Supplementary Material for "Dynamic Principal Component Analysis in High Dimensions"

S.1 Algorithm

Practitioners may use the retraction-based proximal gradient method (ManPG) (Chen et al., 2020) to solve our manifold optimization problem (5). Denote $\mathcal{M} = \mathbb{V}_{p,d}$ and $F(V) = -\mathrm{Tr}\{V(t)^{\mathrm{T}}\hat{\Sigma}(t)V(t)\} + \rho\|V(t)\|_1$ where $f(V) = -\mathrm{Tr}\{V(t)^{\mathrm{T}}\hat{\Sigma}(t)V(t)\}$ is smooth and its gradient is Lipschitz continuous with the Lipschitz constant L and $h(V) = \rho\|V(t)\|_1$. ManPG first computes a descent direction D_k (k-th step) by solving the following problem:

$$\begin{aligned} \min_D \quad & < \triangledown f(V_k), D > + \tfrac{1}{2t}\|D\|_F^2 + h(V_k + D) \\ s.t. \quad & D^{\mathrm{T}}V_k + V_k^{\mathrm{T}}D = 0, \end{aligned} \tag{S.1}$$

where V_k is obtained in the k-th iteration, $t > 0$ is a step size and D is a descent direction of F in the tangent space $T_{V_k}\mathcal{M}$. Based on the Lagrangian function and KKT system, we get that

$$E(\Lambda) = \mathcal{A}_k(D(\Lambda)) = 0, \tag{S.2}$$

where $\mathcal{A}_k(D) = D^{\mathrm{T}}V_k + V_k^{\mathrm{T}}D$, $D(\Lambda) = \mathrm{prox}_{th}(B(\Lambda)) - V_k$ with $B(\Lambda) = V_k - t(\triangledown f(V_k) - \mathcal{A}^*(\Lambda))$, $\mathcal{A}^*(\Lambda))$ denotes the adjoint operator of $\mathcal{A}_k$, where Λ is a $d \times d$ symmetric matrix. The semi-smooth Newton method (SSN) (Xiao et al., 2018) could be used to solve (S.2).

Retraction operation is an important concept in manifold optimization, see Absil et al. (2009) for more details. There are many common retractions for the Stiefel manifold, including exponential mapping, the polar decomposition and the Cayley transformation. For

Algorithm 1 Manifold Proximal Gradient Method (ManPG) for Solving (5).

Input: Initial point $V_0 \in \mathbb{V}_{p,d}$, $\delta \in (0,1)$, $\gamma \in (0,1)$, Lipschitz constant L

for $k \in 0, 1, \dots$ **do**
  Obtain D_k by solving the subproblem (S.1) with $t \in (0, 1/L]$;
  Set $\alpha = 1$
  while $F(\mathrm{Retr}_{V_k}(\alpha D_k)) > F(V_k) - \delta\alpha\|D_k\|_F^2$ **do**
    $\alpha = \gamma\alpha$
  end while
  Set $V_{k+1} = \mathrm{Retr}_{V_k}(\alpha D_k)$
end for

example, the exponential mapping (Edelman et al., 1998) is given by,

$$\mathrm{Retr}_V(tD) = \begin{bmatrix} V & Q \end{bmatrix} \exp\left(t \begin{bmatrix} V^{\mathrm{T}} D & -R^{\mathrm{T}} \\ R & 0 \end{bmatrix} \right) \begin{bmatrix} I_d \\ 0 \end{bmatrix},$$

where $QR = (I_p - VV^{\mathrm{T}})D$ is the unique QR factorization.

S.2 Additional simulation results

The results under $\sigma^2 = 1$ with $p = 100$ and 200 are provided in Tables S.1 and S.2. While the estimation errors of all considered methods under $\sigma^2 = 1$ are smaller than those obtained under $\sigma^2 = 3$, the proposed method still achieves consistently better results under both common and irregular designs. In particular, the DCM and DCM+ methods obtain comparable performance against the proposed approach when $p = 100$ and the number of total observations is large. However, their performance degrades when the dimension increases and the sampling frequency becomes small. In addition, the BJS and DT fail to obtain reasonable estimates with large errors.

Table S.1: Average integrated squared errors and standard deviations over 100 replications for different settings under the irregular design and $\sigma^2 = 1$.

Model		MISE_0	MISE	MISE_{DCM}	MISE_{DCM+}
p=100 n=100	$\bar{m} = 100$	.020 (.013)	.018(.010)	.019(.011)	.019(.010)
	$\bar{m} = 50$	.028(.015)	.026(.012)	.036(.048)	.036(.043)
	$\bar{m} = 20$	.051(.035)	.048(.034)	.092(.078)	.095(.100)
p=100 n=500	$\bar{m} = 20$	.010(.003)	.009(.003)	.010(.002)	.010(.002)
	$\bar{m} = 10$	.020(.003)	.019(.003)	.018(.003)	.018(.004)
	$\bar{m} = 4$	.036(.010)	.035(.009)	.053(.050)	.057(.079)
p=200 n=100	$\bar{m} = 100$	.026 (.014)	.023(.012)	.028(.028)	.025(.016)
	$\bar{m} = 50$	.034(.024)	.033(.025)	.096(.155)	.062(.076)
	$\bar{m} = 20$	.057(.039)	.057(.040)	.320(.253)	.313(.232)
p=200 n=500	$\bar{m} = 20$	.016(.001)	.015(.001)	.014(.003)	.014(.003)
	$\bar{m} = 10$	.022(.005)	.021(.005)	.041(.082)	.040(.079)
	$\bar{m} = 4$	.038(.011)	.037(.013)	.294(.198)	.299(.201)

Table S.2: Average integrated squared errors and standard deviations over 100 replications for different settings under the common design and $\sigma^2 = 1$.

