0\|_F^2 \leq Cp^2.$$

By Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}\|\hat{\Sigma}_0 - \Sigma_0\|_F 1_{B_*^c} \leq & \left(\mathbb{E}\|\hat{\Sigma}_0 - \Sigma_0\|_F^2 \cdot \mathbb{E}1_{B_*^c}^2\right)^{1/2} \\
\overset{(67)}{\leq} & \left(Cp^2 \cdot \gamma r/n^*\right)^{1/2}.
\end{aligned}$$

Similarly to Steps 3 - 5 and based on Lemma 2, one can develop the upper bound for $\|\hat{\Sigma}_0 - \Sigma\|_F$ on the "good event,"

$$\mathbb{E}\|\hat{\Sigma}_0 - \Sigma\|_F \cdot 1_{\{B_* \text{ holds}\}} \leq C\sqrt{p^2 r\gamma/n^*}.$$

Thus,

$$\mathbb{E}\|\hat{\Sigma}_0 - \Sigma\|_F = \mathbb{E}\|\hat{\Sigma}_0 - \Sigma\|_F 1_{\{B_* \text{ holds}\}} + \mathbb{E}\|\hat{\Sigma}_0 - \Sigma\|_F 1_{\{B_*^c \text{ holds}\}} \leq C\sqrt{p^2 r\gamma/n^*}.$$

Finally, since Σ_0 is a p-by-p linear interpolation for G, we finally have

$$\left\|\hat{G} - G\right\|_{HS} \leq \frac{1}{p}\|\hat{\Sigma}_0 - \Sigma_0\|_F + O(p^{-1}) = O(\sqrt{\gamma r/n^*} + p^{-1}),$$

which has finished the proof of Proposition 1. □

C Technical Lemmas

We collect all technical tools that were used in the main context of this paper in this section. We first provide the proof for Lemma 1, which provides an error bound for Wahba's problem (Wahba, 1965).

Proof of Lemma 1. Based on our assumption,

$$\begin{aligned}&\left\|A_2\hat{O} - A_1\right\|_F \leq \left\|A_2O_2^\top O_1 - A_1\right\|_F = \left\|A_2O_2^\top - A_1O_1^\top\right\|_F \\ \leq&\left\|A_2O_2^\top - A\right\|_F + \left\|A_1O_1^\top - A\right\|_F = \|A_2 - AO_2\|_F + \|A_1 - AO_1\|_F \\ \leq& a_1 + a_2.\end{aligned}$$

On the other hand,

$$\begin{aligned}&\left\|A_2\hat{O} - A_1\right\|_F \\ \geq& -\|A_1 - AO_1\|_F - \|A_2\hat{O} - AO_2\hat{O}\|_F + \|AO_1 - AO_2\hat{O}\|_F \\ \geq& -a_1 - a_2 + \sigma_{\min}(A)\|O_1 - O_2\hat{O}\|_F \\ \geq& -a_1 - a_2 + \lambda\|\hat{O} - O_2^\top O_1\|_F.\end{aligned}$$

Therefore,

$$\left\|\hat{O} - O_2^\top O_1\right\|_F \leq \frac{2(a_1 + a_2)}{\lambda}.$$

Finally,

$$\left\|O_2\hat{O} - I\right\|_F \leq \left\|O_2\hat{O} - O_1\right\|_F + \|O_1 - I\|_F \leq \|O_1 - I\|_F + \frac{2(a_1 + a_2)}{\lambda}.$$

□

The following lemma characterizes the least and largest singular value of semi-positive symmetric definite matrix factorization.

