Bernoulli **31**(1), 2025, 649–670
https://doi.org/10.3150/24-BEJ1743

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*

Principal component analysis has been one of prominent techniques in multi-dimensional statistics, and the core is to estimate the eigenvalues in descending ordering with their corresponding eigenvectors. However, when data are collected longitudinally in many scientific applications, the eigenvalues become dynamic over time, and the ordering of them may not be well-defined, for instance, does not coincide with the pointwise ordering. To deal with this issue, we propose a new framework, namely the dynamic principal component analysis. This addresses the identifiability of principal components from a global perspective, and transforms the problem into a regression model for data situated on the orthogonal matrix group space. The one-step unrolling method is exploited to solve the regression problem with a suitably constructed regular base curve. The minimax rate of the proposed estimators is established through theoretical analysis of the one-step unrolling and its connection to smoothing spline in the context of manifold-valued data.

Keywords: Eigenvalue switch; discrete data; one-step unrolling; orthogonal matrix group; smoothing spline

1. Introduction

Principal component analysis has been one of prominent techniques in multi-dimensional data analysis for decades. The traditional principal component analysis, introduced by Pearson (1901), focuses on eigenvectors with leading eigenvalues in correspondence, named as principal components. In this sense, the principal components together cover most of the variability in population and thus attract much attention due to their practical applications in statistical inference (Jolliffe, 2002). When multi-dimensional data are collected longitudinally, the dynamic eigenvectors and eigenvalues become functions of time. Subsequently, the indices of leading eigenvalues may also change with time, resulting in confusion of identifying the principal components. Figure 1 showcases visual illustrations of the dynamic eigenvectors, scaled by their corresponding dynamic eigenvalues, for dimensions $d = 2, 3$. In these examples, the dynamic eigenvalues demonstrate an order-switching phenomenon, leading to variations in the indices associated with the predominant eigenvalues as time progresses.

To address this challenge, we propose for the first time the notion of dynamic principal components whose corresponding eigenvalue functions cover most variability in a global sense that the magnitude ordering of them is determined by their L^p norms, saying the first dynamic eigenvalue λ_1 possesses the greatest L^p norm $\|\lambda_1\|_{L^p}$ among others. This global perspective captures the main characteristic along the whole time course and guarantees the identifiability of the principal components in the dynamic mechanism, which avoids the requirement in Hu and Yao (2024) that the pointwise magnitude order of eigenvalues stay unchanged over time. The adoption of the L^p norm also allows the eigenvalue functions to change discontinuously, including the often-encountered scenario that the variance for a certain eigen-direction changes radically at some point. The concept of dynamic principal components in this paper is fundamentally different from the fields of principal subspace tracking (Bordin

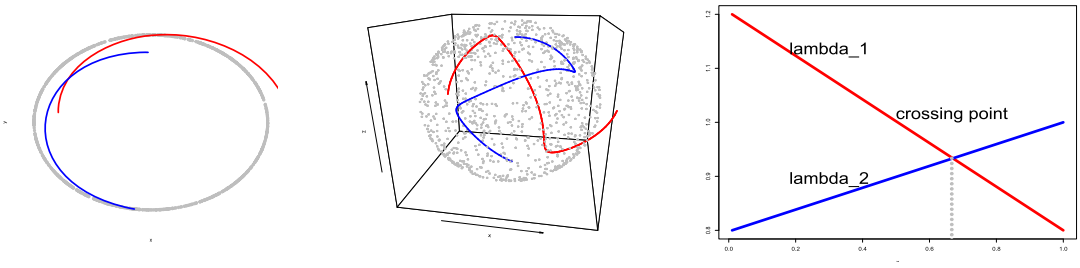


Figure 1. The left and middle plots present the leading two dynamic eigenvectors $\lambda_k(t)e_k(t)$ $(k = 1, 2)$, scaled by their corresponding dynamic eigenvalues, for dimensions $d = 2$ and $d = 3$, respectively. The eigenvalues for both plots are $\lambda_1 = 1.2 - 0.4t$ and $\lambda_2 = 0.8 + 0.2t$, and they intersect at $t = 2/3$ as shown in the right plot. Furthermore, the ancillary grey dots illustrate the one-dimensional and two-dimensional unit spheres in the left and middle plots, respectively.

and Bruno, 2020, Srivastava and Klassen, 2004) or the functional principal component analysis (Yao, Müller and Wang, 2005), and also distinguishes from the dynamic version of multidimensional/functional principal component analysis for time series (Brillinger, 1981, Hörmann, Kidziński and Hallin, 2015).

In the dynamic mechanism, time-evolving data are often observed on a grid of time points $\{t_j\}_{j=1}^m$, like the number of daily new cases of Covid-19 or annual GDP of countries, and thus is the main concern of our paper. The multi-dimensional eigen-decomposition of the data on each time point t_j yields an approximation to the eigenvectors, but suffers from identifiability issue in the following sense. The eigenvector with the largest eigenvalue on this particular t_j does not necessarily be a time section of the leading principal component at other times, while it is not desirable to assign the leading eigenvectors by their eigenvalues pointwisely. To overcome this issue, we turn to estimate the whole continuous eigen-frame, consisting of all dynamic eigenvectors over the time course, where the dynamic principal components can be selected by their corresponding eigenvalue function estimates possessing leading L^p norms. In order to properly represent such dynamic principal components, the main challenge is that the eigen-frame estimation is nontrivial and is cast into a regression problem that resorts to curve fitting techniques on the orthogonal matrix space which is a non-flat manifold.

In the literature, fitting data on manifolds has been studied extensively. The studies of smooth interpolation and smoothing spline methods for manifold-valued data have drawn attention from both statisticians and computational mathematicians (Green and Silverman, 1994). A natural way to directly conduct the smooth spline procedures on a manifold is to redefine the fitting errors and the regularity with respect to the intrinsic geometric properties of the manifold. Machado and Leite (2006) explored the way of constructing piecewise geodesic interpolation on a manifold and smoothing spline on sphere, and were then generalized by Rodrigues, Silva Leite and Jakubiak (2005), Jakubiak, Silva Leite and Rodrigues (2006) and Rui and Leite (2007) to smooth interpolation on Lie group and sphere. A representative work in directly fitting smoothing spline on general manifolds was the gradient-descent algorithm proposed in Samir et al. (2012), which was an iteration method and required analytical expressions involving parallel transport and covariant integral along curves. Apart from directly fitting smoothing spline on manifolds, projecting data into a common tangent space was also a useful approach. This projection was realized by a Riemannian logarithm map and allowed one to deal with the data in a local neighborhood (Clark and Thompson, 1984). Jupp and Kent (1987) improved the Riemannian logarithm map and proposed unrolling by iteration method to deal with the manifold-valued data on a connected component and then Kim et al. (2021), Le (2003) extended this to more general

manifolds. This technique transfers the data on a manifold into a common linear space where the statistical method developed for Euclidean spaces can be adopted. The above-mentioned works considered the fit of smoothing spline's for given deterministic points from the computational mathematics perspective. By contrast, in statistics, data are treated as random samples from some distribution and the spline fitting is usually viewed as the mean estimation of the underlying population. Statisticians have shown that the spline curve estimation converges to the mean of the population (Cai and Yuan, 2011) in Euclidean case. However, there is no statistical research showing whether it is the case for manifolds, which turns out rather challenging to prove so if one operates smoothing spline directly on manifolds or adopts iterative procedures. Hence we propose a new method named one-step unrolling that improves upon and distinguishes from the unrolling by iteration method to realize the eigen-frame estimation on manifold. To overcome theoretical obstacles and derive the convergence rate, we suggest a novel way to choose the regular base curve that is essential for the performance of the proposed method. As a result, our method avoids the iterative procedure in the unrolling by iteration method (Kim et al., 2021, Le, 2003) or the gradient-descent algorithm (Samir et al., 2012), thus saving computing time.

In summary, we propose the following approach to realize dynamic principal component analysis. In the first step, the standard multi-dimensional principal component analysis is performed on the observed time points and produces elements representing the eigen-frame consisting of all eigenvectors. Subsequently, we cast the problem of dynamic eigen-frame estimation into a regression problem for data situated on a manifold. This is the first work to consider the dynamic principal component analysis from the perspective of orthogonal matrix group without regular conditions on the eigenvalues and allow discontinuity/order-switching to the best we have known. In the second step, we propose a new smoothing spline technique, namely the one-step unrolling for manifold-valued data, to produce the estimation of eigen-frame for the whole time interval, which reduces computation and attains the minimax rate of convergence as in Euclidean case. In the third step, the eigenvalue function in correspondence to each estimated dynamic eigenvector function is recovered and the dynamic principal components are determined by their L^p norms.