Model		MISE_0	MISE	MISE_{DCM}	MISE_{DCM+}	MISE_{BJS}	MISE_{DT}
p=100 n=100	m=100	.024(.016)	.020(.012)	.024(.013)	.023(.012)	.252(.051)	.149(.041)
	m=50	.028(.021)	.024(.016)	.031(.021)	.031(.025)		
	m=20	.072(.066)	.072(.070)	.101(.117)	.075(.073)		
p=200 n=100	m=100	.025(.014)	.022(.011)	.038(.053)	.033(.043)	.411(.054)	.146(.041)
	m=50	.032(.031)	.030(.029)	.149(.153)	.124(.154)		
	m=20	.090(.086)	.091(.090)	.585(.292)	.532(.318)		

S.3 Proofs of main results

proof of Theorem 2. We provide the proofs of theoretical results using (4) under the common design. Recall that $\hat{\Sigma}(t) = \sum_{l=1}^{m} w_l S_l$, where $S_l = \sum_{i=1}^{n}(\mathbf{y}_{il} - \bar{\mathbf{y}}_l)(\mathbf{y}_{il} - \bar{\mathbf{y}}_l)^{\mathrm{T}}/n = \sum_{i=1}^{n} \mathbf{y}_{il}\mathbf{y}_{il}^{\mathrm{T}}/n - \bar{\mathbf{y}}_l\bar{\mathbf{y}}_l^{\mathrm{T}}$. Without loss of generality, we assume $\mu(t) = 0$ and it can be shown that $\bar{\mathbf{y}}_l\bar{\mathbf{y}}_l^{\mathrm{T}}$ is a higher order term that is negligible (Vu and Lei, 2013). Therefore, we will ignore this term and focus on the dominating $\sum_{i=1}^{n} \mathbf{y}_{il}\mathbf{y}_{il}^{\mathrm{T}}/n$ term in our proofs below.

Note that

$$\begin{aligned}
&\left\| \sum_{l=1}^{m} w_l \left(\sum_{i=1}^{n} y_{il} y_{il}^{\mathrm{T}}/n \right) - \Sigma(t) - \sigma^2 I_p \right\|_{\infty} \\
= &\max_{j,k} \left| \sum_{l=1}^{m} w_l \left\{ \frac{1}{n}\sum_{i=1}^{n} x_{ijl}x_{ikl} - \sigma_{jk}(t) + \frac{2}{n}\sum_{i=1}^{n} x_{ijl}\epsilon_{ikl} + \frac{1}{n}\sum_{i=1}^{n} \epsilon_{ijl}\epsilon_{ikl} - \sigma^2 1(j=k) \right\} \right| \\
\leq &\max_{j,k} \left| \frac{1}{n}\sum_{l=1}^{m}\sum_{i=1}^{n} w_l x_{ijl}x_{ikl} - \sigma_{jk}(t) \right| + \\
&\max_{j,k} \left| \frac{2}{n}\sum_{i=1}^{n}\sum_{l=1}^{m} w_l x_{ijl}\epsilon_{ikl} \right| + \max_{j,k} \left| \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{m} w_l \epsilon_{ijl}\epsilon_{ikl} - \sigma^2 1(j=k) \right|.
\end{aligned}$$

Denote

$$I = \max_{j,k} \left| \frac{1}{n}\sum_{l=1}^{m}\sum_{i=1}^{n} w_l x_{ijl}x_{ikl} - \sigma_{jk}(t) \right|.$$

Then,

$$\begin{aligned}
I &= \max_{j,k} \left| \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{m} w_l x_{ijl}x_{ikl} - \sigma_{jk}(t) \right| \\
&= \max_{j,k} \left| \frac{n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{m}\{\tilde{w}_l x_{ijl}x_{ikl} - E(\tilde{w}_l x_{ijl}x_{ikl})\}}{R_{0,c}R_{2,c} - R_{1,c}^2} + \frac{n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{m}\{E(\tilde{w}_l x_{ijl}x_{ikl}) - \tilde{w}_l\sigma_{jk}(t)\}}{R_{0,c}R_{2,c} - R_{1,c}^2} \right| \\
&\leq \frac{\max_{j,k}\left| n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{m}\{\tilde{w}_l x_{ijl}x_{ikl} - E(\tilde{w}_l x_{ijl}x_{ikl})\} \right|}{|R_{0,c}R_{2,c} - R_{1,c}^2|} + \frac{\max_{j,k}\left| \sum_{l=1}^{m}\{E(\tilde{w}_l x_{ijl}x_{ikl}) - \tilde{w}_l\sigma_{jk}(t)\} \right|}{|R_{0,c}R_{2,c} - R_{1,c}^2|} \\
&= I_1 + I_2,
\end{aligned}$$

where $\tilde{w}_l = R_{2,c}K_h(t_l - t) - R_{1,c}K_h(t_l - t)(t_l - t)$, and the second equality holds because $\sum_{l=1}^{m} \tilde{w}_l/(R_{0,c}R_{2,c} - R_{1,c}^2) = 1$.

Let $\tilde{w}'_{il} = n\tilde{w}_l$, we have

$$\tilde{w}'_{il} = R_2 K_h(t_{il} - t) - R_1 K_h(t_{il} - t)(t_{il} - t),$$

where $t_{il} = t_l$ and $R_\ell = \sum_{i=1}^n \sum_{l=1}^m K_h(t_{il} - t)(t_{il} - t)^\ell$. We have $\sum_{i=1}^n \sum_{l=1}^m \tilde{w}'_{il}(t_{il} - t) = 0$. Consequently,

$$I_1 = \frac{\max_{j,k} \left| \sum_{i=1}^n \sum_{l=1}^m \{\tilde{w}'_{il} x_{ijl} x_{ikl} - E(\tilde{w}'_{il} x_{ijl} x_{ikl})\} \right|}{n^2 |R_{0,c} R_{2,c} - R_{1,c}^2|}.$$