Lemma 3. *Suppose a positive semidefinite matrix $A \in \mathbb{R}^{p\times p}$ can be decomposed as $A = HDH^\top$. Here $D \in \mathbb{R}^{r\times r}$ is a non-negative diagonal matrix and $H \in \mathbb{R}^{p\times r}$ is a general matrix that is not necessarily orthogonal. Then*

$$\left(\max_i D_{ii}\right)\sigma_r^2(H) \geq \sigma_r(A) \geq \left(\min_i D_{ii}\right)\sigma_r^2(H),$$

$$\left(\max_i D_{ii}\right)\|H\|^2 \geq \|A\| \geq \left(\min_i D_{ii}\right)\|H\|^2.$$

Proof of Lemma 3. Suppose the singular value decomposition of H is $H = U_H D_H V_H^\top$, where $U_H \in \mathbb{O}_{p,r}$, $D_H \in \mathbb{R}^{r\times r}$ is diagonal with non-increasing non-negative entries, $V_H \in \mathbb{O}_{p,r}$. Then,

$$\begin{aligned}\|A\| = &\max_{\|u\|_2\le 1} u^\top A u = \max_{\|u\|_2\le 1} u^\top H D H^\top u \ge \left(U_{H,[:,1]}^\top H\right) D \left(H^\top U_{H,[:,1]}\right)\\ \ge&\sigma_r(D)\cdot \left\|H^\top U_{H,[:,1]}\right\|_2^2 = \min_{1\le i\le r} D_{ii}\cdot \sigma_1^2(H),\end{aligned}$$

$$\begin{aligned}\|A\| = &\max_{\|u\|_2\le 1} u^\top A u = \max_{\|u\|_2\le 1} u^\top H D H^\top u\\ \le& \max_{\|u\|_2\le 1} \|u\|_2\cdot \|H\|\cdot\|D\|\cdot\|H^\top\|\cdot\|u\|_2 = \left(\max_i D_{ii}\right)\|H\|^2.\end{aligned}$$

On the other hand, without loss of generality we assume $D_{rr} = \min_i D_{ii}$, then

$$\begin{aligned}\sigma_r(A) =&\sigma_r\left(U_H\left(D_H V_H^\top D V_H D_H\right)U_H^\top\right) = \sigma_r\left(D_H V_H^\top D V_H D_H\right)\\ =&\min_{u\in\mathbb{R}^r:\|u\|_2=1} u^\top D_H V_H^\top D V_H D_H u \le e_r^\top D_H V_H^\top D V_H D_H e_r\\ \le&\|D\|\cdot\|e_r^\top D_H V_H^\top\|_2^2 = \left(\max_i D_{ii}\right)\sigma_r^2(H),\end{aligned}$$

$$\begin{aligned}\sigma_r(A) =&\sigma_r\left(U_H\left(D_H V_H^\top D V_H D_H\right)U_H^\top\right) = \sigma_r\left(D_H V_H^\top D V_H D_H\right)\\ \ge&\sigma_{\min}^2(D_H V_H^\top)\sigma_{\min}(D) = \left(\min_i D_{ii}\right)\sigma_r^2(H).\end{aligned}$$

These have finished the proof for this lemma. □

Lemma 4. *Suppose $\Sigma = \Sigma_0 + \sigma^2 I \in \mathbb{R}^{b\times b}$. Here, Σ_0 is positive semi-definite, $\Sigma_0 = \Sigma_0^{(r)} + \Sigma_0^{(-r)}$, $\Sigma_0^{(r)}$ is a rank-r matrix. Suppose $\hat\Sigma$ is another rank-r symmetric matrix satisfying $\|\hat\Sigma - \Sigma\|_F \le \lambda$. Suppose $\hat U\hat D\hat U^\top$ is the eigenvalue decomposition and*

$$\hat\Sigma_0 = \sum_{i=1}^r \hat U_{[:,i]}\left\{(\hat D_{ii} - \hat\sigma^2)\vee 0\right\}(\hat U_{[:,i]})^\top,\ \text{where } \hat\sigma^2 = \left(\frac{1}{b-r}\sum_{i=r+1}^b \hat D_{ii}\right)\vee 0, \tag{69}$$

then the following inequality holds,

$$\left\|\hat\Sigma_0 - \Sigma_0\right\|_F \le C\sqrt{b/(b-r)}\left(\lambda + \|\Sigma_0^{(-r)}\|_F\right). \tag{70}$$

for uniform constant $C > 0$.

