Supplement to "Dynamic principal component analysis from a global perspective"

LINGXUAN SHAO[1,a] and FANG YAO[2,b]

[1]Department of Statistics and Data Science, School of Management, Fudan University, Shanghai, China, [a]shao_lingxuan@fudan.edu.cn
[2]Department of Probability and Statistics, Center for Statistical Science, School of Mathematical Sciences, Peking University, Beijing, China, [b]fyao@math.pku.edu.cn

This supplementary material includes computation algorithm, technical proofs, additional simulations and an extension to general Lie group for "Dynamic principal component analysis from a global perspective".

S.1. Computation algorithm

In this section, we present the computational algorithm for conducting dynamic principal component analysis on the centered data $\{X_i(t_j)\}_{1\le i\le n, 1\le j\le m}$. This discussion is particularly beneficial for readers interested in implementing our one-step unrolling method without delving into every detail of the paper. The steps of the algorithm are outlined as follows:
Step 1: compute the sample covariance $\widehat{\Sigma}(t_j) = n^{-1}\sum_{i=1}^n \left(X_i(t_j)\right)\left(X_i(t_j)\right)^T$ and the orthogonal matrix $\widehat{P(t_j)}$ that diagonalizes $\widehat{\Sigma}(t_j)$ on each observed time point t_j.
Step 2: select $\mathcal{P}_1 = \widehat{P(t_1)}$ and recursively $\mathcal{P}_j = \arg\min_{P\in \widehat{P(t_j)}\mathcal{G}} d_M(\mathcal{P}_{j-1}, P)$ for $2\le j\le m$ on M, where $\mathcal{G}$ denotes the signed permutation group.
Step 3: construct regular base curve Γ that relies on $\{\mathcal{P}_j\}_{1\le j\le m}$. This construction commences with the development of $\Gamma_g(t)$, which is characterized as a piece-wise geodesic curve linking $\{\mathcal{P}_{j_i}\}_{1\le i\le D}$ in the interval $[t_{j_1}, t_{j_D}]$ and extended onto $[0, t_{j_1}]$ and $[t_{j_D}, 1]$ geodesically in the boundaries. To be specific, Γ_g is geodesic on $[0, t_{j_2}]$ by fixing $\Gamma_g(t_{j_1}) = \mathcal{P}_{j_1}$ and $\Gamma_g(t_{j_2}) = \mathcal{P}_{j_2}$, and geodesic on $[t_{j_{D-1}}, 1]$ by fixing $\Gamma_g(t_{j_{D-1}}) = \mathcal{P}_{j_{D-1}}$ and $\Gamma_g(t_{j_D}) = \mathcal{P}_{j_D}$. Then, the function $\Gamma_g(t)$ satisfies the differential equation $\Gamma_g'(t) = z_g(t)\Gamma_g(t)$ with

$$z_g(t) = a_1 I_{\{0\le t\le t_{j_1}\}} + \textstyle\sum_{i=1}^{D-1} a_i I_{\{t_{j_i}\le t\le t_{j_{i+1}}\}} + a_{D-1} I_{\{t_{j_D}\le t\le 1\}}$$

where the set $\{a_i\}_{1\le i\le D-1}$ is identified as elements in $T_I M$. Analogously, the regular base curve Γ is defined as the solution to the subsequent differential equation on M

$$\tfrac{d}{dt}\Gamma(t) = z(t)\Gamma(t), \quad \Gamma(t_{j_1}) = \mathcal{P}_{j_1}.$$

This matrix ordinary differential equation, featuring a time-varying coefficient $z(t)$, is solved numerically using a discretization method.
Step 4: utilize the unrolling and unwrapping techniques, along with the right multiplication $(R_{\Gamma(0)})^{-1}$, to transform the data into a common linear space $T_I M$, which corresponds to the anti-symmetric matrix space. The transformed regular base curve $\tilde{\Gamma}$ satisfies the matrix ordinary differential equation

$$\tfrac{d}{dt}\tilde{\Gamma}(t) = (R_{\Gamma(0)})^{-1}\mathrm{Par}_{\Gamma(t)}^{\Gamma(0)}\left(\tfrac{d}{dt}\Gamma(t)\right)$$

and initial condition $\tilde{\Gamma}(0) = 0$, which is also solved numerically. The transformed point $\tilde{Q}_j \in T_I M$ of $\mathcal{P}_j$ is

$$\tilde{Q}_j = \tilde{\Gamma}(t_j) + (R_{\Gamma(0)})^{-1} \mathrm{Par}_{\Gamma(t_j)}^{\Gamma(0)} \left(\mathrm{Log}_{\Gamma(t_j)}(\mathcal{P}_j) \right),$$

where the parallel transport characterized by Lemma 5 also encompasses a matrix ordinary differential equation that is again solved numerically.

Step 5: apply the one-dimensional smoothing spline fitting to the data $\{\tilde{Q}_j\}_{1 \le j \le m}$ entry-wisely. To be specific, for a certain entry indexed by the pair (k, l), we consider the data $\{(t_j, \tilde{Q}_j^{kl})\}_{j=1,\ldots,m}$ where $\tilde{Q}_j^{kl}$ is the entry of $\tilde{Q}_j$ in the kth row and lth column, and obtain a standard smoothing spline estimate $\tilde{\alpha}^{kl}$

$$\tilde{\alpha}^{kl}(t) = \underset{f \in W^{2,2}([0,1])}{\arg\min} \left(\frac{1}{m} \sum_{j=1}^m \left(f(t_j) - \tilde{Q}_j^{kl} \right)^2 + \rho \int_0^1 \left| \frac{d^2}{ds^2} f(s) \right|^2 ds \right).$$

Assemble the entries $\tilde{\alpha}^{kl}$ together to get an curve $\tilde{\alpha}$ lying on $T_I M$, and move $\tilde{\alpha}$ back to obtain the dynamic eigen-frame estimate $\widehat{\alpha}$ on M.

$$\widehat{\alpha} = \mathrm{Exp}_{\Gamma(t)} \circ \mathrm{Par}_{\Gamma(0)}^{\Gamma(t)} \circ R_{\Gamma(0)}(\tilde{\alpha}(t) - \tilde{\Gamma}(t)).$$

This curve $\widehat{\alpha}$ is a representative of α in Algorithm 1.

Step 6: select the leading dynamic components via the L^p norms. Compute $\widehat{\lambda_{\mathcal{E}_l}(t_j)} = (\mathcal{P}_j^T \widehat{\Sigma}(t_j) \mathcal{P}_j)_{ll}$, where $\widehat{\Sigma}(t_j)$ is the sample covariance, and $(\cdot)_{ll}$ corresponds to the element in the lth column and the lth row of a matrix. Their L^p norms are estimated by $R_l = \left(\sum_{j=2}^m |\widehat{\lambda_{\mathcal{E}_l}(t_j)}|^p (t_j - t_{j-1}) \right)^{1/p}$ and are ordered as $R_{l_1} \ge \cdots \ge R_{l_d}$. Finally, the kth global principal component e_k is estimated by $\hat{e}_k = \mathcal{E}_{l_k}$, where $\mathcal{E}_{l_k}$ is the l_kth column of $\widehat{\alpha}$.

In concluding this section, we note that the minimization problem

$$\mathcal{P}_j = \arg\min_{P \in \widehat{P(t_j)} \mathcal{G}} d_M(\mathcal{P}_{j-1}, P)$$

outlined in **Step 2** is addressed using Algorithm S.1, which circumvents the need to search the entire signed permutation group $\mathcal{G}$. For large sample sizes n and high sampling frequency m, the algorithm consistently yields the minimizer, as substantiated by Lemma S.1, with a polynomial time complexity in terms of n and m.

Algorithm S.1. Algorithm of the minimization problem in **Step 2**.

Input: Two matrices $\widehat{P(t_j)}$ and $\mathcal{P}_{j-1}$.
Output: The minimizer $\mathcal{P}_j = \arg\max_{P \in \widehat{P(t_j)} \mathcal{G}} d_M(\mathcal{P}_{j-1}, P)$.

– (1) Compute $A = (\widehat{P(t_j)})^{-1} \mathcal{P}_{j-1}$.
– (2) Find the index pair $(i_1, j_1) = \arg\max_{1 \le i, j \le d} |A_{ij}|$.
– (3) Suppose the indexes $\{(i_k, j_k)\}_{k=1}^l$ are determined. Find a new index pair
$(i_{l+1}, j_{l+1}) = \arg\max_{1 \le i, j \le d; i \ne i_1, \ldots, i_l; j \ne j_1, \ldots, j_l} |A_{ij}|$.
– (4) Continue repeating step (3) until d pairs of indices $\{(i_k, j_k)\}_{k=1}^d$ are identified.
– (5) Define $G \in \mathcal{G}$ with entries $G_{i_k j_k} = \mathrm{sign}(A_{i_k j_k})$ for $1 \le k \le d$ and other entries being zero.
– (6) Output $\mathcal{P}_j = \widehat{P(t_j)} G$.

Lemma S.1. *Algorithm S.1 efficiently solves the minimization problem* $\arg\min_{P \in \widehat{P(t_j)}\mathcal{G}} d_M(\mathcal{P}_{j-1}, P)$, *demonstrating polynomial time complexity for large sample sizes n and sampling frequency m.*

Proof of Lemma S.1. The minimizer of $\arg\min_{P \in \widehat{P(t_j)}\mathcal{G}} d_M(\mathcal{P}_{j-1}, P)$ obviously lies in the same connected component with $\mathcal{P}_{j-1}$ in the orthogonal matrix group M. Therefore, it is equivalent to consider $\arg\min_{P\in\mathcal{G}} d_M((\widehat{P(t_j)})^{-1}\mathcal{P}_{j-1}, P)$ according to Lemma 7.