Using Lemma S.1 and similar arguments in Lemma 5 by replacing $\bar{m}$ with m, it is straightforward to obtain

$$\max_{j,k} \left| \sum_{i=1}^n \sum_{l=1}^m \{\tilde{w}'_{il} x_{ijl} x_{ikl} - E(\tilde{w}'_{il} x_{ijl} x_{ikl})\} \right| = O_p\{(\log p)^{1/2} (n^3 m^3 h^3 + n^3 m^4 h^4)^{1/2}\}.$$

According to Lemma S.1(a), we have $R_{0,c} R_{2,c} - R_{1,c}^2 \asymp m^2 h^2$. Thus,

$$I_1 = O_p \left\{ \left(\frac{\log p}{nmh} + \frac{\log p}{n} \right)^{1/2} \right\}.$$

Next we bound the approximation term I_2. Under the common fixed design, $\max_{0 \le j \le m} |t_{j+1} - t_j| \le C m^{-1}$. We have for each t, $|\{j : |t_j - t| \le h\}| = O(mh)$. Note that

$$\begin{aligned}
II &= \max_{j,k} \left| \sum_{l=1}^m E(\tilde{w}_l x_{ijl} x_{ikl}) - \tilde{w}_l \sigma_{jk}(t) \right| \\
&= \max_{j,k} \left| \sum_{l=1}^m \left[\tilde{w}_l \left\{ \sigma_{jk}(t) + \sigma_{jk}^{(1)}(t)(t_l - t) + \frac{\sigma_{jk}^{(2)}(\xi)}{2} (t_l - t)^2 \right\} - \tilde{w}_l \sigma_{jk}(t) \right] \right| \\
&\le \frac{\max_{j,k} |\sigma_{jk}^{(2)}(\xi)|}{2} \left| \sum_{l=1}^m \tilde{w}_l (t_l - t)^2 \right| = O(m^2 h^4).
\end{aligned}$$

where ξ is between t and t_l. The last inequality holds since $\sum_{l=1}^m \tilde{w}_l(t_l - t) = 0$. Note hat

$$\sum_{l=1}^m \tilde{w}_l (t_l - t)^2 = R_{2,c} \sum_{l=1}^m K_h(t_l - t)(t_l - t)^2 - R_{1,c} \sum_{l=1}^m K_h(t_l - t)(t_l - t)^3 = O(m^2 h^4).$$

Since $R_{0,c} R_{2,c} - R_{1,c}^2 \asymp m^2 h^2$ from Lemma S.1(a), $I_2 = O(h^2)$. Analogously, we can obtain the rates of other terms. Combining together and by Lemma 3, we complete the proof. □

Proof of Corollary 2. Recall that

$$d\{U(t), \hat{U}(t)\} = O_p\left[\left\{\left(\frac{\log p}{nmh} + \frac{\log p}{n}\right)^{1/2} + h^2\right\}^{1-q/2}\right].$$

A trade-off between the variance term $\{\log p/(nmh)\}^{1/2}$ and the bias term h^2 gives the optimal bandwidth $h = O[\{\log p/(nm)\}^{1/5}]$. However, by Assumption 9, the bandwidth h is at least of the order $1/m$. To illustrate the effect of the bandwidth on the final result, we define the function

$$r(h) = \left(\frac{\log p}{nmh} + \frac{\log p}{n}\right)^{1/2} + h^2.$$

Obviously, the function $r(h)$ decreases when $h \leq \{\log p/(nm)\}^{1/5}$ and increases when $h \geq \{\log p/(nm)\}^{1/5}$. Together with the condition $h \geq 1/m$, if $1/m > \{\log p/(nm)\}^{1/5}$, then the function $r(h)$ attains the minimum when $h = 1/m$. Otherwise if $1/m \leq \{\log p/(nm)\}^{1/5}$, then the function $r(h)$ attains the minimum at $h = \{\log p/(nm)\}^{1/5}$.

Based on the above analysis, we obtain the optimal bandwidth and the corresponding convergence rates under different sampling frequencies.

- If $\{\log p/(nm)\}^{1/5} \ll 1/m$, that is, $m/(n/\log p)^{1/4} \to 0$, then the optimal bandwidth is $h = O(1/m)$ and

$$d\{U(t), \hat{U}(t)\} = O_p\left\{\left(\frac{1}{m^2}\right)^{1-q/2}\right\}.$$

- If $\{\log p/(nm)\}^{1/5} \asymp 1/m$, that is $m/(n/\log p)^{1/4} \to C$, then the optimal bandwidth is $h = O\{(\log p/n)^{1/4}\} = O(1/m)$ and

$$d\{U(t), \hat{U}(t)\} = O_p\left\{\left(\frac{1}{m^2}\right)^{1-q/2}\right\}.$$

- If $\{\log p/(nm)\}^{1/5} \gg 1/m$, that is, $m/(n/\log p)^{1/4} \to \infty$, then the optimal bandwidth is $h = o\{(\log p/n)^{1/4}\}$ with $mh \to \infty$, and

$$d\{U(t), \hat{U}(t)\} = O_p\left\{\left(\frac{\log p}{n}\right)^{1/2-q/4}\right\}.$$

This completes the proof.

□

S.4 Auxiliary lemmas and proofs

Lemma S.1. *Under the common design, we have*

(a) $R_{\ell,c} \asymp mh^{\ell}$, $\ell = 0, 1, 2$. *Moreover, we have* $R_{2,c}R_{0,c} - R_{1,c}^2 \asymp mh^2$.

(b) $\sum_{l=1}^{m} E\left[\{R_{2,c}K_h(t_l - t) - R_{1,c}K_h(t_l - t)(t_l - t)\} x_{ijl}x_{ikl}\right]^2 = O(m^3h^3)$.