Proof of Lemma 4. Since $\hat\Sigma = \sum_{i=1}^b \hat U_{[:,i]}\hat D_{ii}\hat U_{[:,i]}^\top$ is the eigenvalue decomposition of $\hat\Sigma$, we also have the following eigenvalue decomposition for $\hat\Sigma - \sigma^2 I_{b\times b}$,

$$\hat\Sigma - \sigma^2 I_{b\times b} = \Sigma^{(r)} + \Sigma_0^{(-r)} + (\hat\Sigma - \Sigma) = \sum_{i=1}^b \hat U_{[:,i]}(\hat D_{ii} - \sigma^2)\hat U_{[:,i]}^\top.$$

Additionally, since $\Sigma_0^{(r)}$ is positive semi-definite, we can write down the eigenvalue decomposition $\Sigma_0^{(r)} = \sum_{i=1}^r U_{[:,i]} D_{ii} U_{[:,i]}^\top$, where $U \in \mathbb{O}_{b,r}$, $D \in \mathbb{R}^{r\times r}$ is non-negative diagonal. By Lemma 5,

$$\begin{aligned}
\|\{\hat{D}_{ii} - \sigma^2\}_{i=r+1}^b\|_2 = & \left(\sum_{i=r+1}^b \left(\hat{D}_{ii} - \sigma^2\right)^2\right)^{1/2} \\
\leq & \left(\sum_{i=1}^r (\hat{D}_{ii} - \sigma^2 - D_{ii})^2 + \sum_{i=r+1}^b (\hat{D}_{ii} - \sigma^2)^2\right)^{1/2} \\
\leq & \|\hat{\Sigma} - \sigma^2 I_{b\times b} - \Sigma_0^{(r)}\|_F = \|\Sigma_0^{(-r)} + \hat{\Sigma} - \Sigma\|_F \leq \lambda + \|\Sigma_0^{(-r)}\|_F.
\end{aligned} \tag{71}$$

Then

$$\begin{aligned}
\left|\hat{\sigma}^2 - \sigma^2\right| = & \left|\left(\frac{1}{b-r}\sum_{i=r+1}^b \hat{D}_{ii}\right) \vee 0 - \sigma^2\right| \leq \left|\left(\frac{1}{b-r}\sum_{i=r+1}^b \hat{D}_{ii}\right) - \sigma^2\right| \\
\leq & \frac{1}{b-r}\sum_{i=r+1}^b \left|\hat{D}_{ii} - \sigma^2\right| \leq \frac{1}{\sqrt{b-r}}\left(\sum_{i=r+1}^b (\hat{D}_{ii} - \sigma^2)^2\right)^{1/2} \\
\leq & \frac{1}{\sqrt{b-r}}\left(\lambda + \|\Sigma_0^{(-r)}\|_F\right).
\end{aligned} \tag{72}$$

Thus

$$\begin{aligned}
& \left\|(\hat{\Sigma} - \hat{\sigma}^2 I_{b\times b}) - \Sigma_0\right\|_F \leq \left\|\hat{\Sigma} - \Sigma\right\|_F + \|\hat{\sigma}^2 I_{b\times b} - \sigma^2 I_{b\times b}\|_F \\
\leq & \sqrt{b/(b-r)}\left(\lambda + \|\Sigma_0^{(-r)}\|_F\right) + \lambda.
\end{aligned} \tag{73}$$

On the other hand, note that $\hat{\Sigma}_0 - \hat{\sigma}^2 I_{b\times b} = \sum_{i=1}^b \hat{U}_{[:,i]}(\hat{\Sigma}_{ii} - \hat{\sigma}^2)\hat{U}_{[:,i]}^\top$ and $\hat{U}_{[:,1]}, \ldots, \hat{U}_{[:,b]}$ are orthonormal, the following inequality holds,

$$\begin{aligned}
\left\|\hat{\Sigma}_0 - \Sigma_0\right\|_F \leq & \left\|\hat{\Sigma}_0 - (\hat{\Sigma} - \hat{\sigma}^2 I_{b\times b})\right\|_F + \left\|\hat{\Sigma} - \hat{\sigma}^2 I_{b\times b} - \Sigma_0\right\|_F \\
\overset{(73)}{\leq} & \left\|\sum_{i=1}^r \hat{U}_{[:,i]}\left\{(\hat{D}_{ii} - \hat{\sigma}^2) \vee 0\right\}\hat{U}_{[:,i]}^\top - \sum_{i=r+1}^b \hat{U}_{[:.i]}(\hat{D}_{ii} - \hat{\sigma}^2)\hat{U}_{[:.i]}^\top\right\|_F \\
& + \sqrt{b/(b-r)}\left(\lambda + \|\Sigma_0^{(-r)}\|_F\right).
\end{aligned} \tag{74}$$