The proof of Theorem 1 suggests that there exist a series of $\{G_{n,j}\}_n \subset \mathcal{G}$ such that $\|\widehat{P(t_j)}G_{n,j} - P(t_j)\|_{HS} \to 0$, and also a series of $\{G_{n,j-1}\}_n \subset \mathcal{G}$ such that $\|\mathcal{P}_{j-1}G_{n,j-1} - P(t_{j-1})\|_{HS} \to 0$ since $\mathcal{P}_{j-1} \in [\widehat{P(t_{j-1})}]$. Combine above two convergences to see

$$\|(\mathcal{P}_{j-1})^{-1}\widehat{P(t_j)} - G_{n,j-1}(P(t_{j-1}))^{-1}P(t_j)(G_{n,j})^{-1}\|_{HS} \to 0$$

For any $\varepsilon > 0$, we have $\|(P(t_{j-1}))^{-1}P(t_j) - I_d\| < \varepsilon$ for large m due to the smoothness of the dynamic eigen-frame P in Assumption 1. Therefore, we see

$$\|(\mathcal{P}_{j-1})^{-1}\widehat{P(t_j)} - G_{n,j-1}(G_{n,j})^{-1}\|_{HS} < 2\varepsilon$$

for large n and m. It suggests that $(\mathcal{P}_{j-1})^{-1}\widehat{P(t_j)}$ is arbitrarily close to this signed permutation matrix $G_{n,j-1}(G_{n,j})^{-1} \in \mathcal{G}$ for any $\varepsilon > 0$. Therefore, the minimizer is

$$G_{n,j-1}(G_{n,j})^{-1} = \arg\min_{P\in\mathcal{G}} d_M((\widehat{P(t_j)})^{-1}\mathcal{P}_{j-1}, P)$$

since $\mathcal{G}$ is a distinct group on M (Lemma 9).

It is observed that $G_{n,j-1}(G_{n,j})^{-1}$ comprises n entries of ± 1 and $n(n-1)$ zero entries. Identifying the positions of these n nonzero entries together with their signs is adequate for reconstructing $G_{n,j-1}(G_{n,j})^{-1}$. The convergence of $(\widehat{P(t_j)})^{-1}\mathcal{P}_{j-1}$ to $G_{n,j-1}(G_{n,j})^{-1}$ in the Hilbert–Schmidt norm implies point-wise convergence. Consequently, the indices of the n entries with the largest absolute values in $(\widehat{P(t_j)})^{-1}\mathcal{P}_{j-1}$ correspond to the indices of ± 1 in $G_{n,j-1}(G_{n,j})^{-1}$. Using this principle, Algorithm S.1 determines the minimizer, addressing the stated minimization problem efficiently for large values of n and m within polynomial time complexity.

□

S.2. Technical proofs

This section presents the technical proofs of the lemmas, propositions and theorems.

Proof of Lemma 3. Suppose that $D = P^T V P$ where P is an orthogonal matrix and D is a diagonal matrix with distinct diagonal elements. It is obvious that for any $G \in \mathcal{G}$, $G^T D G = (PG)^T V (PG)$ is also a diagonal matrix. Besides, if there exits a $G \in M$ such that $G^T D G$ is a diagonal matrix, then we have

$$G^T D G v_k = \lambda_k v_k,$$

where $v_k = (0, 0, \ldots, 1 \ldots, 0)^T$ and λ_k is the kth diagonal element of $G^T D G$ for all $1 \le l \le d$. This fact leads to $D(Gv_k) = \lambda_k (Gv_k)$ and (Gv_k) is an eigenvector of D. Since all the eigen-spaces $\{\mathcal{V}_k\}_{1\le k\le d}$ of D are one-dimensional ($\mathcal{V}_k = \mathbb{R}v_k$), then $Gv_k = \pm v_l$ for some $1 \le l \le d$ and hence $G \in \mathcal{G}$. Therefore, all the orthogonal matrices that diagonalize V form exactly one equivalence class $[P] \in N$. □

Proof of Lemma 4. Suppose that $\gamma(t)$ is a curve on M satisfying $\gamma(0) = P$ and $\gamma'(0) = u$. By the definition of tangent map $d(L_G)_P$, we have

$$d(L_G)_P(u) = \frac{d}{dt} L_G(\gamma(t))\Big|_{t=0} = G\frac{d}{dt}\gamma(0) = Gu.$$

Thus, for any $u, v \in T_P M$, $\langle Gu, Gv\rangle = \mathrm{tr}(v^T G^T G u) = \mathrm{tr}(v^T u) = \langle u, v\rangle$. The expression for $d(R_G)_P$ is derived in a similar way. □

Proof of Lemma 5. Since $R(t) \in T_I M$, $v_l(\frac{1}{2}R), v_r(-\frac{1}{2}R)$ are curves on M and $v_r(-\frac{1}{2}R)v_l(\frac{1}{2}R) = I$. Proving $u(t)$ is geodesic is equivalent to showing that $u'(t) \perp T_{\gamma(t)} M$. From section A we have $T_{\gamma(t)} M = \{S\gamma(t) \mid S \in T_I M\}$. Denote $A_1 = v_l(\frac{1}{2}R)u(0)\gamma(0)^{-1}v_r(-\frac{1}{2}R)$ and $A_2 = \gamma'(t)\gamma(t)^{-1}$ where $A_1, A_2 \in T_I M$. Then

$$\begin{aligned}
\langle \tfrac{d}{dt}u(t), S\gamma(t)\rangle =& \langle \tfrac{1}{2}\gamma'(t)\gamma(t)^{-1}A_1\gamma(t) - \tfrac{1}{2}A_1\gamma'(t) + A_1(t)\gamma'(t), S\gamma(t)\rangle \\
=& \tfrac{1}{2}\langle \gamma'(t)\gamma(t)^{-1}A_1\gamma(t) + A_1\gamma'(t), S\gamma(t)\rangle \\
=& \tfrac{1}{2}\langle A_2A_1 + A_1A_2, S\rangle.
\end{aligned}$$

Since $S, A_1, A_2, S \in T_I M$,

$$\mathrm{tr}(SA_1A_2) = \mathrm{tr}(A_2^T A_1^T S^T) = \mathrm{tr}(-A_2A_1S) = \mathrm{tr}(-SA_2A_1).$$

This gives $\langle u'(t), \gamma(t)S\rangle = 0$ for all $S \in T_I M$ and hence $u'(t) \perp T_{\gamma(t)} M$. □

Proof of Lemma 6. According to Lemma 5, $\gamma'(t)$ is a parallel vector field along $\gamma(t)$ and hence $\gamma(t)$ is a geodesic curve, which leads to the explicit expression of the Riemannian exponential map Exp_P. The length of the geodesic curve γ from $t = a$ to $t = b$ is

$$Len(a, b, \gamma(t)) = \int_a^b \|\gamma'(t)\|_{\mathrm{HS}} dt = (b - a)\|S\|_{\mathrm{HS}}.$$

□

Proof of Lemma 7. According to Lemma 6, the geodesic line linking I and $P_2^T P_1$ is the curve $\gamma(t) = \exp(t \log(P_2^T P_1))$ where $\log(P_2^T P_1)$ is well defined since $\|P_2^T P_1 - I\|_{\mathrm{HS}} < 1$. In addition, the Riemannian logarithm map is bijective on $B(I, 1) \cap M$ and this gives $d_M(P_2^T P_1, I) = \|\log(P_2^T P_1)\|_{\mathrm{HS}}$. The left multiplication L_{P_2} is isometric on M and subsequently $d_M(P_1, P_2) = d_M(I, P_2^T P_1)$. □

Proof of Lemma 8. Recall the definition of the distance between $[P_1]$ and $[P_2]$ on N

$$d_N([P_1], [P_2]) = \min_{\{r(t): r(0)=[P_1], r(1)=[P(2)]\}} Len(0, 1, r).$$

For any $\varepsilon > 0$, there exists $r_\varepsilon(t)$ such that $Len(0, 1, r_\varepsilon) < d_N([P_1], [P_2]) + \varepsilon$, $r_\varepsilon(0) = [P_1]$ and $r_\varepsilon(1) = [P(2)]$. According to lifting Lemma 2, there exists a curve $\gamma_\varepsilon(t)$ on M such that $\phi(\gamma_\varepsilon(t)) = r_\varepsilon(t)$. This indicates that $\gamma_\varepsilon(0) = P_1G_1$ and $\gamma_\varepsilon(1) = P_2G_2$ for some $G_1, G_2 \in \mathcal{G}$. Since ϕ preserves metric, we have $Len(0, 1, r_\varepsilon) = Len(0, 1, \gamma_\varepsilon)$ and

$$\begin{aligned}
&\min_{G\in\mathcal{G}} d_M(P_1, P_2G) = \min_{G_1,G_2\in\mathcal{G}} d_M(P_1G_1, P_2G_2) \\
&\le Len(0, 1, \gamma_\varepsilon) = Len(0, 1, r_\varepsilon) < d_N([P_1], [P_2]) + \varepsilon.
\end{aligned}$$

Similarly, one can show that $d_N([P_1],[P_2]) < \min_{G\in\mathcal{G}} d_M(P_1, P_2G) + \varepsilon$ and subsequently

$$d_N([P_1],[P_2]) = \min_{G\in\mathcal{G}} d_M(P_1, P_2G).$$

$\square$

Proof of Lemma 9. For any $P \in M$ and $G_1, G_2 \in \mathcal{G}$, $\|PG_1 - PG_2\|_{\mathrm{HS}} = \|I - G_2^{-1}G_1\|_{\mathrm{HS}}$. The set $\{G_2^{-1}G_1\}_{G_1\neq G_2}$ are finite and their norms have a positive lower bound. In addition, since the left multiplication $L_{G_2^{-1}P^{-1}}$ preserves the distance, $d_M(PG_1, PG_2) = d_M(G_2^{-1}G_1, I)$ also has a positive lower bound. Denote the lower bound by R_1.