(c) $\sum_{l\neq l'} E\left(\tilde{w}_l x_{ijl}x_{ikl}\tilde{w}_{l'}x_{ijl'}x_{ikl'}\right) = O(m^4h^4)$, *where* $\tilde{w}_l = R_{2,c}K_h(t_l - t) - R_{1,c}K_h(t_l - t)(t_l - t)$.

Proof of Lemma S.1. (a) Recall that $R_{\ell,c} = \sum_{l=1}^{m} K_h(t_l - t)(t_l - t)^{\ell}$, $\ell = 0, 1, 2$. Under the common fixed design, $\max_{0\leq j\leq m} |t_{j+1} - t_j| \leq Cm^{-1}$. We have for each t, $|\{j : |t_j - t| \leq h\}| = O(mh)$. Then,

$$\sum_{l=1}^{m} K_h(t_l - t)(t_l - t)^{\ell} = \sum_{l:|t_l - t|\leq h} h^{-1}K\left(\frac{t_l - t}{h}\right)(t_l - t)^{\ell} \asymp mh^{\ell},$$

which is concluded from the properties of the kernel function in Assumption 5.

(b) Note that

$$\begin{aligned}
& \sum_{l=1}^{m} E\left[\{R_{2,c}K_h(t_l - t) - R_{1,c}K_h(t_l - t)(t_l - t)\}^2 x_{ijl}^2 x_{ikl}^2\right] \\
= & \sum_{l=1}^{m} E\{R_{2,c}^2 K_h^2(t_l - t)x_{ijl}^2 x_{ikl}^2\} - 2\sum_{l=1}^{m} E\{R_{1,c}R_{2,c}K_h^2(t_l - t)(t_l - t)x_{ijl}^2 x_{ikl}^2\} + \\
& \sum_{l=1}^{m} E\{R_{1,c}^2 K_h^2(t_l - t)(t_l - t)^2 x_{ijl}^2 x_{ikl}^2\} \\
= & I_1 + I_2 + I_3.
\end{aligned}$$

To bound the term I_1, observe that

$$\begin{aligned} I_1 &= \sum_{l=1}^m E\left[\left\{\sum_{l=1}^m K_h^2(t_l-t)(t_l-t)^4\right\} K_h^2(t_l-t)x_{ijl}^2 x_{ikl}^2\right] + \\ &\quad \sum_{l=1}^m E\left[\left\{\sum_{l\neq l'} K_h(t_l-t)(t_l-t)^2 K_h(t_{l'}-t)(t_{l'}-t)^2\right\} K_h^2(t_l-t)x_{ijl}^2 x_{ikl}^2\right] \\ &= O(m^2h^2+m^3h^3), \end{aligned}$$

since $|\{l : |t_l-t|\le h\}| = O(mh)$. Similarly, we have $I_2 = O(m^3h^3)$ and $I_3 = O(m^3h^3)$.

(c) Since $\sup_t Ex_j^4(t) < \infty$ for $j=1,\ldots,p$, so it suffices to prove that $\sum_{l\neq l'} E(\tilde{w}_l\tilde{w}_{l'}) = O(m^4h^4)$. Observe that

$$\begin{aligned} &\sum_{l\neq l'} E(\tilde{w}_l\tilde{w}_{l'}) \\ =& \sum_{l\neq l'} E\left[\{R_{2,c}K_h(t_l-t) - R_{1,c}K_h(t_l-t)(t_l-t)\}\{R_{2,c}K_h(t_{l'}-t) - R_{1,c}K_h(t_{l'}-t)(t_{l'}-t)\}\right] \\ =& \sum_{l\neq l'} E\{R_{2,c}^2 K_h(t_l-t)K_h(t_{l'}-t)\} - \sum_{l\neq l'} E\{R_{1,c}R_{2,c}K_h(t_l-t)(t_l-t)K_h(t_{l'}-t)\} - \\ & \sum_{l\neq l'} E\{R_{1,c}R_{2,c}K_h(t_{l'}-t)(t_{l'}-t)K_h(t_l-t)\} + \sum_{l\neq l'} E\{R_{1,c}^2 K_h(t_l-t)(t_l-t)K_h(t_{l'}-t)(t_{l'}-t)\} \\ =& M_1+M_2+M_3+M_4. \end{aligned}$$

Note that

$$\begin{aligned} M_1 &= \sum_{l\neq l'} E\left[\left\{\sum_{l=1}^m K_h(t_l-t)(t_l-t)^2\right\}^2 K_h(t_l-t)K_h(t_{l'}-t)\right] \\ &= \sum_{l\neq l'} E\left[\left\{\sum_{l=1}^m K_h^2(t_l-t)(t_l-t)^4\right\} K_h(t_l-t)K_h(t_{l'}-t)\right] + \\ &\quad \sum_{l\neq l'} E\left[\left\{\sum_{l\neq l'} K_h(t_l-t)(t_l-t)^2 K_h(t_{l'}-t)(t_{l'}-t)^2\right\} K_h(t_l-t)K_h(t_{l'}-t)\right] \\ &= O(m^3h^3+m^4h^4). \end{aligned}$$

In a similar way, other terms are quantified. Combining them together yields the final result. □

Lemma S.2. *(a)* $|a+b|^q \le |a|^q + |b|^q$, *for* $0<q<1$.

(b) Let $\Pi = UU^{\mathrm{T}}$ with $U^{\mathrm{T}}U = I_d$. Then, $\|\mathrm{vec}(\Pi)\|_q^q \le \|\mathrm{vec}(U)\|_q^{2q}$ for $0 < q < 1$.

Proof of Lemma S.2. (a) The inequality trivially holds either $a = 0$ or $b = 0$. Thus, we focus on the case where $a, b \neq 0$. It suffices to show that $(|a/b| + 1)^q \le |a/b|^q + 1$.

Define $f(x) = x^q + 1 - (x+1)^q$, $x > 0$. By the analysis of its derivative, we have $f^{(1)}(x) > 0$. Consequently, we have $f(x) \ge 0$ for $x > 0$, which completes the proof.