In particular,

$$\begin{aligned}
& \left\|\sum_{i=1}^r \hat{U}_{[:,i]}\left\{(\hat{D}_{ii} - \hat{\sigma}^2) \vee 0\right\}\hat{U}_{[:,i]}^\top - \sum_{i=r+1}^b \hat{U}_{[:.i]}(\hat{D}_{ii} - \hat{\sigma}^2)\hat{U}_{[:.i]}^\top\right\|_F^2 \\
= & \sum_{i=1}^r \left\{\{(\hat{D}_{ii} - \hat{\sigma}^2) \vee 0\} - (\hat{D}_{ii} - \hat{\sigma}^2)\right\}^2 + \sum_{i=r+1}^b \left|\hat{D}_{ii} - \hat{\sigma}^2\right|^2.
\end{aligned} \tag{75}$$

Here, we note that the i-th eigenvalue of Σ satisfies $\lambda_i(\Sigma) = D_{ii} + \sigma^2, D_{ii} \geq 0$ for $1 \leq i \leq r$, so

$$\begin{aligned}
&\sum_{i=1}^{r} \left\{ \{(\hat{D}_{ii} - \hat{\sigma}^2) \vee 0\} - (\hat{D}_{ii} - \hat{\sigma}^2) \right\}^2 = \sum_{1 \leq i \leq r} \left\{ \left(\hat{\sigma}^2 - \hat{D}_{ii}\right)_+ \right\}^2 \\
\leq & 3 \sum_{1 \leq i \leq r} \left\{ \left(\hat{\sigma}^2 - \sigma^2\right)_+ \right\}^2 + 3 \sum_{1 \leq i \leq r} \left\{ \left(\sigma^2 - (D_{ii} + \sigma^2)\right)_+ \right\}^2 \\
& + 3 \sum_{1 \leq i \leq r} \left\{ \left(\lambda_i(\Sigma) - \hat{D}_{ii}\right)_+ \right\}^2 \\
\overset{(72)}{\leq} & \frac{3r}{b-r} \left(\lambda + \|\Sigma_0^{(-r)}\|_F\right)^2 + 0 + 3 \sum_{i=1}^{b} \left\{ \lambda_i(\Sigma) - \hat{D}_{ii} \right\}^2 \\
\overset{\text{Lemma 5}}{\leq} & \frac{3b}{b-r} \left(\lambda + \|\Sigma_0^{(-r)}\|_F\right)^2 + 3\|\hat{\Sigma} - \Sigma\|_F^2 \leq \frac{3b}{b-r} \left(\lambda + \|\Sigma_0^{(-r)}\|_F\right)^2 + 3\lambda^2;
\end{aligned}$$

$$\begin{aligned}
&\sum_{r+1 \leq i \leq b} \left| \hat{D}_{ii} - \hat{\sigma}^2 \right|^2 \leq \sum_{r+1 \leq i \leq b} \left\{ 2 \left| \hat{D}_{ii} - \sigma^2 \right|^2 + 2 \left| \sigma^2 - \hat{\sigma}^2 \right|^2 \right\} \\
\overset{(71)(72)}{\leq} & 2 \left(\lambda + \|\Sigma_0^{(-r)}\|_F\right)^2 + \frac{2b}{b-r} \left(\lambda + \|\Sigma_0^{(-r)}\|_F\right)^2.
\end{aligned}$$