2. Model and motivation

2.1. Model setup

Let $X : [0,1] \times \Omega \to \mathbb{R}^d$ be the centered random process and of second order in the sense that $E(X(t)) = 0$ and $\Sigma(t) = E(X(t)X(t)^T) < \infty$ for all time points $t \in [0,1]$. For each fixed t_0, standard principal component analysis yields $X(t_0) = \sum_{k=1}^{d} \psi^k(t_0) e_k(t_0)$, where $e_k(t_0)^T e_l(t_0) = 0$ and $\mathrm{cov}(\psi^k(t_0), \psi^l(t_0)) = 0$ for $k \neq l$. All the smooth time-varying eigenvectors $\{e_k(\cdot)\}_{k=1}^{d}$ compose the dynamic eigen-frame. The time-varying eigenvalues $\lambda_k(t) = \mathrm{var}(\psi^k(t))$ are ordered by the global L^p norms $\|\lambda_k\|_{L^p} = \{\int_0^1 \lambda_k^p(t)dt\}^{\frac{1}{p}}$, saying $\|\lambda_1\|_{L^p} \geq \cdots \geq \|\lambda_d\|_{L^p}$. We stress that the eigenvalues at one specific time point t_0 does not necessarily satisfy $\lambda_1(t_0) \geq \cdots \geq \lambda_d(t_0)$. We expect the dynamic eigen-frame to vary smoothly over time while eigenvalues are relaxed to be discontinuous. In reality, the independent and identically distributed realizations $X_1, \ldots, X_n$ of sample size n are often observed intermittently. To accommodate this practice, we assume that the observed data are sampled on a common grid, i.e., the observed time points $\{t_j\}_{j=1}^{m}$ and samples $\{X_i(t_j)\}_{1 \leq i \leq n, 1 \leq j \leq m}$. Subsequently, the observations admit the dynamic representation $X_i(t_j) = \sum_{k=1}^{d} \psi_i^k(t_j) e_k(t_j)$. This paper addresses the analysis of multi-dimensional dynamic data wherein the dimension d is fixed and less than the sample size n, as opposed to high-dimensional scenarios.

The observed process considered may be composed of an underlying process X^* and an uncorrelated random noise ε. In theory, only $\mathrm{var}(X) = \mathrm{var}(X^*) + \mathrm{var}(\varepsilon)$ are estimable from the samples, while

$\text{var}(X^*)$ and $\text{var}(\varepsilon)$ are unidentifiable with no further assumptions. In practice, the data are often pre-processed such that every component has the similar scale prior to performing eigen-analysis, which may lead to a common scale of noise over different components. In particular, if the noise possesses an isotropic variance structure $\text{var}(\varepsilon) = \sigma^2(t)I_d$, the dynamic eigen-frame $\{e_k(\cdot)\}_{k=1}^d$ and dynamic principal components of the observed process are necessarily identical to those of the underlying process, since the scalar matrix of noise variance does not affect the matrix eigen-decomposition. To avoid this type of assumptions, we aim to recover the eigen-frame and the dynamic components of the observed process X rather than the underlying process X^*.

In this paper, we highlight that the notion of dynamic principal component analysis fundamentally diverges from that of functional principal component analysis, as elucidated below.

- Firstly, the dynamic principal component analysis can be considered as an extension of the classical multi-dimensional principal component analysis, wherein it incorporates the aspect of temporal evolution. Conceptually, dynamic principal component analysis views $X_1(t), \ldots, X_n(t)$ at each time t as a collection of vectors, thereby effectively capturing the time-varying features of dynamic eigenvectors and eigenvalues. In contrast, functional dynamic component analysis treats each random process as an atomic object, with the information about varying time being concealed within the corresponding eigenfunction.
- Secondly, dynamic principal component analysis applies the eigen-decomposition to the vectors $X_1(t), \ldots, X_n(t)$ in $\mathbb{R}^d$ to obtain the representation $X_i(t) = \sum_{k=1}^d \psi_i^k(t)e_k(t)$. For a fixed t, $\{e_k(t)\}_{k=1}^d$ form an orthonormal basis of $\mathbb{R}^d$. By contrast, functional principal component analysis requires eigen-analysis in an infinite-dimensional function space $L^2([0,1];\mathbb{R}^d)$ to represent $X_i(t) = \sum_{k=1}^\infty \xi_{ik}\varphi_k(t)$, where the functions $\{\varphi_k(\cdot)\}_{k=1}^\infty$ form an orthonormal basis of $L^2([0,1];\mathbb{R}^d)$, the space of square integrable functions on $[0,1]$ to $\mathbb{R}^d$, and $\{\xi_{ik}\}_{k=1}^\infty$ are uncorrelated random variables. Hence, dynamic principal component analysis possesses a computational advantage compared to functional principal component analysis, as the search for a basis in the relatively simpler space $\mathbb{R}^d$ is less complex than in the more intricate space $L^2([0,1];\mathbb{R}^d)$.
- Thirdly, the dynamic principal component analysis methodology proposed in this paper necessitates solely the smoothness of the dynamic eigen-frame $\{e_k(\cdot)\}_{k=1}^d$, permitting discontinuities in individual realizations X. This contrasts with the smoothness constraints required for all realizations in functional data analysis (Yao, Müller and Wang, 2005).

2.2. Motivation and reformulation

We begin with finding the eigen-frame $\{e_k(t)\}_{k=1}^d$ for a given time point t via the standard matrix eigen-decomposition solving the eigen-equation $\Sigma(t)e(t) = \lambda(t)e(t)$ for an eigenvector $e(t)$ and an eigenvalue $\lambda(t)$. From matrix analysis, it is equivalent to finding a diagonal matrix $D(t)$ and an orthogonal matrix $P(t)$ to diagonalize the covariance matrix $\Sigma(t)$, i.e., $P^T(t)\Sigma(t)P(t) = D(t)$ and $P(t)^T P(t) = I_d$. A solution is the diagonal matrix $D(t)$ with diagonal elements $\lambda_1(t), \ldots, \lambda_d(t)$, and the orthogonal matrix $P(t) = (e_1(t), \ldots, e_k(t))$, i.e., the matrix formed by concatenating the column vectors $e_1(t), \ldots, e_d(t)$. However, the above method to derive dynamic eigen-frame suffers from non-identifiability and consequent issues. The orthogonal matrix is identifiable only up to column permutation and the signs of $e_1(t), \ldots, e_d(t)$. Specifically, for any permutation τ of $\{1, \ldots, d\}$ and variables $s_1, \ldots, s_d \in \{-1, +1\}$, the matrix $\{s_1 e_{\tau(1)}(t), \ldots, s_d e_{\tau(d)}(t)\}$ diagonalizes $\Sigma(t)$ along with the diagonal matrix with the diagonal elements $\lambda_{\tau(1)}(t), \ldots, \lambda_{\tau(d)}(t)$. The non-identifiability can not be overcome by the ordering (ordered by size) of the eigenvalues $\{\lambda_k(t)\}_{k=1}^d$ since they do not necessarily descend over k on this given time point

t. Hence, despite the eigenvector $e_k(t)$ being well-defined and ordered according to the L^p norm associated with its corresponding dynamical eigenvalue, it cannot be ascertained through the matrix eigen-decomposition of the covariance matrix $\Sigma(t)$. This non-identifiability creates substantial difficulty on interpolating the estimates at the grid time points $t_1,\ldots,t_m$, i.e. once the estimates of eigen-frame at the time points $t_1,\ldots,t_m$ are obtained, it is difficult to interpolate them (or adopt other smoothing techniques) to gain the time-varying estimator on the whole time interval. It is noteworthy that Hu and Yao (2024) adopted a restrictive assumption that the dynamic eigenvalues $\lambda_1(t) \geq \cdots \geq \lambda_d(t)$ hold for all $t \in [0,1]$, ensuring that the ordering of dynamic eigenvalues remains static with the passage of time, thereby preventing non-identifiability issues. We have relaxed this restrictive assumption to facilitate its broader practical application. Consequently, a novel estimation approach is required to address the resulting non-identifiability.

The above discussion shows that the non-identifiability stems from non-uniqueness of the diagonalization of a covariance matrix in the $d \times d$ orthogonal matrix space $M = O(d) = \{P \in \mathbb{R}^{d\times d} : P^TP = PP^T = I_d\}$. This motivates us to consider the equivalence class formed by matrices obtained through permuting columns and/or changing the sign of some elements in a column of P. The permutation of the columns of a $d \times d$ matrix P is realized by right multiplication for a $d \times d$ permutation matrix from $\mathcal{G}_1 = \{G \in \mathbb{R}^{d\times d} : G \text{ is a permutation matrix}\} \subset M$, while the change of the signs of elements in a column is realized by right multiplication of a matrix from the family $\mathcal{G}_2 = \{G \in \mathbb{R}^{d\times d} : G \text{ is diagonal, the diagonal elements are } -1, \text{or } +1\} \subset M$. The group generated by $\mathcal{G}_1$ and $\mathcal{G}_2$, meaning the smallest subgroup of M that contains $\mathcal{G}_1$ and $\mathcal{G}_2$, is denoted by $\mathcal{G}$ that is a finite subgroup of M and called the signed permutation group in literature. To be specific, each element $G \in \mathcal{G}$ is generated by a permutation τ of $\{1,2,3,\ldots,d\}$, and the entry of that matrix in the kth row and $\tau(k)$th column is ± 1 for each $1 \leq k \leq d$ while others are 0.