The covering map $\phi : M \to N$ restricted on the ball $B_M(PG; \frac{1}{2}R_1)$ is in fact injective. Suppose that $Q_1, Q_2 \in B_M(PG; \frac{1}{2}R_1)$ such that $Q_1 \neq Q_2$ and $\phi(Q_1) = \phi(Q_2)$. Then there exists $G \in \mathcal{G}$ such that $Q_1 = Q_2G$. By previous discussion, $d_M(Q_1, Q_2) > R_1$. However,

$$d_M(Q_1, Q_2) \leq d_M(Q_1, PG) + d_M(PG, Q_2) < R_1,$$

which leads to contradiction. Thus, $\phi\big|_{B_M(PG;\frac{1}{2}R_1)}$ is injective. Therefore, the local isometric map ϕ_{PG} can be defined on the ball $B_M(PG; \frac{1}{2}R_1)$. $\square$

Proof of Lemma 10. Since $\phi(\gamma_1(t)) = \phi(\gamma_2(t)) = r(t)$, there exists $G \in \mathcal{G}$ such that $\gamma_1(t)G = \gamma_2(t)$ and subsequently

$$R^{-1}_{\gamma_2(0)}(\gamma_2(t)) = \gamma_2(t)\gamma_2^{-1}(0) = \gamma_1(t)GG^T\gamma_1^{-1}(0) = R^{-1}_{\gamma_1(0)}(\gamma_1(t)).$$

In addition, the right multiplication R_G satisfies $\phi \circ R_G = \phi$. The tangent map on $\gamma_1(0)$ is

$$d\phi\big|_{\gamma_2(0)} \circ R_G\big|_{\gamma_1(0)} = d\phi\big|_{\gamma_1(0)},$$

which further implies that

$$R_G^{-1}\big|_{\gamma_1(0)} \circ d\phi^{-1}\big|_{\gamma_2(0)} = d\phi^{-1}\big|_{\gamma_1(0)}.$$

Since $G = \gamma_1^{-1}(0)\gamma_2(0)$, we then have

$$R^{-1}_{\gamma_1(0)} \circ d\phi^{-1}_{\gamma_1(0)} = R^{-1}_{\gamma_2(0)} \circ d\phi^{-1}_{\gamma_2(0)}.$$

$\square$

Proof of Lemma 11. By Lagrange mean value inequality,

$$\sup_{1\leq i\leq D} \|P(t_{j_{i+1}}) - P(t_{j_i})\|_{\mathrm{HS}} < (\sup_{0\leq t\leq 1} \|P'(t)\|_{\mathrm{HS}})(t_{j_{i+1}} - t_{j_i}) < \frac{C_4 \sup_{0\leq t\leq 1} \|P'(t)\|_{\mathrm{HS}}}{D} \to 0.$$

Hence for sufficiently large D, Lemma 1 suggests

$$\sup_{1\leq i\leq D} d_M(P(t_{j_{i+1}}), P(t_{j_i})) < \tfrac{1}{3}R, \quad \sup_{1\leq i\leq D} d_N([P(t_{j_{i+1}})], [P(t_{j_i})]) < \tfrac{1}{3}R.$$

According to Theorem 1, $\sup_{1\leq j\leq m} E\, d_N([Q_j], [P(t_j)])^2 = O(n^{-1})$ and subsequently

$$\begin{aligned}
&\mathrm{pr}\{\sup_{1\leq i\leq D} d_N([Q_{j_i}], [Q_{j_{i+1}}]) < R\} \geq \mathrm{pr}\{\sup_{1\leq i\leq D} d_N([Q_{j_i}], [P(t_{j_i})]) < \tfrac{1}{3}R\} \\
&\geq 1 - \textstyle\sum_{i=1}^{D} \mathrm{pr}\{d_N([Q_{j_i}], [P(t_{j_i})]) \geq \tfrac{1}{3}R\} \geq 1 - \sum_{i=1}^{D} \frac{9}{R^2} \sup_{1\leq i\leq D} E\, d_N([Q_{j_i}], [P(t_{j_i})])^2 \\
&= 1 - O(\tfrac{D}{n}).
\end{aligned}$$

Since $D = O(n^{1/4}) = o(n)$,

$$\lim_{n\to+\infty} \mathrm{pr}\{\sup_{1\le i\le D} d_N([Q_{j_i}], [Q_{j_{i+1}}]) < R\} = 1.$$

Similar arguments lead to $\lim_{n\to+\infty} \mathrm{pr}\{\sup_{1\le i\le D} d_M(\mathcal{P}_{j_i}, \mathcal{P}_{j_{i+1}}) < R\} = 1.$ □

Proof of Lemma 12. Suppose that γ is lift of r satisfying $\phi(\gamma) = r$ and $r = [\gamma] = \gamma\mathcal{G}$. Take derivative on both sides of $\gamma(t)\gamma(t)^T = I$, we have

$$\big(\tfrac{d}{dt}\gamma(t)\big)\gamma(t)^T + \gamma(t)\big(\tfrac{d}{dt}\gamma(t)\big)^T = 0.$$

Define $z(t) = \gamma'(t)\gamma(t)^T \in T_I M$, we have $\gamma'(t) = z(t)\gamma(t)$ and $\gamma'(t)\mathcal{G} = z(t)\gamma(t)\mathcal{G}$. View r as the notation of equivalent class of $\gamma\mathcal{G}$, then

$$r'(t) = z(t)r(t).$$

If $r(t)$ is a geodesic line with $r(0) = [Q_0]$ and $r(1) = [Q_1]$, choose the lift curve $\gamma(t)$ on M with $\gamma(0) = Q_0$ and $\gamma(1) = Q_1$, then γ has the following expression according to Lemma 6,

$$\gamma(t) = \exp(t\log(Q_1 Q_0^{-1}))Q_0.$$

Therefore, $z(t) = \log(Q_1 Q_0^{-1})$ and does not rely on t. □

Proof of Lemma 13. Proposition 3 suggests that $\lim_{n\to\infty} \mathrm{pr}\{\sup_{t\in[0,1]} d_N([P(t)], r(t)) < R\} = 1$. Consider the case when $\sup_{t\in[0,1]} d_N([P(t)], r(t)) < R$ holds. This implies that there exists a lifting curve $\gamma(t)$ of $r(t)$ such that $\sup_{t\in[0,1]} d_M(P(t), \gamma(t)) < R$. Then $\tilde{r}(t)$ and $\tilde{P}(t)$ satisfy

$$\tfrac{d}{dt}\tilde{r}(t) = R^{-1}_{\gamma(0)}\mathrm{Par}^{\gamma(0)}_{\gamma(t)}\big(\tfrac{d}{dt}\gamma(t)\big), \quad \tilde{P}(t) = \tilde{r}(t) + R^{-1}_{\gamma(0)}\mathrm{Par}^{\gamma(0)}_{\gamma(t)}\{\mathrm{Log}_{\gamma(t)}(P(t))\}.$$

Lemma 6 suggests that $\mathrm{Log}_A(B) = \log(BA^{-1})A$ when $d_M(A, B) < R$, while Lemma 5 characterize the parallel transport along the curve γ. Thus,

$$\begin{aligned}\tfrac{d}{dt}\tilde{r}(t) &= v_l(-\tfrac{1}{2}\gamma'(t)\gamma(t)^{-1})\gamma'(t)\gamma(t)^{-1}v_r(\tfrac{1}{2}\gamma'(t)\gamma(t)^{-1})\\ \tilde{P}(t) &= \tilde{r}(t) + v_l(-\tfrac{1}{2}\gamma'(t)\gamma(t)^{-1})\log(P(t)\gamma^{-1}(t))v_r(\tfrac{1}{2}\gamma'(t)\gamma(t)^{-1}).\end{aligned}$$

Since $\|\gamma'(t)\|_{\mathrm{HS}} = \|r'(t)\|_{\mathrm{HS}}$ and $\|\gamma''(t)\|_{\mathrm{HS}} = \|r''(t)\|_{\mathrm{HS}}$, Proposition 2 suggests that

$$\sup_{0\le t\le 1}(\|\tfrac{d}{dt}\tilde{r}(t)\|_{\mathrm{HS}} + \|\tfrac{d^2}{dt^2}\tilde{r}(t)\|_{\mathrm{HS}}) \le \sup_{0\le t\le 1}(\|r'(t)\|_{\mathrm{HS}} + \|r'(t)\|^2_{\mathrm{HS}} + \|r''(t)\|_{\mathrm{HS}}) = O_p(1).$$