In summary, we have

$$\left\| \hat{\Sigma}_0 - \Sigma_0 \right\|_F \leq C \sqrt{b/(b-r)} \left(\lambda + \|\Sigma_0^{(-r)}\|_F\right).$$

for some uniform constant $C > 0$. $\square$

Lemma 5. *Suppose $A, B \in \mathbb{R}^{d \times d}$ are two symmetric matrices. $\lambda_j(A)$ and $\lambda_j(B)$ represent the j-th eigenvalues of A and B, respectively. Then*

$$\|A - B\|_F^2 \geq \sum_{j=1}^{d} \left(\lambda_j(A) - \lambda_j(B)\right)^2. \tag{76}$$

Proof of Lemma 5. Since

$$\begin{aligned}
\|A - B\|_F^2 =& \operatorname{tr}\left((A - B)^\top (A - B)\right) \\
=& \|A\|_F^2 + \|B\|_F^2 - 2\operatorname{tr}(A^\top B) = \sum_{j=1}^{d} \lambda_j^2(A) + \sum_{j=1}^{2} (B) - 2\operatorname{tr}(A^\top B),
\end{aligned}$$

we only need to show

$$\operatorname{tr}(A^\top B) \leq \sum_{j=1}^{d} \lambda_j(A) \lambda_j(B). \tag{77}$$

Suppose the eigenvalue decomposition of B is $B = UDU^\top$, with $D = \mathrm{diag}(\lambda_1(B), \ldots, \lambda_d(B))$. Let $U_{\{j\}} = U_{[:,1:j]}$, then

$$B = \sum_{j=1}^{d} U_{[:,j]} U_{[:,j]}^\top \cdot \lambda_j(B) = \sum_{j=1}^{d} U_{\{j\}} U_{\{j\}}^\top \cdot (\lambda_j(B) - \lambda_{j+1}(B)),$$

thus,

$$\begin{aligned}\mathrm{tr}(A^\top B) &= \sum_{j=1}^{d} \mathrm{tr}\left(A^\top U_{\{j\}} U_{\{j\}}^\top\right) \cdot (\lambda_j(B) - \lambda_{j+1}(B)) \\ &\overset{\text{Lemma 6}}{\leq} \sum_{j=1}^{d} \mathrm{tr}\left(\sum_{i=1}^{j} \lambda_i(A)\right) (\lambda_j(B) - \lambda_{j+1}(B)) \\ &= \sum_{j=1}^{d} \lambda_j(A)\lambda_j(B),\end{aligned}$$

which has finished the proof of this lemma. □

Lemma 6. *Suppose $A \in \mathbb{R}^{d \times d}$ is symmetric, $U_{\{j\}} \in \mathbb{O}_{d,j}$, then*

$$\mathrm{tr}\left(A^\top U_{\{j\}} U_{\{j\}}^\top\right) \leq \sum_{i=1}^{j} \lambda_i(A)$$

Proof of Lemma 6. Without loss of generality we can assume $A = \mathrm{diag}(\lambda_1(A), \ldots, \lambda_d(A))$. Since $U_{\{j\}} \in \mathbb{O}_{d,j}$, we have

$$0 \leq (U_{\{j\}})_{ii} \leq 1, \quad \sum_{i=1}^{d} (U_{\{j\}})_{ii} = j,$$

then by rearrangement inequality,

$$\mathrm{tr}\left(A^\top U_{\{j\}} U_{\{j\}}^\top\right) = \sum_{i=1}^{d} \lambda_i(A) \left(U_{\{j\}} U_{\{j\}}^\top\right)_{ii} \leq \sum_{i=1}^{j} \lambda_i(A).$$

□

The following lemma characterizes the square-root factorization perturbation. The proof involves Abel's summation identity in Lemmas 8 and 9, which is highly non-trivial.

Lemma 7. *Suppose $\hat{M}, M \in \mathbb{R}^{p \times r}$ are two matrices with the same dimension, then there exists an orthogonal matrix $O \in \mathbb{O}_r$ such that*

$$\left\|\hat{M} - MO\right\|_F^2 \leq \frac{\|\hat{M}\hat{M}^\top - MM^\top\|_F^2}{\sigma_r(M)\sigma_r(\hat{M})} \wedge \left(\|\hat{M}\|_F^2 + \|M\|_F^2\right). \tag{78}$$

Proof of Lemma 7. Suppose $M^\top \hat{M}$ has singular value decomposition: $M^\top \hat{M} = U\Sigma V^\top$, where $U_M, V_M \in \mathbb{O}_r$, $\Sigma \in \mathbb{R}^{r\times r}$. We will show that when $O = UV^\top$ (namely the solution to Wahba's problem), (78) holds.