The group $\mathcal{G}$ induces a quotient space N of M, defined by $N = M/\mathcal{G} = \{[P] : P \in M\}$, where $[P] = P\mathcal{G} = \{PG : G \in \mathcal{G}\}$ is an equivalence class containing P, i.e, if for a symmetric matrix V, P^TVP is diagonal, then $(PG)^TV(PG)$ is also diagonal for any $G \in \mathcal{G}$. The geometry of the quotient space N is detailed in Appendix A. Essentially, we identify all orthogonal matrices within an equivalence class while distinguishing them across equivalence classes. When the multiplicity of the eigenvalues of V is one, there is a unique equivalence class in N such that every P in this class diagonalizes V (Lemma 3).

Above discussion motivates us to formulate the matrix eigen-decomposition as the diagonalization of covariance matrix $\Sigma(t)$ by the element $[P(t)] \in N$ to overcome the identifiability issue. In the context of dynamic principal component analysis, we identify the equivalence class $[P(t)]$ that is a curve on N as the eigen-frame projection from M to N. For the data $\{X_i(t_j)\}_{1\leq i\leq n, 1\leq j\leq m}$, standard matrix eigen-decomposition is adopted to obtain the estimates of $[P(t)]$ on the gird time points $\{t_j\}_{j=1}^m$. To be specific, suppose that the sample covariance matrix at t_j is

$$\widehat{\Sigma}(t_j) = n^{-1}\sum_{i=1}^n \big(X_i(t_j)\big)\big(X_i(t_j)\big)^T$$

and the matrix $\widehat{P(t_j)}$ diagonalizes $\widehat{\Sigma}(t_j)$. In practice, the produced matrix $\widehat{P(t_j)}$ is arranged by the magnitude ordering of the eigenvalues on t_j, which does not match the true value $P(t_j)$ in general. We set $[Q_j] = [\widehat{P(t_j)}]$ on N, and consider them as rough estimates of $[P(t)]$ at discrete time points $t = t_1,\ldots,t_m$. The rough estimates $\{[Q_j]\}_{1\leq j\leq m}$ can also be viewed as "observations" of the underlying eigen-frame projection $[P(t)]$ on the manifold N. The next step is to smooth these rough estimates by a newly-proposed smoothing spline method to obtain a smooth estimate $\alpha(t)$ of $[P(t)]$ for the whole time interval $t \in [0,1]$.

3. Dynamic principal components estimation

3.1. Unrolling and unwrapping

In this section, we discuss enhancements to the unrolling and unwrapping techniques (Jupp and Kent, 1987, Kim et al., 2021, Le, 2003) by introducing a regular base curve r on N. The construction and properties of this regular base curve r, based on a subset of the rough estimates $\{[Q_j]\}_{1\le j\le m}$, are purely geometric; thus, the details are deferred to Appendix B. In general, we perform geodesic interpolation on this subset and modify the resulting piecewise geodesic curve to ensure continuous differentiability with bounded second order derivative, as demonstrated in Proposition 2. The Riemannian logarithm maps, $\mathrm{Log}_{r(t)}([P(t)])$ and $\mathrm{Log}_{r(t_j)}([Q_j])$, employed in the subsequent operations are well-posed with a probability tending to one, as corroborated by Proposition 3.

Denote the parallel transport along $r(t)$ by $\mathrm{Par}_{r(t)}^{r(s)} : T_{r(t)}N \to T_{r(s)}N$. The unrolling of r onto $T_{r(0)}N$ is a curve $\acute{r}$ on $T_{r(0)}N$ satisfying

$$\tfrac{d}{dt}\acute{r}(t) = \mathrm{Par}_{r(t)}^{r(0)}\big(\tfrac{d}{dt}r(t)\big), \quad \acute{r}(0) = 0.$$

The left panel of Figure 2 provides a graphic illustration of the unrolling operation, where the tangent vector $r'(t_0)$ in $T_{r(t_0)}N$ is parallelly transported into $T_{r(0)}N$ as $\mathrm{Par}_{r(t_0)}^{r(0)} r'(t_0)$ that is identical to the tangent vector $\acute{r}'(t_0)$.

The unwrapping of a point $[Q_1] \in N$ at a time point t_1 with respect to the curve r is a vector on $T_{r(0)}N$ defined as

$$[\acute{Q}_1] = \acute{r}(t_1) + \mathrm{Par}_{r(t_1)}^{r(0)}\big(\mathrm{Log}_{r(t_1)}([Q_1])\big).$$

Illustrated in the right panel of Figure 2, the tangent vector $\mathrm{Log}_{r(t_1)}([Q_1])$ in $T_{r(t_1)}N$ is parallelly transported into $T_{r(0)}N$ as $\mathrm{Par}_{r(t_1)}^{r(0)}\big(\mathrm{Log}_{r(t_1)}([Q_1])\big)$ that is identical to $[\acute{Q}_1] - \acute{r}(t_1)$. In addition, the unwrapping of a line $[P(t)] \in N$ along $r(t)$ is a curve on $T_{r(0)}N$ defined pointwisely by

$$[\acute{P}(t)] = \acute{r}(t) + \mathrm{Par}_{r(t)}^{r(0)}\big(\mathrm{Log}_{r(t)}([P(t)])\big).$$

Conversely, the wrapping of a line $\acute{\alpha}(t)$ on $T_{r(0)}N$ into N is the inverse of unwrapping curve $\alpha(t)$ such that

$$\alpha(t) = \mathrm{Exp}_{r(t)}\big\{\mathrm{Par}_{r(0)}^{r(t)}\big(\acute{\alpha}(t) - \acute{r}(t)\big)\big\}.$$

The regular base curve r links the manifold N with the tangent space $T_{r(0)}N$, i.e., all observations $[Q_j]$ are transformed and transported along it to the space $T_{r(0)}N$.

3.2. Estimation of eigen-frame

With the help of the unrolling and unwrapping techniques, the points $\{[Q_j]\}_{1\le j\le m}$ on N are unwrapped into the tangent space $T_{r(0)}N$ along a regular base curve r, denoted by $\{[\acute{Q}_j]\}_{1\le j\le m}$. Now we aim to map $\{[\acute{Q}_j]\}_{1\le j\le m}$ to the tangent space $T_I M$, the space of anti-symmetric matrices, and eliminate the influence of the regular base curve selection.

Let γ be a lifting of r into M satisfying $\phi(\gamma(t)) = r(t)$, where ϕ refers to the covering map from M to N that is detailed in Lemma 2. The choice of the lifting curve γ is not unique but it does not affect the

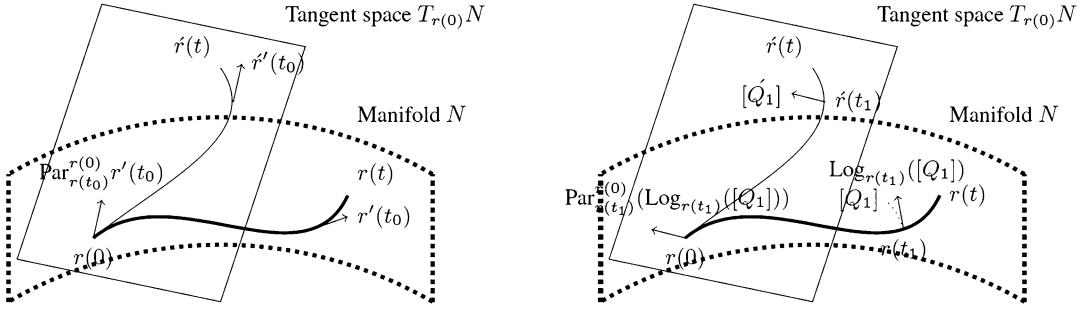


Figure 2. A diagrammatic sketch of unrolling operation (left) and unwrapping operation (right). The dotted bending parallelogram and the solid parallelogram represent the manifold N and the tangent space $T_{r(0)}N$, respectively. The thick curve is $r(t)$ and the thin curve is the unrolling of $r(t)$, named $\acute{r}(t)$.

smoothing spline result on N (Lemma 10). Next, the map $(R_{\gamma(0)})^{-1}$ denoting the inverse map of right multiplication is also necessary in our estimation process, where $(R_{\gamma(0)})^{-1}u = u(\gamma_0)^{-1}$ for any tangent vector $u \in T_{\gamma(0)}M$. With the aid of the following two isometric mappings

$$(d\phi_{\gamma(0)})^{-1} : T_{r(0)}N \to T_{\gamma(0)}M, \qquad (R_{\gamma(0)})^{-1} : T_{\gamma(0)}M \to T_I M,$$

$\{[\acute{Q}_j]\}_{1\le j\le m}$ lying on $T_{r(0)}N$ can be isometrically moved into $T_I M$ by the map $R^{-1}_{\gamma(0)} \circ (d\phi_{\gamma(0)})^{-1}$. These mapped points on $T_I M$ are denoted by $\tilde{Q}_j = R^{-1}_{\gamma(0)} \circ d\phi^{-1}_{\gamma(0)}([\acute{Q}_j])$. We then 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 $\{(t_j, \tilde{Q}_j^{kl})\}_{j=1,\dots,m}$ 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 as follows:

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

where $W^{2,2}([0,1]) = \{g : [0,1] \to \mathbb{R} \mid g, g^{(1)}, g^{(2)} \in L^2([0,1])\}$ is the twice order Sobolev–Hilbert space, and $\rho > 0$ is a tuning parameter representing a trade-off between the fidelity to the data and the smoothness of the estimated function. The computation of the one-dimensional smoothing spline estimates $\{\tilde{\alpha}^{kl}(t)\}_{1\le k,l\le d}$ can be found in Cai and Yuan (2011) and the tuning parameter $\rho > 0$ is selected by K-fold cross validation.