(b) Let $\tilde{u}_j$ is the j-th row of U. According to the Cauchy-Schwarz inequality,

$$|\tilde{u}_i^{\mathrm{T}}\tilde{u}_j| \le \|\tilde{u}_i\|\|\tilde{u}_j\| \le \|\tilde{u}_i\|_1\|\tilde{u}_j\|_1.$$

Denote by u_{jk} the entry in the j-th row, k-th column of U. Since $U^{\mathrm{T}}U = I_d$, we have $|u_{jk}| \le 1$. Then,

$$\|\mathrm{vec}(\Pi)\|_q^q = \sum_{i,j=1,\ldots,p} |\tilde{u}_i^{\mathrm{T}}\tilde{u}_j|^q \le \sum_{i,j} \|\tilde{u}_i\|_1^q\|\tilde{u}_j\|_1^q \le \left(\sum_{j=1}^{p}\sum_{k=1}^{d} |u_{jk}|^q\right)^2.$$

□

Proof of Lemma 1. Recall that $\Gamma = \Sigma + \sigma^2 I_p$, $\Pi = UU^{\mathrm{T}}$ and $\hat{\Pi}^0 = \hat{U}^0\hat{U}^{0\mathrm{T}}$. Note that by Corollary 4.1 in Vu and Lei (2013),

$$\begin{aligned}\frac{1}{2}\|\Pi - \hat{\Pi}^0\|_F^2 &\le \frac{<\Gamma - \hat{\Sigma}, \Pi - \hat{\Pi}^0> - \rho(\|\hat{U}^0\|_1 - \|U\|_1)}{\lambda_d - \lambda_{d+1}} \\ &\le \frac{\|\Gamma - \hat{\Sigma}\|_\infty \|\Pi - \hat{\Pi}^0\|_1 + \rho\|U\|_1 - \rho\|\hat{U}^0\|_1}{\lambda_d - \lambda_{d+1}}.\end{aligned}$$

By choosing $\rho \asymp \|\Gamma - \hat{\Sigma}\|_\infty$, we obtain $\|\Pi - \hat{\Pi}^0\|_F^2 \lesssim \|\Gamma - \hat{\Sigma}\|_\infty$. From proofs of Theorems 1 and 2, we have $\|\Gamma - \hat{\Sigma}\|_\infty = o_p(1)$. Therefore, $\|\Pi - \hat{\Pi}^0\|_F^2 = o_p(1)$.

It is clear that

$$\begin{aligned}\{j : \Pi_{jj} = 0, \hat{\Pi}_{jj}^0 \ge \gamma\} &\subseteq \{j : |\Pi_{jj} - \hat{\Pi}_{jj}^0| \ge \gamma\}, \\ \{j : \Pi_{jj} \ge 2\gamma, \hat{\Pi}_{jj}^0 < \gamma\} &\subseteq \{j : |\Pi_{jj} - \hat{\Pi}_{jj}^0| \ge \gamma\},\end{aligned}$$

and $\{j : \Pi_{jj} = 0, \hat{\Pi}_{jj}^0 \ge \gamma\} \cap \{j : \Pi_{jj} \ge 2\gamma, \hat{\Pi}_{jj}^0 < \gamma\} = \emptyset$. By Markov's inequality,

$$|\{j : \Pi_{jj} = 0, \hat{\Pi}_{jj}^0 \ge \gamma\}| + |\{j : \Pi_{jj} \ge 2\gamma, \hat{\Pi}_{jj}^0 < \gamma\}| \le \frac{\|\Pi - \hat{\Pi}^0\|_F^2}{\gamma^2}.$$

If $\gamma > \|\Pi - \hat{\Pi}^0\|_F$ and $\min_{j\in J}\Pi_{jj} \geq 2\gamma$, then $J = \hat{J}$. $\square$

Proof of Lemma 3. Denote $d^2\{U, \hat{U}\} := \hat{\epsilon}^2$ and $\delta = \lambda_d - \lambda_{d+1}$. The d-dimensional principal subspace of Γ is spanned by $U = \begin{pmatrix} U_J \\ 0 \end{pmatrix}$, where U_J is an orthonormal matrix. Then U_J spans the d-dimensional principal subspace of Γ_{JJ}. Define the event $I_n = \{\hat{J} = J\}$ with probability tending to 1 from Lemma 1. Note that $\hat{U} = \begin{pmatrix} \hat{U}_{\hat{J}} \\ 0 \end{pmatrix}$. On the event I_n, $\hat{U}_J$ is an optimal solution to the problem (6) and U_J is feasible. Then by Corollary 4.1 in Vu and Lei (2013), we have on the event I_n,

$$\begin{aligned}
\delta\hat{\epsilon}^2 &\leq\ < \hat{\Sigma}_{JJ} - \Gamma_{JJ}, \hat{U}_J\hat{U}_J^{\mathrm{T}} - U_JU_J^{\mathrm{T}} > -\rho(\|\hat{U}_J\|_1 - \|U_J\|_1) \\
&\leq \|\hat{\Sigma} - \Gamma\|_\infty \|\hat{U}_J\hat{U}_J^{\mathrm{T}} - U_JU_J^{\mathrm{T}}\|_1 - \rho(\hat{U}_J\|_1 - \|U_J\|_1) \\
&\leq \|\hat{\Sigma} - \Gamma\|_\infty \|\hat{U}_J\hat{U}_J^{\mathrm{T}} - U_JU_J^{\mathrm{T}}\|_1 + \rho\|\hat{U}_J - U_J\|_1 \\
&= I + II. \qquad \text{(S.3)}
\end{aligned}$$