Define $x_i = \sigma_i(\hat{M}), y_i = \sigma_i(M)$, $z_i = \sigma_i(\hat{M}^\top M)$, by Lemma 8, we know

$$x_1 \geq \cdots \geq x_r \geq 0, \quad y_1 \geq \cdots \geq y_r \geq 0, \quad z_1 \geq \cdots \geq z_r \geq 0,$$

and $\sum_{i=1}^s z_i \leq \sum_{i=1}^s x_i y_i$ for all $1 \leq s \leq r$. Then by both inequalities of Lemma 9,

$$\begin{aligned}
&\sum_{i=1}^r (x_i^4 + y_i^4 - 2z_i^2) - \sum_{i=1}^r \left(x_i^2 + y_i^2 - 2z_i\right) x_r y_r \\
\geq& 2\sum_{i=1}^r (x_i^2 y_i^2 - z_i^2) - 2\sum_{i=1}^r (x_i y_i - z_i) x_r y_r \geq 0.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\left\|\hat{M} - MO\right\|_F^2 = \mathrm{tr}\left(\hat{M}\hat{M}^\top + MM^\top - \hat{M}O^\top M^\top - MO\hat{M}^\top\right) \\
=& \|\hat{M}\|_F^2 + \|M\|_F^2 - 2\mathrm{tr}(O^\top M^\top \hat{M}) = \|\hat{M}\|_F^2 + \|M\|_F^2 - 2\mathrm{tr}(VU^\top U\Sigma V) \\
=& \sum_{i=1}^r \left(\sigma_i^2(\hat{M}) + \sigma_i^2(M) - 2\sigma_i(M^\top \hat{M})\right) = \sum_{i=1}^r (x_i^2 + y_i^2 - 2z_i);
\end{aligned}$$

$$\begin{aligned}
&\left\|\hat{M}\hat{M}^\top - MM^\top\right\|_F^2 \\
=& \mathrm{tr}\left(\hat{M}\hat{M}^\top\hat{M}\hat{M}^\top + MM^\top MM^\top - \hat{M}\hat{M}^\top MM^\top - MM^\top\hat{M}\hat{M}^\top\right) \\
=& \|\hat{M}\hat{M}^\top\|_F^2 + \|MM^\top\|_F^2 - 2\|M^\top\hat{M}\|_F^2 \\
=& \sum_{i=1}^r \left(\sigma_i^4(\hat{M}) + \sigma_i^4(M) - 2\sigma_i^2(M^\top\hat{M})\right) = \sum_{i=1}^r (x_i^4 + y_i^4 - 2z_i^2),
\end{aligned}$$

which means

$$\begin{aligned}
&\left\|\hat{M} - MO\right\|_F^2 \sigma_r(M)\sigma_r\hat{M} = x_r y_r \sum_{i=1}^r (x_i^2 + y_i^2 - 2z_i) \leq \sum_{i=1}^r \left(x_i^4 + y_i^4 - 2z_i^2\right) \\
\leq& \left\|\hat{M}\hat{M}^\top - MM^\top\right\|_F^2,
\end{aligned}$$

$$\min_{O\in\mathbb{O}_r} \left\|\hat{M} - MO\right\|_F^2 \leq \frac{\|\hat{M}\hat{M}^\top - MM^\top\|_F^2}{\sigma_r(M)\sigma_r(\hat{M})}.$$

In addition,

$$\min_O \left\|\hat{M} - MO\right\|_F^2 \leq \frac{1}{2}\left(\left\|\hat{M} - IM\right\|_F^2 + \left\|\hat{M} + IM\right\|_F^2\right) = \|\hat{M}\|_F^2 + \|M\|_F^2.$$