The anti-symmetry of $\{\tilde{Q}_j\}_{1\le j\le m}$ suggests that $\tilde{Q}_j^{kl} = -\tilde{Q}_j^{lk}$ and thus $\tilde{\alpha}^{lk}(t) = -\tilde{\alpha}^{kl}(t)$ for all $t \in [0,1]$. Assembling the entries $\tilde{\alpha}^{kl}$ together, we obtain an curve $\tilde{\alpha}$ that lies on $T_I M$. Finally, we move $\tilde{\alpha}$ back to $T_{r(0)}N$ by $\phi \circ R_{\gamma(0)}$ to get a curve $\acute{\alpha}$ on $T_{r(0)}N$ and then wrap $\acute{\alpha}$ along the regular base curve r to obtain α, which is the one-step unrolling estimator of the dynamic eigen-frame $[P(t)]$. Figure 3 summarizes the relations among quantities that arise from our estimation procedures.

The entry-wise smoothing spline estimator $\tilde{\alpha}$ resides entirely in the vector space $T_I M$, and the resulting final estimator α is guaranteed to be confined to the manifold N. This is also the reason why we go through this long journey via unrolling and unwrapping techniques to move all rough estimates $\{[Q_j]\}_{1\le j\le m}$ into a common linear space $T_I M$. In contrast, if the algorithm is directly applied to each entry for the data $\{[Q_j]\}_{1\le j\le m}$ on the non-linear manifold N, the estimator α may go out of that manifold and becomes invalid.

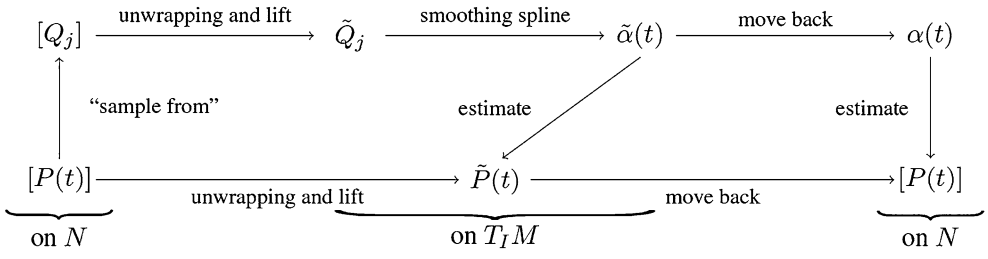


Figure 3. The relationships of important quantities in the estimation.

3.3. Simplified approach

We propose a simplified and practical way to compute the estimate $\alpha(t)$ since the "unwrapping and lift" procedures of $\{[Q_j]\}_{1\le j\le m}$ in Figure 3 require complicated computation on equivalence classes shown as the complex way in Figure 4. We turn to choose new representatives denoted by $\{\mathcal{P}_j\}_{1\le j\le m}$ for the rough estimates $\{[Q_j]\}_{1\le j\le m}$ in the computation procedure. $\mathcal{P}_1$ can be chosen arbitrarily in the equivalence class of $[Q_1]$ and $\mathcal{P}_j$ is selected recursively for $2 \le j \le m$

$$\mathcal{P}_j = \arg\min_{P\in[Q_j]} d_M(\mathcal{P}_{j-1}, P). \tag{2}$$

Therefore, the data $\{\widehat{P(t_j)}\}_{1\le j\le m}$ and $\{[Q_j]\}_{1\le j\le m}$ have new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$ that satisfy $d_N([Q_j],[Q_{j+1}]) = d_M(\mathcal{P}_j,\mathcal{P}_{j+1})$ (Lemma 8). The new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$ are viewed as the isometric copies of the rough estimates $\{[Q_j]\}_{1\le j\le m}$ and can replace the rough estimates in the computation process, playing an important role in the simplified way in Figure 4 explained as follows.

In this simplified approach, the regular base curve Γ on M is obtained based on the new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$ according to Section B. We address that $\gamma \subset M$ introduced in Section 3.2 is an arbitrary lift of r while $\Gamma \subset M$ is obtained from the new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$, which eventually depends on $\mathcal{P}_1$. Proposition 1 suggests that they are both the lifts from $r \subset N$ onto M, meaning $\phi(\Gamma) = \phi(\gamma) = r$. Based on the regular base curve $\Gamma \subset M$, the techniques of unrolling and unwrapping on M are adapted to unwrap the new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$ to the tangent space $T_{\Gamma(0)}M$ along Γ to obtain $\{\acute{\mathcal{P}}_j\}_{1\le j\le m}$. Followed by $R^{-1}_{\Gamma(0)}$, $\{\acute{\mathcal{P}}_j\}_{1\le j\le m}$ are mapped into $T_I M$ to obtain $\{R^{-1}_{\Gamma(0)}\acute{\mathcal{P}}_j\}_{1\le j\le m}$ that are exactly $\{\tilde{Q}_j\}_{1\le j\le m}$ in the complex way, as guaranteed in Proposition 1.

Proposition 1. *Following the notations in the above description, the regular base curve* Γ *satisfies* $\phi(\Gamma) = r$. *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 equivalent. Therefore, these two sets of estimates agree with each other.*

In the step of unrolling and unwrapping techniques, the computation of the parallel transport on M is thoroughly interpreted by Lemma 5 through ordinary differential equation whose numerical solutions can be computed by standard discretization method (Shampine, Gladwell and Thompson, 2003). This gives polynomial time complexity on m, and we successfully avoid including N in the unwrapping and lift computation procedures and reduce the run time. We present a concise summary of the practical procedures (i.e., the simplified way) for computing the estimate $\alpha(t)$ in Algorithm 1, while a more comprehensive version can be found in Supplementary Material (Shao and Yao, 2024).

Despite of the simplified way, the quotient space N is necessary in our framework. It is the set of equivalence classes and thus acts as the base stone to select new representatives for each $[Q_j]$. Besides, the quotient space N provides a Riemannian manifold to describe the one-step unrolling estimate

Input: The observations $\{(t_j, X_i(t_j))\}_{1\le i\le n, 1\le j\le m}$.
Output: The estimate $\alpha(t)$.
– (1) Obtain rough estimates $\{[Q(t_j)]\}_{1\le j\le m}$ by matrix eigen-decomposition from input;
– (2) Select new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$ according to (2);
– (3) Obtain $\{\tilde{Q}_j\}_{1\le j\le m}$ on $T_I M$ following the simplified way in Figure 4;
– (4) Compute $\tilde{\alpha}(t)$ entrywisely from $\{\tilde{Q}_j\}_{1\le j\le m}$ according to (1);
– (5) Move back $\tilde{\alpha}(t)$ to obtain $\alpha(t)$ on N.

Algorithm 1: Algorithm of the simplified way to compute the estimate α.

in a mathematically simple and reader-friendly way, possessing vital importance to constructing the convergence property of the proposed estimate in Theorem 2. It is worth noting that, while the proposed one-step unrolling method addresses the curve-fitting problem only on the orthogonal matrix group M (or its quotient space N), this method can be extended to a general Lie group, as illustrated in Supplementary Material (Shao and Yao, 2024).

3.4. Selection of principal components

In fact, the new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$ determine an ordering of the eigenvectors on each observed time points $\{t_j\}_{1\le j\le m}$, but may not be the ordering of L^p norms of the corresponding eigenvalues. Assuming the lth column of $\mathcal{P}_j$ is represented by $\mathcal{E}_{l,j}$, for each l, the set $\{\mathcal{E}_{l,j}\}_{1\le j\le m}$ comprises estimates of one dynamic eigenvector within the collection $\{e_1,\ldots,e_d\}$ at the observed time points $\{t_j\}_{1\le j\le m}$. Therefore, the smooth spline interpolation $\mathcal{E}_l$, based on $\{\mathcal{E}_{l,j}\}_{1\le j\le m}$, is the smoothed estimate for that specific dynamic eigenvector. Therefore, the estimates of eigenvalues $\lambda_{\mathcal{E}_l}$ corresponding to $\mathcal{E}_l$ on the observed time points $\{t_j\}_{1\le j\le m}$ are defined as

$$\widehat{\lambda_{\mathcal{E}_l}(t_j)} = (\mathcal{P}_j^T \widehat{\Sigma}(t_j)\mathcal{P}_j)_{ll},$$

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. Subsequently, the L^p norm of the eigenvalue function $\lambda_{\mathcal{E}_l}$ is estimated by the Riemannian sum

$$R_l = \left(\sum_{j=2}^m |\widehat{\lambda_{\mathcal{E}_l}(t_j)}|^p (t_j - t_{j-1})\right)^{1/p}.$$

Then dynamic components are selected by the magnitude ordering of the estimated L^p norms. Specifically, the estimated L^p norms $\{R_l\}_{l=1}^d$ are arranged in descending order such that $R_{l_1} \ge \cdots \ge R_{l_d}$. Here, the subscript l_k denotes that R_{l_k} represents the kth largest estimated L^p norm within the set $\{R_l\}_{l=1}^d$. Consequently, the kth global principal component e_k is estimated by $\hat{e}_k = \mathcal{E}_{l_k}$.