Introduce the shorthand notation $\hat{\Delta} = \mathrm{vec}(\hat{U}_J\hat{U}_J^{\mathrm{T}} - U_JU_J^{\mathrm{T}})$, where $\mathrm{vec}(A)$ denotes the vector of length p^2 obtained by stacking the columns of a $p\times p$ matrix A. Using a standard argument of bounding l_1 norm by the l_q and l_2 norms, we have for all $\tau > 0$ and $0 < q \leq 1$,

$$\|\hat{\Delta}\|_1 \leq \tau^{-q/2}\|\hat{\Delta}\|_2\|\hat{\Delta}\|_q^{q/2} + \tau^{1-q}\|\hat{\Delta}\|_q^q, \qquad \text{(S.4)}$$

and when $q = 0$,

$$\|\hat{\Delta}\|_1 \leq \|\hat{\Delta}\|_2\|\hat{\Delta}\|_0^{1/2}. \qquad \text{(S.5)}$$

First we need to bound the term $\|\hat{\Delta}\|_q^q$. Denote by $\tilde{u}_l \in \mathbb{R}^d, l = 1, \ldots, p$, the l-th row of U.

- Case 1: $q = 0$. We have $\|\mathrm{vec}(U_J)\|_0 \leq dR_0$ since $U \in \mathcal{U}(0, R_0)$ and also $\|\mathrm{vec}(\hat{U}_J)\|_0 \leq dR_0$. Thus, we have $\|\hat{\Delta}\|_0 \leq d^2R_0^2$.

- Case 2: $0 < q \le 1$. Note that

$$
\begin{aligned}
\|\mathrm{vec}(\hat{U}_J\hat{U}_J^{\mathrm{T}} - U_JU_J^{\mathrm{T}})\|_q^q &= \sum_{i\in J}\sum_{j\in J}|\hat{\Pi}_{ij} - \Pi_{ij}|^q \\
&\le \sum_{i\in J}\sum_{j\in J}|\hat{\Pi}_{ij}|^q + \sum_{i\in J}\sum_{j\in J}|\Pi_{ij}|^q \\
&\le \|\mathrm{vec}(\hat{U}_J)\|_q^{2q} + \|\mathrm{vec}(U_J)\|_q^{2q} \\
&\le (\|\mathrm{vec}(\hat{U}_J - U_J)\|_q^q + \|\mathrm{vec}(U_J)\|_q^q)^2 + \|\mathrm{vec}(U_J)\|_q^{2q} \\
&\le 2\left(\sum_{i\in J}\frac{\Pi_{ii}^q}{2^q\gamma^q}\sum_{l=1}^{d}|\hat{u}_{il} - u_{il}|^q\right)^2 + 3\|\mathrm{vec}(U_J)\|_q^{2q} \\
&\le 2\left\{2^{-q}d\max_{i,l}\frac{|\hat{u}_{il} - u_{il}|^q}{\gamma^q}\sum_{i\in J}\Pi_{ii}^q\right\}^2 + 3\|\mathrm{vec}(U_J)\|_q^{2q} \\
&\le 2^{1-2q}d^4R_q^2 + 3d^2R_q^2,
\end{aligned}
$$

  where u_{ij} is the i-th row, j-th column element of U, the first and third inequalities hold because $|a+b|^q \le |a|^q + |b|^q$ for $0 < q < 1$ in Lemma S.2(a), the second inequality holds for $\|\mathrm{vec}(\Pi_J)\|_q^q \le \|\mathrm{vec}(U_J)\|_q^{2q}$ in Lemma S.2(b), the forth inequality holds due to the fact that $\min_{j\in J}\Pi_{jj} \ge 2\gamma$, and the last inequality holds since $\max_{i,l}|\hat{u}_{il} - u_{il}|/\gamma < 1$ from Lemma 1, $\|\mathrm{vec}(U_J)\|_q^q \le dR_q$ and $\sum_{i\in J}\Pi_{ii}^q \le \|\mathrm{vec}(U_J)\|_q^q$.

From the above discussion, we can similarly obtain $\|\mathrm{vec}(\hat{U}_J - U_J)\|_0 \le dR_0$ and $\|\mathrm{vec}(\hat{U}_J - U_J)\|_q^q \le 2^{-q}d^2R_q$ when $0 < q \le 1$. Define $C_q = Cd^2R_q$ for some constant $C > 0$. Then by (S.4) and (S.5), for $0 \le q \le 1$,

$$\|\hat{\Delta}\|_1 \le C_q\tau^{-q/2}\|\hat{\Delta}\|_2 + C_q^2\tau^{1-q}.$$

In a similar way, we have

$$\|\mathrm{vec}(\hat{U}_J - U_J)\|_1 \le C_q\tau^{-q/2}\|\hat{\Delta}\|_2 + C_q^2\tau^{1-q},$$

since $\|\mathrm{vec}(\hat{U}_J - U_J)\|_2 \le \|\hat{\Delta}\|_2$ according to Proposition 2.2 in Vu and Lei (2013).

Let $\tau = \rho/\delta$. On the event I_n, choose $\rho \asymp \|\hat{\Sigma} - \Gamma\|_\infty$, then by (S.3), we have

$$\hat{\epsilon}^2 \le 2\sqrt{2}C_q\tau^{1-q/2}\hat{\epsilon} + 2C_q^2\tau^{2-q}.$$

Hence, $\hat{\epsilon} \leq (2+\sqrt{2})C_q\tau^{1-q/2}$.

$\square$

Proof of Lemma 4. (a) Recall that $R_\ell = \sum_{i=1}^n \sum_{l=1}^{m_i} K_h(t_{il}-t)(t_{il}-t)^\ell$, $\ell = 0,1,2$. We have

$$ER_\ell = \sum_{i=1}^n \sum_{l=1}^{m_i} \int K(u_{il})u_{il}^\ell h^\ell f(u_{il})du_{il} = O(n\bar{m}h^\ell).$$

$$\begin{aligned} var(R_\ell) &\leq \sum_{i=1}^n \sum_{l=1}^{m_i} E\{K_h(t_{il}-t)(t_{il}-t)^\ell\}^2 \\ &= \sum_{i=1}^n \sum_{l=1}^{m_i} \int K^2(u_{il})u_{il}^{2\ell} h^{2\ell-1} f(u_{il})du_{il} = O(n\bar{m}h^{2\ell-1}). \end{aligned}$$

Combining the above arguments together yields the desired results.