Therefore, we have finished the proof of this lemma. □

Lemma 8. *Suppose $M, \hat{M} \in \mathbb{R}^{p\times r}$ are two matrices of the same dimensions, we have the following inequality for Ky Fan s-norm of $M^\top \hat{M}$ (Fan, 1950) for any $s \geq 1$,*

$$\|M^\top \hat{M}\|_{k_s} = \sum_{i=1}^{s} \sigma_i(M^\top \hat{M}) \leq \sum_{i=1}^{s} \sigma_i(M)\sigma_i(\hat{M}).$$

Proof of Lemma 8. We first note the following property for Ky Fan norm (Fan, 1950),

$$\|X\|_{k_s} = \sum_{i=1}^{s} \sigma_i(X) = \max_{\substack{U\in\mathbb{O}_{p,s}\\ V\in\mathbb{O}_{r,s}}} \mathrm{tr}(U^\top X V).$$

Let $\hat{M} = U_{\hat{M}}\Sigma_{\hat{M}}V_{\hat{M}}^\top$ be the singular value decomposition, then $(\Sigma_{\hat{M}})_{ii} = \sigma_i(\hat{M})$. Now for any $U, V \in \mathbb{O}_{r,s}$,

$$\begin{aligned}
\mathrm{tr}\left(U^\top M^\top \hat{M} V\right) =& \mathrm{tr}\left(U^\top M^\top U_{\hat{M}}\Sigma_{\hat{M}}V_{\hat{M}}^\top V\right) = \mathrm{tr}\left(V_{\hat{M}}^\top V U^\top M^\top U_{\hat{M}}\Sigma_{\hat{M}}\right)\\
=& \sum_{i=1}^{r}(\Sigma_{\hat{M}})_{ii}\left(U_{\hat{M}}^\top M U V^\top V_{\hat{M}}\right)_{ii} = \sum_{i=1}^{r}\sigma_i(\hat{M})\left(U_{\hat{M}}^\top M U V^\top V_{\hat{M}}\right)_{ii}\\
=& \sum_{i=1}^{r}\left\{\left(\sigma_i(\hat{M}) - \sigma_{i+1}(\hat{M})\right)\sum_{j=1}^{i}\left(U_{\hat{M}}^\top M U V^\top V_{\hat{M}}\right)_{jj}\right\},
\end{aligned}$$

where is last equality is due to the Abel's summation formula[1]. Note that $U_{\hat{M}}^\top M U V^\top V_{\hat{M}}$ is a s-by-s projection of M, so it has smaller Ky Fan norms than M. Then

$$\begin{aligned}
\text{when } i \leq s, \quad \sum_{j=1}^{i}\left(U_{\hat{M}}^\top M V^\top U V_{\hat{M}}\right)_{jj} =& \sum_{j=1}^{i} e_j^\top U_{\hat{M}}^\top M U V^\top V_{\hat{M}} e_j\\
\leq& \|U_{\hat{M}}^\top M U V^\top V_{\hat{M}}\|_{k_i} \leq \sum_{j=1}^{i}\sigma_j(M);
\end{aligned}$$

$$\text{when } i > s, \quad \sum_{j=1}^{i}\left(U_{\hat{M}}^\top M U V^\top V_{\hat{M}}\right)_{jj} \leq \|U_{\hat{M}}^\top M U V^\top V_{\hat{M}}\| \leq \sum_{j=1}^{s}\sigma_j(M).$$

[1] https://en.wikipedia.org/wiki/Summation_by_parts

Thus,

$$
\begin{aligned}
\operatorname{tr}\left(U^\top M^\top \hat{M} V\right) &\leq \sum_{i=1}^{s}\left\{\left(\sigma_i(\hat{M})-\sigma_{i+1}(\hat{M})\right)\sum_{j=1}^{i}\left(U_{\hat{M}}^\top M U V^\top V_{\hat{M}}\right)_{jj}\right\} \\
&+\sum_{i=s+1}^{r}\left\{\left(\sigma_i(\hat{M})-\sigma_{i+1}(\hat{M})\right)\sum_{j=1}^{i}\left(U_{\hat{M}}^\top M U V^\top V_{\hat{M}}\right)_{jj}\right\} \\
&\leq \sum_{i=1}^{s}\left\{\left(\sigma_i(\hat{M})-\sigma_{i+1}(\hat{M})\right)\sum_{j=1}^{i}\sigma_j(M)\right\} \\
&+\sum_{i=s+1}^{r}\left\{\left(\sigma_i(\hat{M})-\sigma_{i+1}(\hat{M})\right)\sum_{j=1}^{s}\sigma_j(M)\right\} \\
&=\sum_{i=1}^{s}\sigma_i(\hat{M})\sigma_i(M),
\end{aligned}
$$

since U and V are arbitrarily chosen from $\mathbb{O}_{r,s}$, we have finished the proof for this lemma. □