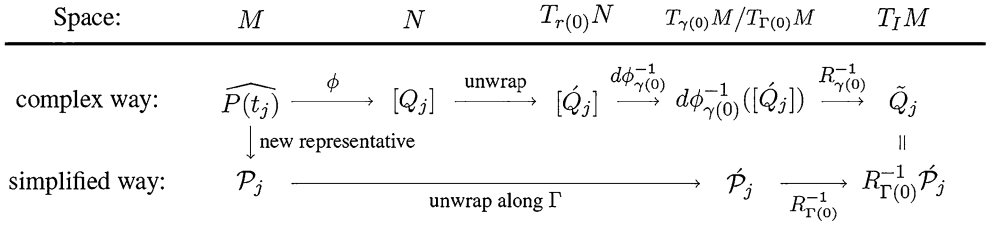


Figure 4. The graphic illustration of the complex v.s. simplified way of lifting the rough estimates to $T_I M$.

4. Theoretical properties

We start with the following assumption on the underlying population X and the eigenvalues $\{\lambda_k(t_j)\}$. The single multiplicity requirement that guarantees the eigenvectors to be uniquely identified is a common condition in the field of principal component analysis (Hsing and Eubank, 2015, Pearson, 1901).

Assumption 1. The population X has finite fourth moment $\sup_{t\in[0,1]} E\,\|X(t)\|^4 < \infty$ and the time-varying eigenvectors are twice differentiable $e_k \in C^2[0,1]$ for $1 \le k \le d$. The variance function Σ has distinct eigenvalues $\kappa = \min_{1\le j\le m, k\neq l} |\lambda_k(t_j) - \lambda_l(t_j)| > 0$ on the observed time points $\{t_j\}_{1\le j\le d}$.

Our first result concerns the mean square errors and rates of the rough estimates $[Q_j]$ at the observed time points, which rely on the fourth moment condition and is essential for the next theorem.

Theorem 1. *If Assumption 1 holds, then the rough estimates admit*

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

With the aid of the regularity of the base curve r from Proposition 2, the convergence rate of the estimator α and the consistency of dynamic principal components selection are established as follows. This convergence rate shares the minimax rate of the smoothing spline estimator under the regular design points in Euclidean case (Cai and Yuan, 2011). This suggests that, with our suitably chosen regular base curve, the proposed estimator enjoys the minimax convergence rate even the data $\{[Q_j]\}_{1\le j\le m}$ are located in the manifold N. This rate exhibits a phase transition at $m \asymp n^{1/4}$. While below this threshold, the sampling frequency m determines the convergence rate of α. Otherwise, the sample size n dominates.

Theorem 2. *If $e_1(\cdot),\ldots,e_d(\cdot) \in C^2[0,1]$ and Assumptions 1 and 2 hold, then $\alpha(t)$ converges to $[P(t)]$ on manifold N in the L^2 norm sense such that*

$$\int_0^1 d_N^2\big([P(t)],\alpha(t)\big)\,dt = O_p(m^{-4}+n^{-1})$$

with the choice of tuning parameter $\rho \asymp m^{-4}+n^{-1}$. Following the notations in Section 3.4, we have $\lim_{m\to\infty} R_l = \|\lambda_{\mathcal{E}_l}\|_{L^p}$ for every $1 \le l \le d$. The estimated dynamic principal component is of the same rate for every $1 \le k \le d$

$$\int_0^1 \|\hat{e}_k(t) - e_k(t)\|^2 dt = O_p(m^{-4}+n^{-1}).$$

We address that the proof of Theorem 2 in Supplementary Material (Shao and Yao, 2024) is nontrivial. The difficulty stems from the analysis of the object $\tilde{P}(t) = R^{-1}_{\gamma(0)} \circ d\phi^{-1}_{\gamma(0)}([\acute{P}(t)])$ on $T_I M$ that is the transformation of the eigen-frame population $P(t)$ and acts as an important medium in the proof. To see this, $\tilde{P}(t)$ depends on not only the random regular base curve r but also the parallel transport that is characterized by Lemma 5 in terms of matrix ordinary differential equations. The entanglement between the randomness of r and lack of an explicit form of the parallel transport substantially complicates our analysis and poses technical challenges.

This is the first manifold-valued smoothing spline method with theoretical convergence result guarantee to the best of our knowledge. As a matter of fact, our smoothing method, with suitable construction of the regular base curve, avoids the iterative procedure (with less computation and smaller estimation error) that facilitates the theoretical analysis, and thus can be viewed as an extension of the

unrolling by iteration method (Kim et al., 2021). The construction of the regular base curve improves the piecewise geodesic curve by choosing a subset and smoothing first derivative. By contrast, the unrolling by iteration method begins with the piecewise geodesic interpolation and does unrolling/unwrapping procedures iteratively. To be specific, the initial estimate α_0 was the piece-wise geodesic interpolation of $\{[Q_j]\}_{1\le j\le m}$ and it was used as the base curve to do unrolling and unwrapping procedures to obtain the estimate α_1 that was again used as the base curve. This operation was repeated to obtain a series of estimates $\alpha_1, \alpha_2, \alpha_3, \ldots$ until convergence. Since the regular base curve obtained by our construction is smooth (relative to the piecewise geodesic interpolation), we only need one step of unrolling and unwrapping procedure to obtain a well-performing estimator with theoretical guarantee, thus named as one-step unrolling.

If the dynamic eigenvectors $e_1(\cdot), \ldots, e_d(\cdot)$ possess higher order smoothness, e.g., belong to $C^p[0,1]$, we may slightly modify the estimation procedure by substituting $\int_0^1 |f''(s)|^2 ds$ with $\int_0^1 |f^{(p)}(s)|^2 ds$ in (1). We shall also replace the twice differentiable regular base curve with a pth differentiable one, which can be constructed by recursively applying the method described in section B. Then the rate in Theorem 2 can be shown $O_p(m^{-2p} + n^{-1})$ by slightly modifying the proof. This also coincides with the result of classic higher order smoothing splines (Cai and Yuan, 2011).

5. Simulation studies

5.1. Assessment of estimation and convergence

We assess the numerical performance of the proposed one-step unrolling method and compare it with other existing smoothing methods on manifolds. Consider the case $d = 3$ for illustration and generate data according to $X_i(t_j) = \sum_{k=1}^3 \psi_i^k(t_j) e_k(t_j)$, where $\{\psi_i^k(t_j)\}_i$ are independent and identically distributed as Normal$(0, \lambda_k(t_j))$, and the observed time points $\{t_j\}_{1\le j\le m}$ are generated equidistantly. The non-Gaussian scores are studied in Supplementary Material (Shao and Yao, 2024) for space economy. The eigenvalues $\{\lambda_k(t)\}_{k=1}^d$ are discontinuous and swap over time in our setting $\lambda_k(t) = k^2 I_{\{0\le t<0.5\}} + (d+1-k)^2 I_{\{0.5<t\le 1\}}$. The dynamic eigen-frame $\{e_k(t)\}_{1\le k\le d}$ is parameterized by the three-dimensional polar coordinates

$$\begin{aligned} e_1(t) &= \big(\cos(\theta), \sin(\theta)\sin(\phi), \sin(\theta)\cos(\phi)\big), \\ e_2(t) &= \big(-\sin(\theta), \cos(\theta)\sin(\phi), \cos(\theta)\cos(\phi)\big), \\ e_3(t) &= \big(0, \cos(\phi), -\sin(\phi)\big), \end{aligned} \tag{3}$$

where $\theta = \theta(t) = \sin(2\pi t)$ and $\phi = \phi(t) = \sin(\pi t)/2$.

To examine the convergence in Theorem 2 in terms of m and n, we use the mean square error of the smooth estimator α, given by $\int_0^1 E\, d_N^2(\alpha(s), [P(s)])ds$ assessed by Monte Carlo average over 100 runs. The tuning parameters are selected by 5-fold cross validation. The results of mean square errors of our proposed one-step unrolling method presented in the upper zone of Table 1 lead to the following observations. (i) Mean square errors decreases as m and/or n increases. (ii) The run time increases as m goes large but remains stable when n varies. This is expected since the sample size n only affects the time for computing the rough estimates $\{[Q_j]\}_{\{1\le j\le m\}}$, which is relatively small and negligible to the total run time, while the sampling frequency m determines most of the run time in the unrolling and smoothing spline parts.