(b) Note that

$$\begin{aligned} & E\left[\{R_2K_h(t_{il}-t) - R_1K_h(t_{il}-t)(t_{il}-t)\}^2 x_{ijl}^2 x_{ikl}^2\right] \\ = \; & E\{R_2^2K_h^2(t_{il}-t)x_{ijl}^2x_{ikl}^2\} - 2E\{R_1R_2K_h^2(t_{il}-t)(t_{il}-t)x_{ijl}^2x_{ikl}^2\} + E\{R_1^2K_h^2(t_{il}-t)(t_{il}-t)^2x_{ijl}^2x_{ikl}^2\} \\ = \; & I_1 + I_2 + I_3. \end{aligned}$$

To bound the term I_1, observe that

$$\begin{aligned} I_1 &= E\left[\left\{\sum_{i=1}^n \sum_{l=1}^{m_i} K_h^2(t_{il}-t)(t_{il}-t)^4\right\} K_h^2(t_{il}-t)x_{ijl}^2x_{ikl}^2\right] + \\ & \quad E\left[\left\{\sum_{(i,l)\neq(i',l')} K_h(t_{il}-t)(t_{il}-t)^2K_h(t_{i'l'}-t)(t_{i'l'}-t)^2\right\} K_h^2(t_{il}-t)x_{ijl}^2x_{ikl}^2\right] \\ &= O(n^2\bar{m}^2h^3), \end{aligned}$$

by the change of variables and Assumptions 2 and 5. Similarly, we have $I_2 = O(n^2\bar{m}^2h^3)$ and $I_3 = O(n^2\bar{m}^2h^3)$.

(c) Since $\sup_t Ex_j^4(t) < \infty$ for $j = 1, \ldots, p$, so it suffices to prove that $E(\tilde{w}_{il}\tilde{w}_{i'l'}) =$

$O(n^2\bar{m}^2h^4)$. Observe that

$$\begin{aligned} E(\tilde{w}_{il}\tilde{w}_{i'l'}) &= E\left[\{R_2K_h(t_{il}-t) - R_1K_h(t_{il}-t)(t_{il}-t)\}\{R_2K_h(t_{i'l'}-t) - R_1K_h(t_{i'l'}-t)(t_{i'l'}-t)\}\right] \\ &= E\{R_2^2K_h(t_{il}-t)K_h(t_{i'l'}-t)\} - E\{R_1R_2K_h(t_{il}-t)(t_{il}-t)K_h(t_{i'l'}-t)\} - \\ &\quad E\{R_1R_2K_h(t_{i'l'}-t)(t_{i'l'}-t)K_h(t_{il}-t)\} + E\{R_1^2K_h(t_{il}-t)(t_{il}-t)K_h(t_{i'l'}-t)(t_{i'l'}-t)\} \\ &= M_1 + M_2 + M_3 + M_4. \end{aligned}$$

Note that

$$\begin{aligned} M_1 &= E\left[\left\{\sum_{i=1}^{n}\sum_{l=1}^{m_i} K_h(t_{il}-t)(t_{il}-t)^2\right\}^2 K_h(t_{il}-t)K_h(t_{i'l'}-t)\right] \\ &= E\left[\left\{\sum_{i=1}^{n}\sum_{l=1}^{m_i} K_h^2(t_{il}-t)(t_{il}-t)^4\right\} K_h(t_{il}-t)K_h(t_{i'l'}-t)\right] + \\ &\quad E\left[\left\{\sum_{(i,l)\neq(i',l')} K_h(t_{il}-t)(t_{il}-t)^2K_h(t_{i'l'}-t)(t_{i'l'}-t)^2\right\} K_h(t_{il}-t)K_h(t_{i'l'}-t)\right] \\ &= O(n^2\bar{m}^2h^4). \end{aligned}$$

Other terms can be quantified similarly and the details are omitted. Combining them together yields the final result. □

Proof of Lemma 5. Define $W_{ijk} := \sum_{l=1}^{m_i}\{\tilde{w}_{il}x_{ijl}x_{ikl} - E(\tilde{w}_{il}x_{ijl}x_{ikl})\}$, for $1 \le i \le n, 1 \le l \le m_i, 1 \le j,k \le p$. Then from Lemma 4, we have

$$\begin{aligned} var\{W_{ijk}(t,h)\} &\le \sum_{l=1}^{m_i} E\left(\tilde{w}_{il}^2x_{ijl}^2x_{ikl}^2\right) + \sum_{l\neq l'} E\left(\tilde{w}_{il}x_{ijl}x_{ikl}\tilde{w}_{il'}x_{il'j}x_{il'k}\right) \\ &= O(m_in^2\bar{m}^2h^3 + m_i^2n^2\bar{m}^2h^4). \end{aligned}$$