Since $\|\log(P(t)\gamma^{-1}(t))\|_{\mathrm{HS}} = d_M(P(t), \gamma(t)) < R$ is bounded, we claim that

$$\|\tfrac{d}{dt}\log(P(t)\gamma^{-1}(t))\|_{\mathrm{HS}} \le C\|\gamma'(t)\|_{\mathrm{HS}}$$

and

$$\|\tfrac{d^2}{dt^2}\log(P(t)\gamma^{-1}(t))\|_{\mathrm{HS}} \le C(\|\gamma'(t)\|^2_{\mathrm{HS}} + \|\gamma''(t)\|_{\mathrm{HS}})$$

for a constant C. Straightforward derivation yields

$$\sup_{0\le t\le 1}(\|\tfrac{d}{dt}\tilde{P}(t)\|_{\mathrm{HS}} + \|\tfrac{d^2}{dt^2}\tilde{P}(t)\|_{\mathrm{HS}}) \le \sup_{0\le t\le 1}(\|r'(t)\|_{\mathrm{HS}} + \|r'(t)\|^2_{\mathrm{HS}} + \|r''(t)\|_{\mathrm{HS}}) = O_p(1).$$

□

Proof of Lemma 14. Without loss of generosity, suppose that $t < s$ and there exists a finite partition such that $[t, s] = \cup_{i=0}^{N} [\tau_i, \tau_{i+1}]$ where $\tau_0 = t$ and $\tau_{N+1} = s$. Since r is continuous on $[0, 1]$, one can choose N large enough such that $r\big|_{[\tau_i, \tau_{i+1}]} \subset \phi\big(B_M(\Gamma(\tau_i), R_1)\big) = B_N(r(\tau_i); R_1)$ from Lemma 9. Since $\phi_{\Gamma(\tau_i)}$ is isometric from $B_M(\Gamma(\tau_i); R_1)$ to $B_N(r(\tau_i); R_1)$, it induces an isometric form the tangent bundle of $B_M(\Gamma(\tau_i); R_1)$ to the tangent bundle of $B_N(r(\tau_i); R_1)$. Therefore, we have

$$\left(d\phi_{\Gamma(\tau_{i+1})}\right)^{-1} \circ \mathrm{Par}_{r(\tau_i)}^{r(\tau_{i+1})}(v) = \mathrm{Par}_{\Gamma(\tau_i)}^{\Gamma(\tau_{i+1})} \circ \left(d\phi_{\Gamma(\tau_i)}\right)^{-1}(v).$$

Composing above from $i = 0$ to $i = N$ together leads to the result. □

Proof of Proposition 1. The fact that Γ is a lifting of r results from Proposition 4. It suffices to show that the data $\{\tilde{Q}_j\}_{1\le j\le m}$ on $T_I M$ obtained from the simplified way and the complex way in Figure 4 are the same. In Section 3.2, γ is constructed as a lifting of r into M and Lemma 10 suggests that the choice of lifting does not affect the result. Without loss of generosity, γ can be chosen as Γ according to the first step of this proof. Besides, Lemma 14 leads to

$$d\phi_{\Gamma(0)}\left(\frac{d}{dt}\acute{\Gamma}(t)\right) = d\phi_{\Gamma(0)}\left(\mathrm{Par}_{\Gamma(t)}^{\Gamma(0)}\left(\frac{d}{dt}\Gamma(t)\right)\right) = \mathrm{Par}_{r(t)}^{r(0)}\left(\frac{d}{dt}r(t)\right) = \frac{d}{dt}\acute{r}(t).$$

Since both $\acute{\Gamma}$ and $\acute{r}$ start at 0, the definition of unrolling in Section 3.1 indicates that $d\phi_{\Gamma(0)}\left(\acute{\Gamma}(t)\right) = \acute{r}(t)$. The data $\{\tilde{Q}_j\}_{1\le j\le m}$ on $T_I M$ obtained from the complex way in Figure 4 is computed as

$$\tilde{Q}_j = R_{\Gamma(0)}^{-1}\left(d\phi_{\Gamma(0)}\right)^{-1}\left\{\acute{r}(t_j) + \mathrm{Par}_{r(t_j)}^{r(0)}\left(\mathrm{Log}_{r(t_j)}\left(\phi(\widehat{P(t_j)})\right)\right)\right\}. \tag{S.1}$$

Since r is the regular base curve on N relying on $\{[Q_j]\}_{1\le j\le m}$ and Γ is the regular base curve on M relying on the new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$, (S.1) can be simplified by

$$\tilde{Q}_j = R_{\Gamma(0)}^{-1}\left\{\acute{\Gamma}(t_j) + \mathrm{Par}_{\Gamma(t_j)}^{\Gamma(0)}\left(\mathrm{Log}_{\Gamma(t_j)}(\mathcal{P}_j)\right)\right\} = R_{\Gamma(0)}^{-1}\acute{\mathcal{P}}_j,$$

which verifies our claim. □

Proof of Proposition 2. We claim that $\sup_{1\le i\le D-1}\|a_i\|_{\mathrm{HS}} = O_p(1)$ and $\sup_{1\le i\le D-2}\|a_{i+1} - a_i\|_{\mathrm{HS}} = O_p(D^{-1})$. By Taylor expansion, $P(t_{j_{i+1}}) = P(t_{j_i}) + P'(t_{j_i})h_i + O(h_i^2)$. Thus, according to Lemma 11, $\log\{P(t_{j_{i+1}})P(t_{j_i})^{-1}\}$ is well defined with probability tending to one and

$$\sup_{1\le i\le D-1}\|\tfrac{1}{h_i}\log\{P(t_{j_{i+1}})P(t_{j_i})^{-1}\}\|_{\mathrm{HS}} = \sup_{1\le i\le D-1}\|P'(t_{j_i})P(t_{j_i})^{-1} + O(h_i)\|_{\mathrm{HS}} = O(1).$$

In addition, since $P'(t_{j_{i+1}}) - P'(t_{j_i}) = P''(t_{j_i})h_i + O(h_i^2)$ and

$$P(t_{j_{i+1}})^{-1} - P(t_{j_i})^{-1} = P(t_{j_i})^{-1}P'(t_{j_i})P(t_{j_i})^{-1}h_i + O(h_i^2),$$

we have

$$\begin{aligned}
&\sup_{1\le i\le D-2}\|\tfrac{1}{h_{i+1}}\log\{P(t_{j_{i+2}})P(t_{j_{i+1}})^{-1}\} - \tfrac{1}{h_i}\log\{P(t_{j_{i+1}})P(t_{j_i})^{-1}\}\|_{\mathrm{HS}}\\
=&\sup_{1\le i\le D-2}\|O(h_i) + O(h_{i+1}) + P'(t_{j_i})P(t_{j_i})^{-1}P'(t_{j_i})P(t_{j_i})^{-1}h_i + P''(t_{j_i})P(t_{j_i})^{-1}h_i + O(h_i^2)\|_{\mathrm{HS}}\\
=&O(D^{-1}).
\end{aligned}$$

The fact $\sup_{1\le i\le D} E\, d_N([Q_{j_i}],[P(t_{j_i})]) = O(n^{-1/2})$ from Theorem 1 indicates that there exists a lifting of $[Q_{j_i}]$, denoted by Q_{j_i}, such that $\sup_{1\le i\le D} E\, d_M(Q_{j_i}, P(t_{j_i})) = O(n^{-\frac{1}{2}})$ and $\sup_{1\le i\le D} E\,\|Q_{j_i} - P(t_{j_i})\|_{\mathrm{HS}} = O(n^{-\frac{1}{2}})$. Then according to Lemma 12, a_i is

$$a_i = \tfrac{1}{h_i}\log(Q_{j_{i+1}}Q_{j_i}^{-1}) = \tfrac{1}{h_i}\log\{P(t_{j_{i+1}})P(t_{j_i})^{-1}\} + O_p(Dn^{-1/2}).$$

Due to $D = O(n^{1/4})$, we further deduce that

$$\begin{aligned}
&\sup_{1\le i\le D-1}\|a_i\|_{\mathrm{HS}} = O(1) + O_p(Dn^{-1/2}) = O_p(1),\\
&\sup_{1\le i\le D-2}\|a_{i+1}-a_i\|_{\mathrm{HS}} = O(D^{-1}) + O_p(Dn^{-1/2}) = O_p(D^{-1}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sup_{0\le t\le 1}\|r'(t)\|_{\mathrm{HS}} = \sup_{0\le t\le 1}\|z(t)\|_{\mathrm{HS}} = \sup_{1\le i\le D-1}\|a_i\|_{\mathrm{HS}} = O_p(1),\\
&\sup_{0\le t\le 1}\|r''(t)\|_{\mathrm{HS}} \le \sup_{0\le t\le 1}(\|z'(t)\|_{\mathrm{HS}} + \|z(t)\|_{\mathrm{HS}}^2)\\
&= D\sup_{1\le i\le D-2}\|a_{i+1}-a_i\|_{\mathrm{HS}} + \sup_{1\le i\le D-1}\|a_i\|_{\mathrm{HS}}^2 = O_p(1).
\end{aligned}$$

□

Proof of Proposition 3. We divide the proof into two steps. The first step shows $[P(t)]$ and $[Q_j]$ are close to r_g and the second step shows r_g and r are also close.