We also compare the proposed one-step unrolling with two other methods, unrolling by iteration method (Kim et al., 2021) and the gradient-descent method (Samir et al., 2012), the results of which

Table 1. The Monte Carlo mean, standard deviation of the mean square errors (multiplied by 10^4) and run time of our proposed one-step unrolling method, unrolling by iteration method and gradient-descent method

one-step unrolling	$m = 60$		$m = 80$		$m = 100$	
$n = 100$	46.3 (23.2)	21.4 s	31.9 (13.4)	32.6s	27.5 (13.0)	59.1 s
$n = 300$	15.4 (5.47)	24.2 s	12.2 (4.80)	39.9s	9.41 (3.46)	55.8 s
$n = 600$	7.80 (3.08)	21.8 s	6.35 (2.61)	37.7s	5.00 (1.96)	58.6 s
$n = 1000$	5.09 (1.84)	24.3 s	4.45 (1.59)	38.4s	3.12 (1.15)	59.0 s
unrolling by iteration	$m = 60$		$m = 80$		$m = 100$	
$n = 100$	44.2 (19.7)	44.4 s	34.2 (14.3)	78.7 s	29.1 (14.7)	165.4 s
$n = 300$	16.0 (6.63)	41.2 s	13.8 (4.96)	77.9 s	10.7 (4.45)	135.4 s
$n = 600$	8.24 (3.27)	40.0 s	6.50 (2.42)	76.8 s	5.17 (2.20)	133.0 s
$n = 1000$	5.45 (1.95)	38.6 s	4.39 (1.49)	76.1 s	3.33 (1.12)	129.9 s
gradient-descent	$m = 60$		$m = 80$		$m = 100$	
$n = 100$	434 (68.5)	798.6 s	439 (81.2)	1990.5 s	431 (45.4)	4211.4 s
$n = 300$	156 (28.4)	628.3 s	139 (16.2)	1366.1 s	138 (15.6)	2414.4 s
$n = 600$	68.0 (10.6)	507.4 s	69.5 (8.57)	1119.1 s	67.5 (7.10)	1587.7 s
$n = 1000$	40.6 (8.31)	278.7 s	40.8 (6.08)	585.2 s	41.0 (4.81)	988.4 s

are shown in the middle and lower zones of Table 1, respectively. The tuning parameters of these two methods are also chosen by 5-fold cross validation, and the iterations stop when the relative mean square error is less than 0.1%. From the results in Table 1, we see that mean square errors of our method are smaller than those of the other two methods in most cases. The run time of the gradient-descent method is substantially longer due to its complex computation procedures under Palais metric and the nature of the gradient descent iteration to reach the extremal point. The run time of the unrolling by iteration method is also longer than that of our method, especially when n is small and m is large as expected due to the following reasons. First, when the sample size n decreases, the rough estimates $\{[Q_j]\}_{\{1\le j\le m\}}$ are further away from the underlying function, so is the initial curve α_0 constructed by the piece-wise geodesic interpolation in the unrolling by iteration method, thus taking more iterations to converge. Second, when the sampling frequency m increases, the initial curve α_0 becomes more serrated and thus also needs more iterations to get smoother. By comparison, the construction of our regular base curve only requires a subset of all observed time points such that the smoothness and regularity can be guaranteed according to Proposition 2, and subsequently avoids the poor performance of the initial curve α_0 in the unrolling by iteration method.

5.2. Illustration of eigenvalue patterns

Our proposed method is not affected by the pointwise ordering of eigenvalue functions $\{\lambda_k(t)\}_{k=1}^d$ and allows the dynamic eigenvalues to change discontinuously or switch over time. This feature can not be achieved by smoothing time-varying covariance matrices, referring to estimating the covariance matrix function across time by interpolating the sample covariance matrices at the observed time points via the techniques from Chen and Li (2016), Groisser, Jung and Schwartzman (2017), Yuan et al. (2012), Zhu et al. (2009) and then performing the standard matrix eigen-decomposition subsequently on each time point $t \in [0,1]$, since it inevitably incorporates the eigenvalues and relies on the smoothness of them.

Consider the case $n = 1000$, $m = 100$ and $d = 2$ for simplicity. Generate the data according to $X_i(t_j) = \sum_{k=1}^2 \psi_i^k(t_j)e_k(t_j)$, where $\{\psi_i^k(t_j)\}_i$ are independent and identically distributed as Normal$(0, \lambda_k(t_j))$ and the observed time points $\{t_j\}_{1\le j\le m}$ are generated equidistantly. In the switched case, set the eigenvalues

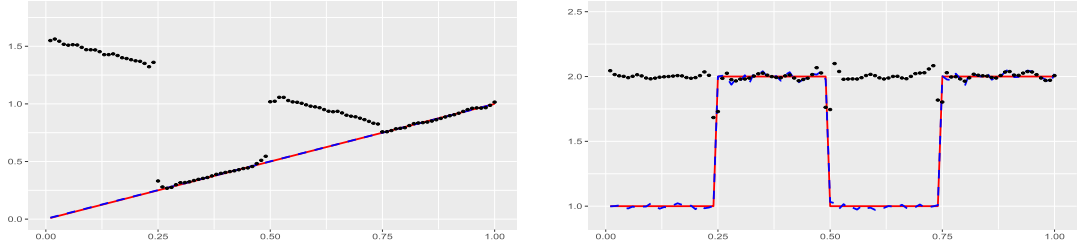


Figure 5. The dynamic eigenvector estimates (parameterized by polar coordinate θ) are shown on the left while the first eigenvalue estimates $\hat{\lambda}_1$ are drawn on the right. The dashed, dotted and solid lines refer to the estimates from our method, the smoothing time-varying covariance method and true values, respectively.

λ_1, λ_2 as

$$\lambda_1(t) = I_{\{0\le t<0.25\}} + 2I_{\{0.25\le t<0.5\}} + I_{\{0.5\le t<0.75\}} + 2I_{\{0.75\le t\le 1\}};$$
$$\lambda_2(t) = 2I_{\{0\le t<0.25\}} + I_{\{0.25\le t<0.5\}} + 2I_{\{0.5\le t<0.75\}} + I_{\{0.75\le t\le 1\}}.$$

The two dimensional special orthogonal matrix group $SO(2)$ is a one-parameter group that can be parameterized by angle θ under the polar coordinate

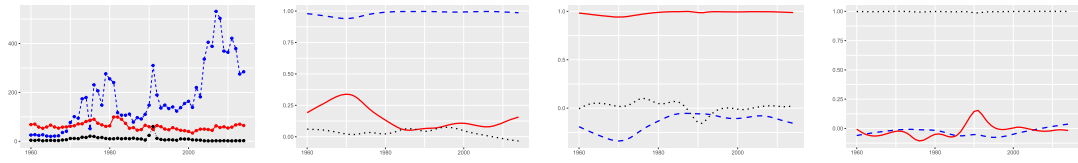
$$S(\theta) = \begin{Bmatrix} \cos(\theta), & -\sin(\theta) \\ \sin(\theta), & \cos(\theta) \end{Bmatrix} \in SO(2).$$

The true value of dynamic eigen-frame $P(t) = \{e_1(t), e_2(t)\}$ is a curve on $SO(2) \subset O(2)$ described by the polar coordinate $\theta_{\text{true}}(t) = t$. Both our method and the smoothing time-varying covariance method are adopted to estimate the eigen-frame $\theta_{\text{true}}(t)$ and eigenvalue $\lambda_1(t)$, with tuning parameters selected by 5-fold cross validation. The panels of Figure 5 suggest that our method successfully recovers the true values of dynamic eigenvector and eigenvalue while the smoothing time-varying covariance method fails over the set $[0, 0.25] \cup [0.5, 0.75]$ that is exactly the switched set $\{t : \lambda_1(t) < \lambda_2(t)\}$.

6. Data application

We apply the proposed dynamic principal component analysis method to study the government expenditures based on the data from World Bank OECD National Accounts, available at https://data.worldbank.org/. The data set considered includes 208 countries and locations from 1960 to 2014 and consists of three major aspects as proportions to total GDP: net import, gross capital formation and military expenditures. Specifically, the net import is the difference between imports and exports of goods and services. Imports (exports) of goods and services represent the value of all goods and other market services received from (provided to) the rest of the world. Gross capital formation, formerly called gross domestic investment, consists of outlays in addition to the fixed assets of the economy plus net changes in the level of inventories. The military expenditures include all current and capital expenditures on the armed forces, including peacekeeping forces, defense ministries, paramilitary forces, and military space activities.

The proposed one-step unrolling method is applied to analyze this data with dimension $d = 3$, sample size $n = 208$ and sampling frequency $m = 55$. The tangent space $T_I SO_3$ is also three-dimensional and thus three operations of one-dimensional smoothing spline are conducted, where the tuning parameters



$\hat{\lambda}_1, \hat{\lambda}_2$ and $\hat{\lambda}_3$. coefficients of $\hat{e}_1$ coefficients of $\hat{e}_2$ coefficients of $\hat{e}_3$

Figure 6. $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$. The left panel presents the estimates of three eigenvalue patterns. The rest panels plot the estimates of three dynamic eigenvectors $\hat{e}_1(t), \hat{e}_2(t), \hat{e}_3(t)$. The three dynamic coefficients for each $\hat{e}_k(t)$ are the weights on (1) net import, (2) gross capital formation and (3) military expenditures plotted in the shape of dashed, solid and dotted lines, respectively.

$\rho = 5.9 \times 10^{-4}, 4.4 \times 10^{-3}, 2.9 \times 10^{-5}$ are selected by 5-fold cross validation. The left panel of Figure 6 presents the estimated three eigenvalue dynamics $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$ ordered by their L^2 norms; other L^p criterion also leads to the same result. It reveals eigenvalue switch phenomena at three years (1970, 1975, 1991), and that the pointwise magnitude order does not always match the global order. The estimates of the corresponding dynamic eigenvectors $\hat{e}_1(t), \hat{e}_2(t), \hat{e}_3(t)$ are plotted in the rest panels of Figure 6. The results of the three coefficients for each dynamic eigenvector show that $\hat{e}_1(t)$ majorly depends on net import while $\hat{e}_2(t)$ on gross capital formation and $\hat{e}_3(t)$ on military expenditures.