Let $b_{ijk} > var(W_{ijk})$ and $b_{ijk} = O\{var(W_{ijk})\} = O(m_in^2\bar{m}^2h^3 + m_i^2n^2\bar{m}^2h^4)$. If there exists a sufficiently small constant $a > 0$ such that $Ee^{aW_{ijk}} < \infty$ holds, then $Ee^{aW_{ijk}} \le e^{a^2b_{ijk}}$ by Theorem 2.13 in Wainwright (2019). To see this, when n is large enough, by the Jensen's inequality and the Cauchy-Schwarz inequality we have

$$
\begin{aligned}
Ee^{a\sum_{l=1}^{m_i}\tilde{w}_{il}x_{ijl}x_{ikl}} &\leq 1+\sum_{r=1}^{\infty}\frac{a^r}{r!}E\left(\sum_{l=1}^{m_i}\tilde{w}_{il}x_{ijl}x_{ikl}\right)^r \\
&\leq 1+\sum_{r=1}^{\infty}\frac{a^r m_i^{r-1}}{r!}\sum_{l=1}^{m_i}E|\tilde{w}_{il}x_{ijl}x_{ikl}|^r \\
&\leq 1+\sum_{r=1}^{\infty}\frac{a^r m_i^{r-1}}{r!}\sum_{l=1}^{m_i}\{E(|\tilde{w}_{il}|^{1/2}x_{ijl})^{2r}+E(|\tilde{w}_{il}|^{1/2}x_{ikl})^{2r}\} \\
&\leq \frac{1}{m_i}\sum_{l=1}^{m_i}\left(1+\sum_{r=1}^{\infty}\frac{(am_i)^r}{r!}E(|\tilde{w}_{il}|x_{ijl}^2)^r+\sum_{r=1}^{\infty}\frac{(am_i)^r}{r!}E(|\tilde{w}_{il}|x_{ikl}^2)^r\right) \\
&\leq \frac{1}{m_i}\sum_{l=1}^{m_i}\left(Ee^{am_i|\tilde{w}_{il}|x_{ijl}^2}+Ee^{am_i|\tilde{w}_{il}|x_{ikl}^2}\right).
\end{aligned}
$$

Denote the event $E_n=\{E_{\mathrm{x}}e^{a\sum_{l=1}^{m_i}\tilde{w}_{il}x_{ijl}x_{ikl}}<\infty\}$, where E_{x} means that the expectation is taken on x conditional on t_{il}. Then, it holds for all i by picking some appropriate $\tilde{w}_{il}$ and a such that $\max_i am_i|\tilde{w}_{il}|$ is sufficiently small, since $x_j^2(t)$ is sub-exponential uniformly in t by Assumption 2.

Define $B:=\sum_{i=1}^n b_{ijk}=O\left(n^3\bar{m}^3h^3+n^3\bar{m}^4h^4\right)$. For sufficiently large n, we have

$$
\begin{aligned}
P\left(\sum_{i=1}^n W_{ijk}\geq\gamma_n\middle|E_n\right) &\leq \exp\{-a\gamma_n\}E\left\{\exp\left(a\sum_{i=1}^n W_{ijk}\right)\middle|E_n\right\} \\
&= \exp\{-a\gamma_n\}\prod_{i=1}^n E\left\{\exp(aW_{ijk})\middle|E_n\right\} \\
&\leq \exp\{-a\gamma_n+Ba^2\}. \qquad\qquad \text{(S.6)}
\end{aligned}
$$

Note that (S.6) is minimized when $a=\gamma_n/(2B)$ and that the minimizer is $\exp\{-\gamma_n^2/(4B)\}$. Thus, there exists some positive constant C such that

$$
P\left(\sum_{i=1}^n W_{ijk}\geq\gamma_n\middle|E_n\right)\leq\exp\left\{-C\gamma_n^2/(n^3\bar{m}^3h^3+n^3\bar{m}^4h^4)\right\}.
$$

Similarly, we obtain

$$
P\left(\sum_{i=1}^n W_{ijk}\leq-\gamma_n\middle|E_n\right)\leq\exp\left\{-C\gamma_n^2/(n^3\bar{m}^3h^3+n^3\bar{m}^4h^4)\right\}.
$$

The following obtained by a simple union bound holds for each $t \in \mathcal{T}$,

$$
\begin{aligned}
& P\left(\max_{j,k}\left|\sum_{i=1}^{n}\sum_{l=1}^{m_i}\left\{\tilde{w}_{il}x_{ijl}x_{ikl} - E\left(\tilde{w}_{il}x_{ijl}x_{ikl}\right)\right\}\right| \geq \gamma_n \middle| E_n\right) \\
\leq \quad & 2p^2\exp\left\{-C\gamma_n^2/(n^3\bar{m}^3h^3 + n^3\bar{m}^4h^4)\right\}.
\end{aligned}
$$

Let $\gamma_n = O\{(\log p)^{1/2}(n^3\bar{m}^3h^3 + n^3\bar{m}^4h^4)^{1/2}\}$. Note that $\tilde{w}_{il} = O_p\{(n^2\bar{m}^2h^3)^{1/2}\}$ from Lemma 4, then with probability tending to 1, the event E_n holds from Assumption 4. Consequently,

$$
\max_{j,k}\left|\sum_{i=1}^{n}\sum_{l=1}^{m_i}\left\{\tilde{w}_{il}x_{ijl}x_{ikl} - E\left(\tilde{w}_{il}x_{ijl}x_{ikl}\right)\right\}\right| = O_p\{(\log p)^{1/2}(n^3\bar{m}^3h^3 + n^3\bar{m}^4h^4)^{1/2}\}.
$$

□

References

Absil, P.-A., Mahony, R., and Sepulchre, R. (2009), *Optimization algorithms on matrix manifolds*, Princeton University Press.

Chen, S., Ma, S., Man-Cho So, A., and Zhang, T. (2020), "Proximal gradient method for nonsmooth optimization over the Stiefel manifold," *SIAM Journal on Optimization*, 30, 210–239.

Edelman, A., Arias, T. A., and Smith, S. T. (1998), "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, 20, 303–353.

Vu, V. Q. and Lei, J. (2013), "Minimax sparse principal subspace estimation in high dimensions," *The Annals of Statistics*, 41, 2905–2947.

Wainwright, M. J. (2019), *High-dimensional statistics: A non-asymptotic viewpoint*, Cambridge University Press.

Xiao, X., Li, Y., Wen, Z., and Zhang, L. (2018), "A regularized semi-smooth Newton method with projection steps for composite convex programs," *Journal of Scientific Computing*, 1–26.