Step 1: For any t, t lies in an interval $[t_{j_i}, t_{j_{i+1}}]$. Then

$$d_N([P(t)],[P(t_{j_i})]) \le d_M(P(t), P(t_{j_i})) \le \int_{t_{j_i}}^{t}\|P'(s)\|_{\mathrm{HS}}ds = O(D^{-1}).$$

By the definition of r_g and the fact from the proof of Proposition 2 that $\sup_{1\le i\le D-1}\|a_i\|_{\mathrm{HS}} = O_p(1)$, we conclude that

$$d_N(r_g(t), r_g(t_{j_i})) = (t - t_{j_i})\|a_i\|_{\mathrm{HS}} = O_p(D^{-1}).$$

According to Theorem 1, $\sup_{1\le j\le m} d_N([P(t_j)], Q_j) = O_p(n^{-1/2})$, $Q_{j_i} = r_g(t_{j_i})$ and by triangular inequality,

$$\begin{aligned}
&\sup_{0\le t\le 1} d_N([P(t)], r_g(t))\\
\le &\sup_{0\le t\le 1} d_N([P(t)],[P(t_{j_i})]) + \sup_{0\le t\le 1} d_N(r_g(t), r_g(t_{j_i})) + \sup_{1\le i\le D} d_N([P(t_{j_i})], r_g(t_{j_i}))\\
= &O_p(D^{-1}) + O_p(n^{-1/2}) = O_p(D^{-1}).
\end{aligned}$$

Similarly, we have

$$\sup_{1\le j\le m} d_N([Q_j], r_g(t_j)) \le \sup_{1\le j\le m} d_N(Q_j, [P(t_j)]) + \sup_{1\le j\le m} d_N([P(t_j)], r_g(t_j)) = O_p(D^{-1}).$$

Step 2: Recall that $r(t)$ and $r_g(t)$ satisfy the differential equations $r'(t) = z(t)r(t)$ and $r_g'(t) = z_g(t)r_g(t)$ with initial values $r(t_{j_1}) = r_g(t_{j_1}) = [Q_{j_1}]$. Now choose lifting curves of $r(t)$ and $r_g(t)$, denoted by $\gamma(t)$ and $\gamma_g(t)$, satisfying

$$\tfrac{d}{dt}\gamma(t) = z(t)\gamma(t), \quad \tfrac{d}{dt}\gamma_g(t) = z_g(t)\gamma_g(t), \quad \gamma(t_{j_1}) = \gamma_g(t_{j_1}) = Q_{j_1}.$$

Then

$$\begin{aligned}\tfrac{d}{dt}\|\gamma(t)-\gamma_g(t)\|_{\mathrm{HS}} \le& \|\tfrac{d}{dt}\gamma(t)-\tfrac{d}{dt}\gamma_g(t)\|_{\mathrm{HS}}\\ \le& \|z(t)(\gamma(t)-\gamma_g(t))+(z(t)-z_g(t))\gamma_g(t)\|_{\mathrm{HS}}\\ \le& (\sup_{1\le i\le D}\|a_i\|_{\mathrm{HS}})\|\gamma(t)-\gamma_g(t)\|_{\mathrm{HS}}+\sqrt{d}\|(z(t)-z_g(t)\|_{\mathrm{HS}}.\end{aligned}$$

According to Grownwall inequality and the fact that $\sup_{1\le i\le D-1}\|a_i\|_{\mathrm{HS}}=O_p(1)$ and $\sup_{1\le i\le D-2}\|a_{i+1}-a_i\|_{\mathrm{HS}}=O_p(D^{-1})$ in the proof of Proposition 2,

$$\begin{aligned}\sup_{0\le t\le 1}\|\gamma(t)-\gamma_g(t)\|_{\mathrm{HS}} \le& \sup_{0\le t\le 1}\int_{t_{j_1}}^{t} e^{(\sup_{1\le i\le D-1}\|a_i\|_{\mathrm{HS}})(t-s)}\sqrt{d}\|(z(s)-z_g(s)\|_{\mathrm{HS}}ds\\ \le& O_p(1)\int_0^1\|z(s)-z_g(s)\|_{\mathrm{HS}}ds\\ \le& O_p(1)\sup_{1\le i\le D-2}\|a_{i+1}-a_i\|_{\mathrm{HS}}=O_p(D^{-1}).\end{aligned}$$

Hence, $\sup_{0\le t\le 1} d_N(r(t),r_g(t)) \le \sup_{0\le t\le 1} d_M(\gamma(t),\gamma_g(t)) = O_p(D^{-1})$ according to Lemma 8. Since $\lim_{n\to\infty} D(n)=\infty$ from Assumption 2, the conclusion of the proposition follows from Steps 1 and 2. □

Proof of Proposition 4.
Recall that r is the regular base curve on N obtained from $\{[Q_{j_i}]\}_{1\le i\le D}$ and Γ is the regular base curve on M relying on the representatives $\{\mathcal{P}_{j_i}\}_{1\le i\le D}$.
Step 1: the first step aims to show that Γ_g is a lifting of r_g. Lemma 9 reveals that $\phi_{\mathcal{P}_{j_i}}$ is isometric on the ball $B_M(\mathcal{P}_{j_i};R_1)$ where R_1 is a constant depending only on the dimension d. Assumption 2 ensures that $t_{j_{i+1}}-t_{j_i}=o(1)$ and the continuity of the eigen-frame promises $d_N([P(t_{j_{i+1}})],[P(t_{j_i})])=o(1)$. Theorem 1 suggests that $\widehat{[P(t_j)]}$ converges to $[P(t_j)]$ uniformly over $1\le j\le m$ and subsequently the inequality $d_N(\widehat{[P(t_{j_i})]},\widehat{[P(t_{j_{i+1}})]})<R_1$ holds with probability tending to one. According to the choices of the representatives $\{\mathcal{P}_{j_i}\}_{1\le i\le D}$ described in Section 3.3, they satisfy $d_M\left(\mathcal{P}_{j_i},\mathcal{P}_{j_{i+1}}\right)<R$ for all $1\le i\le D$ with probability tending to one due to Lemma 11. Since $\phi_{\mathcal{P}_{j_i}}$ is isometric on the ball $B_M(\mathcal{P}_{j_i};R_1)$ and Γ_g is geodesic on $[t_{j_i},t_{j_{i+1}}]$ liking $\mathcal{P}_{j_i}$ and $\mathcal{P}_{j_{i+1}}$, its image $\phi(\Gamma_g\big|_{[t_{j_i},t_{j_{i+1}}]})$ is also geodesic liking $\phi(\mathcal{P}_{j_i})=[Q_{j_i}]$ and $\phi(\mathcal{P}_{j_{i+1}})=[Q_{j_{i+1}}]$, which is exactly $r_g\big|_{[t_{j_i},t_{j_{i+1}}]}$. Similar arguments can be made on the boundaries $[0,t_{j_1}]$ and $[t_{j_D},1]$. Above discussion leads to $\phi(\Gamma_g)=r_g$ and subsequently Γ_g is a lifting of r_g.
Step 2: now we show $z_g(t)=\tilde{z}_g(t)\in T_IM$. Since $r_g'(t)=z_g(t)r_g(t)$ in fact refers to the equivalence class $r_g'(t)\mathcal{G}=z_g(t)r_g(t)\mathcal{G}$, we can choose any representative of $r_g(t)\mathcal{G}$, saying Γ_g, to reformulate it as $\Gamma_g'(t)=z_g(t)\Gamma_g(t)$, and subsequently $z_g(t)=\tilde{z}_g(t)$. Lemma 4 claims that the right-multiplication $R_{(\Gamma_g(t))^{-1}}$ maps elements in $T_{\Gamma_g(t)}M$ into T_IM, and subsequently $z_g(t)=\Gamma_g'(t)(\Gamma_g(t))^{-1}\in T_IM$.
Step 3: it remains to show that Γ is a lifting of r and shares similar properties. Equations (5) and (6) suggest that Γ and r are driven by the same coefficient z and the same initial conditions $[\Gamma(t_{j_1})]=[\mathcal{P}_{j_1}]=r(t_{j_1})=[Q_{j_1}]$. Lift r to M to obtain γ such that $\gamma(t_{j_1})=\mathcal{P}_{j_1}$ and $\phi(\gamma)=r$. Then Γ and γ share the same ordinary differential equation and the same initial condition, and subsequently they are the same $\Gamma=\gamma$ due to the uniqueness of solution in this ordinary differential equation, meaning Γ is also a lifting of r. Therefore, the locally isometric covering map ϕ yields $\|\Gamma'\|=\|r'\|$ and $\|\Gamma''\|=\|r''\|$, and subsequently

$$\sup_{t\in[0,1]}(\|\Gamma'(t)\|_{\mathrm{HS}}+\|\Gamma''(t)\|_{\mathrm{HS}})=O_p(1),\quad \sup_{1\le j\le m} d_N(\mathcal{P}_j,\Gamma(t_j))=O_p(D^{-1})=o_p(1).$$

from Propositions 2 and 3. □

Proof of Theorem 1. Define $V_{n,j} = P(t_j)^T \widehat{\Sigma}(t_j) P(t_j)$. Suppose that $S_{n,j}$ is an orthogonal matrix such that $S_{n,j}^T V_{n,j} S_{n,j}$ is a diagonal matrix and hence $P(t_j) S_{n,j}$ diagonalizes $\widehat{\Sigma}(t_j)$. Since the eigenvalues of $\widehat{\Sigma}(t_j)$ are distinct with probability tending to one (Lemma 3), there exists $G_{n,j} \in \mathcal{G}$ such that

$$P(t_j) S_{n,j} = \widehat{P(t_j)} G_{n,j},$$

where $\widehat{P(t_j)}$ diagonalizes $\widehat{\Sigma}(t_j)$. Assumption 1 guarantees that the fourth moment of $X(t_j)$ is bounded over j and thus the sample covariance $\widehat{\Sigma}(t_j)$ converges to the true value $\Sigma(t_j)$ in both a.s. and L^2 means with rate $n^{-1/2}$. Therefore, the following limit exists and

$$V_j = \lim_{n\to\infty} V_{n,j} = P(t_j)^T \Sigma(t_j) P(t_j) = D(t_j).$$

Besides, the convergence speeds of $\{V_{n,j}\}_{1\le j\le m}$ are uniform over j. According to Theorem 5.1.4 and Corollary 5.1.5 of Hsing and Eubank (2015), if $\|V_{n,j} - V_j\|_{\mathrm{HS}}$ is sufficiently small relative to κ, there exists a constant C (depending on κ) such that $\|I - S_{n,j}\|_{\mathrm{HS}} \le C \|V_{n,j} - V_j\|_{\mathrm{HS}}$. Therefore,

$$\|\widehat{P(t_j)} G_{n,j} - P(t_j)\|_{\mathrm{HS}} = \|P(t_j)(S_{n,j} - I))\|_{\mathrm{HS}} \le \sqrt{d} C \|V_{n,j} - V_j\|_{\mathrm{HS}}.$$