It is interesting to see that the eigenvalue function $\hat{\lambda}_2(t)$ corresponding to gross capital formation is fairly stable while $\hat{\lambda}_1(t)$ corresponding to net import exceeds $\hat{\lambda}_2(t)$ from 1970 and remains larger except for 1975. The $\hat{\lambda}_3(t)$ corresponding to military expenditures is always the smallest except for 1991 when it crossed that of gross capital formation. The switch of eigenvalue dynamics can be interpreted by the changes of variations of these three aspects as follows. First, the rise of $\hat{\lambda}_1$ indicates the growth trend of the variation in net import, which might result from the development of economic globalization in past decades when the economies of most countries and locations become more import-oriented or export-oriented. Second, the track of $\hat{\lambda}_3$ may be explained by the military wars. For example, the jump at year 1991 seems due to the Kuwait War, e.g., inspecting that the military expenditures as proportions to total GDP of Kuwait gives 48.52%, 117.35% and 31.79% in 1990, 1991 and 1992 that lifted the variance of military expenditures. Historical periods including Vietnam war (before/around 1975) and Iraq war (before/around 2003) also saw an increase in the variation of military expenditures.

Appendix A: Geometry of quotient space N

To appreciate the geometry of N defined as a quotient space of M, we start with a brief introduction to the geometry of M. Recall that the space of $d \times d$ matrices can be identified with the space $\mathbb{R}^{d\times d}$, endowed with the Hilbert–Schmidt (or Frobenius) inner product

$$\langle A, B\rangle = \sum_{k,l=1}^{d} a_{kl} b_{kl} = \mathrm{tr}(B^T A), \quad \text{for } A = (a_{kl})_{d\times d} \text{ and } B = (b_{kl})_{d\times d}.$$

The space $\mathbb{R}^{d\times d}$ is also viewed as a Riemannian manifold, whose tangent space at any point is identified with $\mathbb{R}^{d\times d}$ itself and endowed with the above inner product. The space M is then treated as a sub-manifold of $\mathbb{R}^{d\times d}$ whose Riemannian metric is inherited from $\mathbb{R}^{d\times d}$. Moreover, with the matrix multiplication, M is indeed a Lie group. Below we discuss various geometric and group properties of M that are relevant to this paper, while referring readers to Gilmore (2012) for more details of this manifold. For a comprehensive treatment of Riemannian manifolds and Lie groups, we recommend the textbooks (do Carmo, 1992, Knapp, 2013).

To reduce notional burden, we write I_d by I. The tangent space of M at I is $T_I M = \{S \in \mathbb{R}^{d\times d} \mid S + S^T = 0\}$, the space of $d \times d$ anti-symmetric matrices. The tangent space of M at $P \in M$ is characterized by $T_P M = \{PS : S \in T_I M\}$, or equivalently, $T_P M = \{S \in \mathbb{R}^{d\times d} : SP^T + PS^T = 0\}$. For any $G \in M$, the left multiplication operator L_G and the right multiplication operator R_G are defined as $L_G P = GP$ and $R_G P = PG$, respectively. These operators L_G, R_G preserve the metric in the sense of Lemma 4. In other words, the Riemannian metric on it is bi-invariant in light of this Lemma.

As M is compact, it is geodesically complete. The geodesic curve is characterized in Lemma 6, while the distance function d_M can by computed according to Lemma 7. In addition, the Riemannian exponential map Exp_P is well defined on all of $T_P M$ (do Carmo, 1992, Hopf–Rinow Theorem, p. 150), which is $\mathrm{Exp}_P(PS) = P\exp(S)$ for $S \in T_I M$. Exp_P is not injective but locally bijective. Its local reverse $\mathrm{Log}_I = (\mathrm{Exp}_I)^{-1}$ is well defined on $\{P \in M : \|I - P\|_{\mathrm{HS}} < 1\}$.

Since M is a Riemannian sub-manifold of $\mathbb{R}^{d\times d}$, the Levi–Civita connection on M is simply the covariant derivative. To be specific, suppose that $\gamma(t)$ is a smooth curve on M and $X(t)$ is a vector field along $\gamma(t)$, then the Levi–Civita connection on M is given by

$$\nabla_{\gamma'(t)} X(t) = \mathrm{Proj}_{T_{\gamma(t)}M} X'(t)$$

where $\mathrm{Proj}_{T_{\gamma(t)}M}$ means the projection into the subspace $T_{\gamma(t)}M$. The parallel transport associated with the connection can be computed by Lemma 5.

Now we turn to the group $\mathcal{G}$ and the quotient space N. There are three equivalent ways to characterize N. (1) For $P_1, P_2 \in M$, if there exists $G \in \mathcal{G}$ such that $P_1 = P_2 G$, then P_1 and P_2 are in the same equivalence class of N. (2) As $\mathcal{G}$ is a subgroup of M, $[P] = P\mathcal{G}$ is a left-coset. All the left-cosets $\{[P] : P \in M\}$ form the space N. (3) The right-action of $\mathcal{G}$ (i.e., for any $G \in \mathcal{G}$, $R_G : M \to M$) is in fact the signed-permutation group action, which is discrete and induces a covering map $\phi : M \to N$ (do Carmo, 1992).

The tangent vector $[u] = u\mathcal{G} \in T_{[P]}N$ (where $u \in T_P M$) is an equivalence class of vectors $\{uG \in T_{PG}M : G \in \mathcal{G}\}$ and the vector addition in $T_{[P]}N$ is defined as $u_1\mathcal{G} + u_2\mathcal{G} = (u_1 + u_2)\mathcal{G}$. The metric on N can be induced by the covering map ϕ. Specifically, for any $[P] \in N$, there exists a collection of isomorphisms $\{\phi_{PG}\}_{G\in\mathcal{G}}$, $\phi_{PG} : U_{PG} \to U_{[P]}$, where U_{PG} is the neighborhood of $PG \in M$ and $U_{[P]}$ is the neighborhood of $[P] \in N$. The metric on $U_{[P]}$ induced by ϕ_{PG} is

$$\langle [u], [v] \rangle = \langle (d\phi_{PG})^{-1}[u], (d\phi_{PG})^{-1}[v] \rangle = \langle uG, vG \rangle = \langle u, v \rangle$$

where $[u], [v] \in T_{[P]}N, G \in \mathcal{G}$ and $d\phi_{PG}$ denotes the differential of ϕ_{PG}. The above definition of the metric on N does not rely on the choice of G.

Endowed with this metric, the manifold N is geodesically complete (do Carmo, 1992, p. 146). For any $[P_1], [P_2] \in N$, their distance is computed by $d_N([P_1], [P_2]) = \min_{G_1, G_2 \in \mathcal{G}} d_M(P_1 G_1, P_2 G_2)$ (Lemma 8). In addition, the left multiplication $L_P : N \to N$ defined by $L_P[Q] = [PQ]$ is a Riemannian isometric on N for any $P \in M$. Moreover, the Riemannian exponential map on N is locally bijective and induces a diffeomorphism. Therefore, the Riemannian logarithmic map at $[P] \in N$ is locally well defined. In addition, the distance between $\mathrm{Log}_{[P]}(a)$ and $\mathrm{Log}_{[P]}(b)$ is bounded by the distance between a and b on N, as asserted by the following Lemma.

Lemma 1. *There exist constants $R < 1, C_1, C_2$ relying only on the dimension d such that for any $[P] \in N$ and $a, b \in B_N([P]; R)$, the Riemannian logarithmic map $Log_{[P]}(\cdot)$ is well defined within $B_N([P]; R)$ and the following inequalities hold*

$$C_1 d_N(a, b) \le \|Log_{[P]}(a) - Log_{[P]}(b)\|_{HS} \le C_2 d_N(a, b),$$

abbreviated by $d_N(a, b) \asymp \|Log_{[P]}(a) - Log_{[P]}(b)\|_{HS}$.

Proof of Lemma 1. First, choose $R_2 < \frac{1}{2}R_1$, where R_1 is defined in Lemma 9. For any $P \in M$, the two isometric bijections

$$\phi_P : B_M(P;R_2) \to B_N([P];R_2) \quad \text{and} \quad L_P : B_M(I;R_2) \to B_M(P;R_2)$$

suggest that

$$\begin{aligned} d_N(a,b) =& d_M(\phi_P^{-1}(a),\phi_P^{-1}(b)) = d_M(L_P^{-1}\phi_P^{-1}(a),L_P^{-1}\phi_P^{-1}(b)),\\ \mathrm{Log}_{[P]}(a) =& \mathrm{Log}_P(\phi_P^{-1}(a)) = L_P \log(L_P^{-1}\phi_P^{-1}(a)),\\ \mathrm{Log}_{[P]}(b) =& \mathrm{Log}_P(\phi_P^{-1}(b)) = L_P \log(L_P^{-1}\phi_P^{-1}(b)). \end{aligned}$$

The exponential map $\log : T_I M \to M$ is a local diffeomorphism with $d(log)_I$ being the identity map. Thus, there exist constants R, C_1, C_2 $(R < R_2)$ depending only on the dimension d such that for any $\tilde{a}, \tilde{b} \in B_M(I;R)$

$$C_1 d_M(\tilde{a},\tilde{b}) \le \|\log(\tilde{a}) - \log(\tilde{b})\|_{\mathrm{HS}} \le C_2 d_M(\tilde{a},\tilde{b}).$$

Putting $\tilde{a} = L_P^{-1}\phi_P^{-1}(a)$ and $\tilde{b} = L_P^{-1}\phi_P^{-1}(b)$, the above inequalities suggest that for any $P \in M$ and $a, b \in B_N([P];R)$,

$$C_1 d_N(a,b) \le \|\mathrm{Log}_{[P]}(a) - \mathrm{Log}_{[P]}(b)\|_{\mathrm{HS}} \le C_2 d_N(a,b).$$

□

The next Lemma, called the lifting Lemma (Hatcher, 2005, p. 60), is well-known in the field of topology and is important to our development, e.g., we use it to lift all the curves on N onto M for computation.