Lemma 9. *Suppose $\{x_i\}_{i=1}^r, \{y_i\}_{i=1}^r, \{z_i\}_{i=1}^r$ are three sequences of non-negative values satisfying*

$$
x_1 \geq \cdots \geq x_r \geq 0, \quad y_1 \geq \cdots \geq y_r \geq 0, \quad z_1 \geq \cdots \geq z_r \geq 0;
$$
$$
\forall 1 \leq s \leq r, \quad \sum_{i=1}^{s} x_i y_i \geq \sum_{i=1}^{s} z_i.
$$

This means $x_1y_1 \geq z_1$, $x_1y_1 + x_2y_2 \geq z_1 + z_2$, but not necessarily $x_2y_2 \geq z_2$. Then, we must have the following two inequalities,

$$
\sum_{i=1}^{r}\left(x_i^2y_i^2 - z_i^2\right) - \sum_{i=1}^{r}(x_iy_i - z_i)x_ry_r \geq 0.
$$
$$
x_i^4 + y_i^4 - (x_i^2 + y_i^2)x_ry_r - 2x_i^2y_i^2 + 2x_iy_ix_ry_r \geq 0, \quad \forall 1 \leq i \leq r.
$$

Proof of Lemma 9. The key to the first inequality is via Abel's summation formula. First,

$$
\begin{aligned}
&\sum_{i=1}^{r}(x_i^2y_i^2 - z_i^2) + \sum_{i=1}^{r}(x_iy_i - z_i)x_ry_r - \left\{\sum_{i=1}^{r}(x_i^2y_i^2 - z_i^2) - \sum_{i=1}^{r}(x_iy_i - z_i)x_iy_i\right\} \\
=&\sum_{i=1}^{r}(x_iy_i - z_i)(x_iy_i - x_ry_r) = \sum_{i=1}^{r-1}\left\{(x_iy_i - x_{i+1}y_{i+1})\sum_{j=1}^{i}(x_jy_j - z_j)\right\} \geq 0.
\end{aligned}
$$

If we let $x_{r+1} = y_{r+1} = z_{r+1} = 0$, then

$$
\begin{aligned}
&\left\{\sum_{i=1}^{r}(x_i^2y_i^2 - z_i^2) - \sum_{i=1}^{r}(x_iy_i - z_i)x_iy_i\right\} = \sum_{i=1}^{r} z_i(x_iy_i - z_i) \\
=&\sum_{i=1}^{r}(z_i - z_{i+1})\sum_{j=1}^{i}(x_iy_i - z_i) \geq 0.
\end{aligned}
$$

By combining the two inequalities above, we have finished the proof for the first part. In addition, by some algebraic calculation we can show

$$\begin{aligned}
&x_i^4 + y_i^4 - (x_i^2 + y_i^2)x_r y_r - 2\left(x_i^2 y_i^2 - x_i y_i x_r y_r\right) \\
=&x_i^4 + y_i^4 - (x_i^2 + y_i^2)x_i y_i - 2\left(x_i^2 y_i^2 - x_i^2 y_i^2\right) + (x_i^2 + y_i^2 - 2x_i y_i)(x_i y_i - x_r y_r) \\
=&x_i^4 + y_i^4 - x_i^3 y_i - x_i y_i^3 + (x_i - y_i)^2 (x_i y_i - x_r y_r) \\
=&(x_i - y_i)^2 (x_i^2 + x_i y_i + y_i^2) + (x_i - y_i)^2 (x_i y_i - x_r y_r) \geq 0.
\end{aligned}$$

Therefore we have finished the proof for this lemma. □