The convergence of $V_{n,j}$ to V_j gives the convergence of $\widehat{P(t_j)} G_{n,j}$ to $P(t_j)$,

$$\lim_{n\to+\infty} \widehat{P(t_j)} G_{n,j} = P(t_j) \quad a.s., \quad \text{and} \quad \sup_{1\le j\le m} E \|\widehat{P(t_j)} G_{n,j} - P(t_j)\|_{\mathrm{HS}}^2 = O(n^{-1}).$$

Since $G_{n,j}$ satisfies $\lim_{n\to+\infty} \widehat{P(t_j)} G_{n,j} = P(t_j)$ a.s., $\widehat{P(t_j)} G_{n,j}$ and $P(t_j)$ are relatively close to each other when n is large enough. On the other hand, the two components M^+ and M^- of orthogonal matrix space M are two distinct compact sets and thus the distance between them has a positive lower bound, meaning there exists $\varepsilon_M > 0$ such that

$$\inf_{P_1 \in M^+, P_2 \in M^-} d_M(P_1, P_2) > \varepsilon_M.$$

Therefore, as n goes to infinity, $\widehat{P(t_j)} G_{n,j}$ and $P(t_j)$ lie in the same components almost surely and thus Lemma 7 is applicable. The following statements are directly obtained by Lemmas 7 and 8:

$$\sup_{1\le j\le m} E\, d_M(\widehat{P(t_j)} G_{n,j}, P(t_j))^2 = O(n^{-1}), \quad \sup_{1\le j\le m} E\, d_N([Q_j], [P(t_j)])^2 = O(n^{-1}).$$

□

Proof of Theorem 2. We divide the proof into three steps: step one presents a decomposition of $\int_0^1 (\tilde{\alpha}^{kl}(t) - \tilde{P}^{kl}(t))^2 dt$, step two gives a convergence rate of it, and step three gives the convergence rate of $\int_0^1 d_N^2(\alpha(t), [P(t)]) dt$.

Step 1: Let $\tilde{P}(t) = R_{\gamma(0)}^{-1} \circ d\phi_{\gamma(0)}^{-1}([\acute{P}(t)])$ and $\tilde{r}(t) = R_{\gamma(0)}^{-1} \circ d\phi_{\gamma(0)}^{-1}(\acute{r}(t))$. Write $\delta_j = \tilde{\alpha}_j^{kl} - \tilde{P}^{kl}(t_j)$ and let Δ be the linear interpolation of $\{(t_j, \delta_j)\}_{1\le j\le m}$, i.e.,

$$\Delta(t) = \delta_1 I_{t\in[0,t_1]} + \sum_{j=1}^{m-1} \left(\delta_j \frac{t_{j+1}-t}{t_{j+1}-t_j} + \delta_{j+1} \frac{t-t_j}{t_{j+1}-t_j}\right) I_{t\in[t_j,t_{j+1}]} + \delta_m I_{t\in[t_m,1]}.$$

Let $\mathcal{U}$ be the functional operator associated with the twice order spline interpolation, i.e., $\mathcal{U}(g)$ is the solution to

$$\arg\min_{f\in W^{2,2}} \int_0^1 |\tfrac{d^2}{ds^2} f(s)|^2 ds \quad \text{subject to} \quad f(t_j) = g(t_j) \quad (1 \le j \le m).$$

Then $\tilde{\alpha}^{kl}(t) = \mathcal{U}(\Delta(t) + \tilde{P}^{kl}(t))$. Recall that the $\mathcal{U}$ is a linear operator satisfying $\mathcal{U}(g_1+g_2) = \mathcal{U}(g_1) + \mathcal{U}(g_2)$ (DeVore and Lorentz, 1993). By the triangular inequality,

$$\int_0^1 (\tilde{\alpha}^{kl}(t) - \tilde{P}^{kl}(t))^2 dt \le 2\int_0^1 (\mathcal{U}(\tilde{P}^{kl}) - \tilde{P}^{kl})^2 dt + 2\int_0^1 (\mathcal{U}(\Delta))^2 dt. \tag{S.2}$$

Step 2: The first term on the right-hand side in (S.2) represents the approximation error of spline interpolation for $\tilde{P}^{kl}(t)$. It can be bounded by DeVore and Lorentz (1993)

$$\int_0^1 \left(\mathcal{U}(\tilde{P}^{kl}) - \tilde{P}^{kl}\right)^2 dt \le C(\max_j |t_j - t_{j+1}|^4) \int_0^1 (\tfrac{d^2}{dt^2} \tilde{P}^{kl}(t))^2 dt.$$

From Lemma 13 and the assumptions that $\max_{1\le j\le m} |t_j - t_{j-1}| < \frac{C_4}{m}$, we have

$$\int_0^1 (\mathcal{U}(\tilde{P}^{kl}) - \tilde{P}^{kl})^2 dt = O_p(m^{-4}). \tag{S.3}$$

It remains to bound $\int_0^1 \left(\mathcal{U}(\Delta)\right)^2 dt$ in (S.2). The proof of Theorem 2.2 in Cai and Yuan (2011) suggests that $\int_0^1 \left(\mathcal{U}(\Delta)\right)^2 dt \le C\int_0^1 \Delta(t)^2 dt$ for some constant C. Thus, we have

$$\begin{aligned}\int_0^1 (\mathcal{U}(\Delta))^2 dt \le& C\sum_{j=1}^m \delta_j^2 (t_{j+1} - t_{j-1}) \le C\tfrac{1}{m}\sum_{j=1}^m (\tilde{\alpha}^{kl}(t_j) - \tilde{P}^{kl}(t_j))^2 \\ \le& 2C\tfrac{1}{m}\sum_{j=1}^m \{(\tilde{\alpha}^{kl}(t_j) - \tilde{Q}_j^{kl})^2 + (\tilde{P}^{kl}(t_j) - \tilde{Q}_j^{kl})^2\}.\end{aligned}$$

By the definition of $\tilde{\alpha}^{kl}(t)$ and Lemma 13, we have

$$\begin{aligned}\tfrac{1}{m}\sum_{j=1}^m (\tilde{\alpha}^{kl}(t_j) - \tilde{Q}_j^{kl})^2 \le& \tfrac{1}{m}\sum_{j=1}^m (\tilde{\alpha}^{kl}(t_j) - \tilde{Q}_j^{kl})^2 + \rho\int_0^1 (\tfrac{d^2}{dt^2}\tilde{\alpha}^{kl}(t))^2 dt \\ \le& \tfrac{1}{m}\sum_{j=1}^m (\tilde{P}^{kl}(t_j) - \tilde{Q}_j^{kl})^2 + \rho\int_0^1 (\tfrac{d^2}{dt^2}\tilde{P}^{kl}(t))^2 dt.\end{aligned}$$

Then

$$\int_0^1 (\mathcal{U}(\Delta))^2 dt \le 4C[\tfrac{1}{m}\sum_{j=1}^m (\tilde{P}^{kl}(t_j) - \tilde{Q}_j^{kl})^2 + \rho\int_0^1 (\tfrac{d^2}{dt^2}\tilde{P}^{kl}(t))^2 dt].$$

Proposition 3 suggests that $\lim_{n\to\infty} \mathrm{pr}\{\sup_{t\in[0,1]} d_N([P(t)], r(t)) < R\} = 1$. Consider the case when $\sup_{t\in[0,1]} d_N([P(t)], r(t)) < R$ holds. By Theorem 1,

$$\begin{aligned}&\sup_{1\le j\le m} E\,(\tilde{P}^{kl}(t_j) - \tilde{Q}_j^{kl})^2 \le \sup_{1\le j\le m} E\,\|\tilde{P}(t_j) - \tilde{Q}_j\|_{\mathrm{HS}}^2 = \sup_{1\le j\le m} E\,\|[P(\acute{t}_j)] - [\acute{Q}_j]\|_{\mathrm{HS}}^2 \\ =& \sup_{1\le j\le m} E\,\|\mathrm{Log}_{[Q_j]}([P(t_j)])\|_{\mathrm{HS}}^2 = \sup_{1\le j\le m} E\,\{d_N([Q_j],[P(t_j)])^2\} = O(n^{-1}).\end{aligned}$$

Thus, the bound of $\int_0^1 (\mathcal{U}(\Delta))^2 dt$ is $\int_0^1 (\mathcal{U}(\Delta))^2 dt = O_p(n^{-1} + \rho)$. Inserting this and (S.3) into (S.2) and setting $\rho \asymp O(m^{-4} + n^{-1})$, we deduce $\int_0^1 (\tilde{\alpha}^{kl}(t) - \tilde{P}^{kl}(t))^2 dt = O_p(m^{-4} + n^{-1})$. Since the dimension d is fixed,

$$\int_0^1 \|\acute{\alpha}(t) - [P(\acute{t})]\|_{\mathrm{HS}}^2 dt = \sum_{k,l=1}^d \int_0^1 (\tilde{\alpha}^{kl}(t) - \tilde{P}^{kl}(t))^2 dt = O_p(m^{-4} + n^{-1}).$$

Step 3: We claim that $\int_0^1 d_N^2([P(t)],\alpha(t))dt = O_p(m^{-4}+n^{-1})$. It is sufficient to show that for any $\varepsilon > 0$, there exists $M > 0$ such that

$$\limsup_{m,n\to\infty} \mathrm{pr}\{\int_0^1 d_N^2([P(t)],\alpha(t))dt > M(m^{-4}+n^{-1})\} < \varepsilon.$$

Proposition 3 suggests that $\lim_{n\to\infty} \mathrm{pr}\{\sup_{t\in[0,1]} d_N([P(t)],r(t)) < \frac{1}{2}R\} = 1$. Consequently, for sufficiently large n, the probability of the event $\Omega_1 = \{\sup_{t\in[0,1]} d_N([P(t)],r(t)) \geq \frac{1}{2}R\}$ is less than $\varepsilon/3$. On Ω/Ω_1, $\|[\acute{P}(t)] - \acute{r}(t)\|_{\mathrm{HS}} \leq \frac{1}{2}R$. In addition, since $\int_0^1 \|\acute{\alpha}(t) - [\acute{P}(t)]\|_{\mathrm{HS}}^2 dt = O_p(m^{-4}+n^{-1})$, there exists M_1 such that

$$\limsup_{m,n\to\infty} \mathrm{pr}\{\int_0^1 \|\acute{\alpha}(t) - [\acute{P}(t)]\|_{\mathrm{HS}}^2 dt > M_1(m^{-4}+n^{-1})\} < \frac{1}{3}\varepsilon.$$

Define $\Omega_2 = \{\int_0^1 \|\acute{\alpha}(t) - [\acute{P}(t)]\|_{\mathrm{HS}}^2 dt > M_1(m^{-4}+n^{-1})\}$. For m,n sufficiently large, $P(\Omega_2) < \frac{1}{3}\varepsilon$. On $\Omega/(\Omega_1\cup\Omega_2)$, $\int_0^1 \|\acute{\alpha}(t) - [\acute{P}(t)]\|_{\mathrm{HS}}^2 dt \leq M_1(m^{-4}+n^{-1})$ and thus

$$\mathrm{pr}_t\{t : \|\acute{\alpha}(t) - [\acute{P}(t)]\|_{\mathrm{HS}} > \frac{1}{2}R\} \leq \frac{4M_1}{R^2}(m^{-4}+n^{-1}),$$

where pr_t denotes the Lebesgue measure on $[0,1]$. Furthermore, on $\Omega/(\Omega_1\cup\Omega_2)$, we have

$$\mathrm{pr}_t\{t : \|\acute{\alpha}(t) - \acute{r}(t)\|_{\mathrm{HS}} \geq R\} \leq \frac{4M_1}{R^2}(m^{-4}+n^{-1}).$$

According to Lemma 1, when $\omega \in \Omega/(\Omega_1\cup\Omega_2)$ and $\|\acute{\alpha}(t) - \acute{r}(t)\|_{\mathrm{HS}} < R$, we have $C_1 d_N(\alpha(t),r(t)) \leq \|\acute{\alpha}(t) - \acute{r}(t)\|_{\mathrm{HS}}$. Suppose that the diameter of N is $\mathrm{diam}(N) < \infty$. Then when $\omega \in \Omega/(\Omega_1\cup\Omega_2)$,

$$\begin{aligned}\int_0^1 d_N(\alpha(t),r(t))^2 dt =& \int_{\|\acute{\alpha}(t)-\acute{r}(t)\|_{\mathrm{HS}}<R} d_N^2(\alpha(t),r(t))dt + \int_{\|\acute{\alpha}(t)-\acute{r}(t)\|_{\mathrm{HS}}\geq R} d_N^2(\alpha(t),r(t))dt \\ \leq& \int_0^1 \frac{1}{C_1^2}\|\acute{\alpha}(t) - \acute{r}(t)\|_{\mathrm{HS}}^2 dt + \frac{4M_1}{R^2}(m^{-4}+n^{-1})\mathrm{diam}(N)^2 \\ \leq& \frac{1}{C_1^2} M_1(m^{-4}+n^{-1}) + \frac{4M_1}{R^2}(m^{-4}+n^{-1})\mathrm{diam}(N)^2 \\ =& (\frac{M_1}{C_1^2} + \frac{4M_1 \mathrm{diam}(N)^2}{R^2})(m^{-4}+n^{-1}).\end{aligned}$$

Set $M = \frac{M_1}{C_1} + \frac{4M_1 \mathrm{diam}(N)^2}{R^2}$. Then

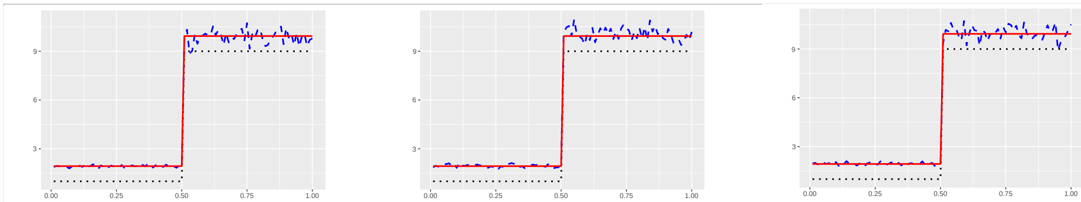
$$\limsup_{m,n\to\infty} \mathrm{pr}\{\int_0^1 d_N^2([P(t)],\alpha(t))dt > M(m^{-4}+n^{-1})\} \leq P(\Omega_1) + P(\Omega_2) < \varepsilon.$$

The convergence of the estimated dynamic components comes directly from that of eigen-frame estimates. □

S.3. Additional simulation on influence of noise

Section 2.1 states that the noise distribution does not affect the eigen-frame when the variance matrix of noise is scalar. This section numerically examines the influences of different noise types on the eigen-frame estimate. Consider the case $d = 3$ for illustration and generate data according to

$$X_i(t_j) = \sum_{k=1}^{3} \psi_i^k(t_j) e_k(t_j) = \sum_{k=1}^{3} \psi_i^{*k}(t_j) e_k(t_j) + \varepsilon_{ij}.$$



Normal noise Uniform noise Laplace noise

Figure S.1: The estimate of the first eigenvalue λ_1 under $n = 1000$ and $m = 100$ for three types of noise. The dashed, dotted and solid lines refer to the estimate $\hat{\lambda}_1$, the true values λ_1^* and λ_1, respectively.

Table S.1. The Monte Carlo mean, standard deviation of mean square errors (multiplied by 10^4) and run time for normal noise, uniform noise and Laplace noise

normal noise	$m = 60$		$m = 80$		$m = 100$	
$n = 100$	71.2 (33.7)	35.2 s	53.7 (23.8)	48.8 s	49.3 (25.7)	81.3 s
$n = 300$	22.7 (10.0)	36.7 s	18.7 (8.35)	57.8 s	14.9 (5.94)	83.8 s
$n = 600$	12.9 (5.42)	39.0 s	10.5 (4.04)	53.5 s	7.75 (2.53)	88.2 s
$n = 1000$	8.50 (3.48)	37.6 s	6.15 (2.38)	62.2 s	5.21 (1.73)	89.0 s
uniform noise	$m = 60$		$m = 80$		$m = 100$	
$n = 100$	71.9 (32.7)	38.6 s	56.8 (25.1)	59.9 s	45.9 (21.1)	80.1 s
$n = 300$	22.9 (8.34)	39.3 s	18.3 (8.68)	58.1 s	15.4 (6.19)	83.4 s
$n = 600$	12.3 (4.23)	40.3 s	9.69 (4.26)	55.3 s	8.28 (2.92)	88.3 s
$n = 1000$	8.24 (3.12)	40.4 s	6.18 (1.81)	52.2 s	5.20 (1.58)	85.7 s
Laplace noise	$m = 60$		$m = 80$		$m = 100$	
$n = 100$	78.5 (36.6)	39.5 s	60.8 (26.2)	49.8 s	46.7 (22.4)	70.4 s
$n = 300$	23.6 (9.70)	32.2 s	17.5 (7.47)	59.8 s	15.6 (5.50)	85.1 s
$n = 600$	12.1 (4.26)	38.3 s	9.71 (4.06)	60.1 s	8.03 (3.02)	86.2 s
$n = 1000$	7.95 (3.18)	37.5 s	6.34 (2.39)	50.3 s	5.60 (2.02)	86.7 s

The dynamic principal components $\{e_k(t)\}_{1\le k\le d}$ are parameterized by the three-dimensional polar coordinate in (3) in the main text with $\theta = \theta(t) = \sin(2\pi t)$ and $\phi = \phi(t) = \sin(\pi t)/2$. The observed time points $\{t_j\}_{1\le j\le m}$ are generated equidistantly. The underlying scores are independent and identically distributed as $\mathrm{Normal}(0, \lambda_k^*(t_j))$ distribution with the underlying eigenvalues $\{\lambda_k^*(t)\}_{k=1}^d$ set as $\lambda_k(t) = k^2 I_{\{0\le t<0.5\}} + (d+1-k)^2 I_{\{0.5<t\le 1\}}$. The vector-valued noise $\varepsilon_{ij} = (\varepsilon_{ij}^1, \varepsilon_{ij}^2, \varepsilon_{ij}^3)$ is set to be the following types with $E\left(|X_1^*(t_j)|^2\right)/E\left(|\varepsilon_{1j}|^2\right) = 5$: normal distribution $\mathrm{Normal}(0, \Sigma_0)$ with $\Sigma_0 = 0.9333 I_d$, uniform distribution $\mathrm{Uniform}([-a, a]^d)$ with $a = 1.67$, and Laplace (double-exponential) distribution $\mathrm{Laplace}(0, b)$ with $b = 0.68$ for each component ε_{ij}^k. Under above noise settings, the contaminated eigenvalues are $\lambda_k(t_j) = \lambda_k^*(t_j) + \sigma^2$ with $\sigma^2 = 0.9333$. Figure S.1 presents the estimates of the first eigenvalue λ_1, suggesting that the estimates $\hat{\lambda}_1$ successfully recovers the true value of eigenvalue, and behaves similar among the three types of noise distributions. The mean square errors $\int_0^1 E\, d_N^2(\alpha(s), [P(s)])ds$ of our proposed estimate under $n = 100, 300, 600, 1000$ and $m = 60, 80, 100$ are assessed by Mento Carlo method with 100 runs. Table S.1 suggests that the mean square error and run time among the three noise types are close as expected since the noise with scalar variance matrix does not affect the eigen-frame part in the eigen-decomposition.

S.4. Extension of one-step unrolling method to general Lie group and beyond

The one-step unrolling method, introduced in Section 3, addresses curve fitting on the orthogonal matrix group M (or its quotient space N). In this section, we demonstrate how this method can be extended to a general Lie group $\mathcal{M}$ for curve fitting and present the fundamental framework for this extension. We begin by extending the unrolling and unwrapping procedures, as proposed by Kim et al. (2020), to a general Lie group $\mathcal{M}$. The left multiplication L_G (or right multiplication, which is also applicable) plays a crucial role in this extension, as it connects the tangent spaces $T_I\mathcal{M}$ and $T_G\mathcal{M}$, where I denotes the identity element in $\mathcal{M}$. The unrolling of a curve r on $\mathcal{M}$ is represented by $\tilde{r}$ on $T_I\mathcal{M}$, satisfying

$$\tfrac{d}{dt}\tilde{r}(t) = L_{r(t)^{-1}}\big(\tfrac{d}{dt}r(t)\big), \qquad \tilde{r}(0) = 0 \in T_I\mathcal{M}.$$

The unwrapping of a point x on $\mathcal{M}$, with respect to a curve r at time t, is obtained as follows

$$\tilde{x} := \tilde{r}(t) + L_{r(t)^{-1}}(\log_{r(t)} x) \in T_I\mathcal{M}.$$

Utilizing the unrolling and unwrapping procedures above, we implement a one-step unrolling technique based on the observations $\{X_j\}_{j=1}^m \subset \mathcal{M}$ and a regular base curve r that will be addressed later

$$\tilde{\mu} := \arg\min_{f:[0,1]\to T_I\mathcal{M}} \tfrac{1}{m}\textstyle\sum_{j=1}^m \|f(t_j) - \tilde{X}_j\|^2_{T_I\mathcal{M}} + \rho\int_0^1 \|f^{(2)}(t)\|^2_{T_I\mathcal{M}}dt.$$

Here, $\|\cdot\|_{T_I\mathcal{M}}$ represents a norm on the linear space $T_I\mathcal{M}$, $\tilde{X}_j$ denotes the unwrapped point of X_j, and ρ serves as the tuning parameter. Following this, we obtain the estimate $\hat{\mu}$ on $\mathcal{M}$ by wrapping the curve $\tilde{\mu}$, which is

$$\hat{\mu}(t) = \exp_{r(t)}\{L_{r(t)}(\tilde{\mu}(t) - \tilde{r}(t))\}.$$

Now, it remains to construct the regular base curve r on $\mathcal{M}$ from the observations $\{X_j\}_{j=1}^m \subset \mathcal{M}$. Following a similar technique to that in Section B, we first select a subset $\{t_{j_i}\}_{i=1}^D$ of the observed time points, which satisfies Assumption 2. Then, we construct the piece-wise geodesic curve $r_g(t)$ that connects the points $\{X_{j_i}\}_{1\le i\le D}$ in the interval $[t_{j_1}, t_{j_D}]$ and extends onto the intervals $[0, t_{j_1}]$ and $[t_{j_D}, 1]$ geodesically at the boundaries. More specifically, r_g is geodesic on the interval $[0, t_{j_2}]$ with $r_g(t_{j_1}) = X_{j_1}$ and $r_g(t_{j_2}) = X_{j_2}$ fixed, and on the interval $[t_{j_{D-1}}, 1]$ with $r_g(t_{j_{D-1}}) = X_{j_{D-1}}$ and $r_g(t_{j_D}) = X_{j_D}$ fixed. Then $r_g(t)$ satisfies the differential equation $r'_g(t) = r_g(t)z_g(t)$ with

$$z_g(t) = a_1 I_{\{0\le t\le t_{j_1}\}} + \textstyle\sum_{i=1}^{D-1} a_i I_{\{t_{j_i}\le t\le t_{j_{i+1}}\}} + a_{D-1} I_{\{t_{j_D}\le t\le 1\}}$$

where $\{a_i\}_{1\le i\le D-1}$ are identified as elements in $T_I\mathcal{M}$ since $z_g(t) = L_{(r_g(t))^{-1}} r'_g(t) \in T_I\mathcal{M}$. The curve $z_g(t)$, which can be viewed as the first derivative of $r_g(t)$, is a step function. To obtain a continuous function, we can transform $z_g(t)$ into $z(t)$ by smoothing out the discontinuous points. Specifically, $z(t)$ can be defined by

$$z(t) = \begin{cases} \frac{a_i - a_{i-1}}{\frac{1}{4}h_i + \frac{1}{4}h_{i-1}}(t - t_i) + \frac{a_i h_{i-1} + a_{i-1}h_i}{h_i + h_{i-1}}, & (t_{j_i} \le t \le t_{j_i} + \frac{1}{4}h_i), \\ a_i, & (t_{j_i} + \frac{1}{4}h_i \le t \le t_{j_i} + \frac{3}{4}h_i), \\ \frac{a_{i+1} - a_i}{\frac{1}{4}h_{i+1} + \frac{1}{4}h_i}(t - t_{i+1}) + \frac{a_{i+1}h_i + a_i h_{i+1}}{h_{i+1} + h_i}, & (t_{j_i} + \frac{3}{4}h_i \le t \le t_{j_{i+1}}), \end{cases}$$

where $h_i = t_{j_{i+1}} - t_{j_i}$. Motivated by the relationship $r_g'(t) = r_g(t)z_g(t)$, we define the regular base curve r as the solution to the following differential equation on $\mathcal{M}$,

$$\frac{d}{dt}r(t) = r(t)z(t), \quad r(t_{j_1}) = X_{j_1}.$$

The aforementioned discussion presents a basic framework for performing one-step unrolling on a general Lie group $\mathcal{M}$.

The above method also holds the promise of extension to a broader class of Riemannian manifolds. The key is to propose an approach to construct a smooth base curve on a Riemannian manifold $\mathcal{M}$. Beginning with observations $\{X_j\}_{j=1}^m$, we select a subset of time points $\{t_{j_i}\}_{i=1}^D$ conforming to Assumption 2. A piece-wise geodesic curve $r_g(t)$ is then constructed, connecting data points $\{X_{j_i}\}_{1\leq i\leq D}$ over the interval $[t_{j_1}, t_{j_D}]$, and is extended beyond this interval through geodesic extrapolation at the endpoints. Locally, the Riemannian logarithm map $\mathrm{Log}_{r_g(t_{j_i})}(\cdot)$ projects the neighborhood of X_{j_i} to an open set in $\mathbb{R}^d$. Specifically, the map renders the image $\mathrm{Log}_{r_g(t_{j_i})}\left\{r_g([t_{j_{i-1}}, t_{j_{i+1}}])\right\}$ as two contiguous line segments. Adjustments are then made around the connection point to smooth the image, resulting in a smooth curve. The cumulation of these local modifications for all $1 \leq i \leq D$ yields the desired smooth regular base curve r. Then the unrolling and estimation operations can be conducted by similar methods.

References

CAI, T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: phase transition. *Ann. Statist.* **39** 2330–2355.

DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation* **303**. Springer Science & Business Media.

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley.

KIM, K.-R., DRYDEN, I. L., LE, H. and SEVERN, K. E. (2020). Smoothing splines on Riemannian manifolds, with applications to 3D shape space. *J. R. Statist. Soc. Ser. B-Statist. Methodol* **83** 108–132.