Lemma 2. *Suppose that $\phi : M \to N$ is a covering map, $r(t)$ is a continuous curve on N, and $P \in M$ satisfies $\phi(P) = r(0)$. Then there exists a unique continuous curve $\gamma(t)$ on M such that $\gamma(0) = P$ and $\phi(\gamma(t)) = r(t)$.*

Appendix B: Construction of regular base curve

The regular base curve plays an important role in our one-step unrolling method. The regular base curve r on N relying on data $\{[Q_j]\}_{1\le j\le m} \subset N$ is the core part in the unrolling and unwrapping procedures in Sections 3.1 and 3.2. Besides, the computation algorithms in Section 3.3 also require another regular base curve Γ that relies on the new representatives $\{\mathcal{P}_j\}_{1\le j\le m}$ on M. This section introduces how to construct r on N and subsequently Γ on M with similar techniques. In the construction of r, it is expected that both the rough estimates $\{[Q_j]\}_{1\le j\le m}$ and their population counterparts $[P(t_j)]$ are close to the regular base curve, i.e., $d_N([Q_j], r(t_j))$ and $d_N([P(t)], r(t))$ are sufficiently small. An intuitive option is the piece-wise geodesic curve that links all points $\{[Q_j]\}_{1\le j\le m}$. However, this curve is not twice differentiable while such differentiability of the regular base curve is required in our theoretical analysis. Therefore, we improve the smooth methods in Jakubiak, Silva Leite and Rodrigues (2006) and Rui and Leite (2007) by using only a subset of the observed time points, allowing us to make the base curve r twice differentiable and control its Sobolev norm.

First we assume the following property of the observed time points $t_1, \ldots, t_m$ in Assumption 2 that allows us to choose a subset of observed time points to construct regular base curve. The condition that $m \to \infty$ is necessary for grid design to achieve consistent estimation, which has been thoroughly

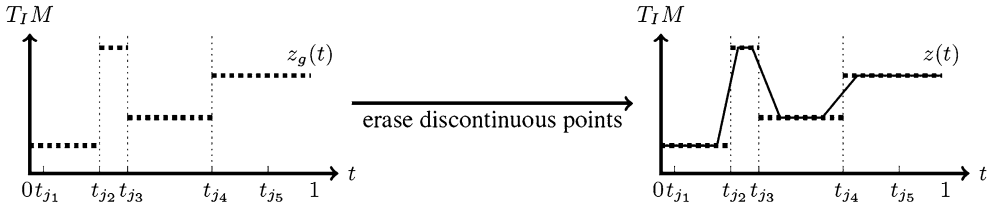


Figure 7. The thick dotted step function is $z_g(t)$ and the solid continuous line is $z(t)$.

demonstrated in Cai and Yuan (2011) for Euclidean spline. This assumption holds in most practical scenarios that the observed time points are rather evenly distributed, especially the equally spaced case.

Assumption 2. Suppose that there exists some $D = D(n) \le m$ with $\lim_{n\to\infty} D(n) = \infty$, a sub-sequence $\{t_{j_i}\}_{1\le i\le D}$ of the observed time points and constants $C_3, C_4 > 0$ such that $C_3/D \le |t_{j_{i+1}} - t_{j_i}| \le C_4/D$ and $D = O(n^{1/4})$. The points $\{t_{j_i}\}_{1\le i\le D}$ are called the interpolation points.

We start constructing the regular base curve with $r_g(t)$ being the piece-wise geodesic curve that links $\{[Q_{j_i}]\}_{1\le i\le D}$ in the interval $[t_{j_1}, t_{j_D}]$ and is extended onto $[0, t_{j_1}]$ and $[t_{j_D}, 1]$ geodesically in the boundaries. To be specific, r_g is geodesic on $[0, t_{j_2}]$ by fixing $r_g(t_{j_1}) = [Q_{j_1}]$ and $r_g(t_{j_2}) = [Q_{j_2}]$, and geodesic on $[t_{j_{D-1}}, 1]$ by fixing $r_g(t_{j_{D-1}}) = [Q_{j_{D-1}}]$ and $r_g(t_{j_D}) = [Q_{j_D}]$. Then $r_g(t)$ satisfies the differential equation $r_g'(t) = z_g(t) r_g(t)$ with

$$z_g(t) = a_1 I_{\{0\le t\le t_{j_1}\}} + \textstyle\sum_{i=1}^{D-1} a_i I_{\{t_{j_i}\le t\le t_{j_{i+1}}\}} + a_{D-1} I_{\{t_{j_D}\le t\le 1\}} \tag{4}$$

where $\{a_i\}_{1\le i\le D-1}$ are identified as elements in $T_I M$ according to Lemma 12. The curve $z_g(t)$, also viewed as the first derivative of $r_g(t)$, is a step function. We can further convert it into a continuous function $z(t)$ by smoothing out discontinuous points, as illustrated in Figure 7. 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 $h_i = t_{j_{i+1}} - t_{j_i}$. Inspired by $r_g'(t) = z_g(t) r_g(t)$, we define the regular base curve r as the solution to the following differential equation on N,

$$\frac{d}{dt} r(t) = z(t) r(t), \quad r(t_{j_1}) = [Q_{j_1}]. \tag{5}$$

The properties of the regular base curve r are important to deriving the convergence rate of our one-step unrolling estimate. The following proposition asserts that our constructed regular base curve r has uniformly bounded first and second derivatives under Hilbert-Schmidt norm with probability tending to one.

Proposition 2. *If $e_1(\cdot),\ldots,e_d(\cdot) \in C^2[0,1]$ and Assumptions 1 and 2 hold, then the regular base curve $r(t)$ defined in (5) satisfies*

$$\sup_{t\in[0,1]}(\|r'(t)\|_{HS} + \|r''(t)\|_{HS}) = O_p(1),$$

where $\|\cdot\|_{HS}$ denotes Hilbert-Schmidt norm.

The unwrapping procedures in Section 3.1 with respect to the regular base curve r depend on the Riemannian logarithmic map that is only defined locally. The following proposition shows that both rough estimates $[Q_j]$ and their population counterparts $[P(t)]$ are sufficiently close to the regular base curve $r(t)$. Therefore, $\mathrm{Log}_{r(t)}[P(t)]$ and $\mathrm{Log}_{r(t_j)}[Q_j]$ are well defined with probability tending to one.

Proposition 3. *If $e_1(\cdot),\ldots,e_d(\cdot) \in C^2[0,1]$ and Assumptions 1 and 2 hold, then*

$$\sup_{0\le t\le 1} d_N([P(t)],r(t)) + \sup_{1\le j\le m} d_N([Q_j],r(t_j)) = O_p(D^{-1}) = o_p(1).$$

The construction of the regular base curve Γ on M is based on the representatives $\{\mathcal{P}_j\}_{1\le j\le m}$ discussed in Section 3.3, utilizing similar techniques. This approach 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 $T_I M$ and corresponds precisely to the coefficients in (4), as stated in Proposition 4. Analogously, the regular base curve Γ is defined using the continuous modification $z(t)$ of $z_g(t)$, which serves as a solution to the subsequent differential equation on M

$$\frac{d}{dt}\Gamma(t) = z(t)\Gamma(t), \quad \Gamma(t_{j_1}) = \mathcal{P}_{j_1}. \tag{6}$$

In fact, the regular base curve Γ on M serves as a lift of r, that is, $\phi(\Gamma) = r$, as ensured by Proposition 4. Additionally, Proposition 4 guarantees that the first and second derivatives of Γ are uniformly bounded under the Hilbert-Schmidt norm. Furthermore, with probability tending to one, $\mathrm{Log}_{\Gamma(t_j)}\mathcal{P}_j$ is well-defined, analogous to r in Propositions 2 and 3.

Proposition 4. *As per the notations in Section B, the coefficients satisfy $z_g(t) = \tilde{z}_g(t) \in T_I M$ in the equations $r_g'(t) = z_g(t)r_g(t)$ and $\Gamma_g'(t) = \tilde{z}_g(t)\Gamma_g(t)$. Additionally, the curves Γ and Γ_g fulfill the conditions $\phi(\Gamma) = r$ and $\phi(\Gamma_g) = r_g$. Moreover, if $e_1(\cdot),\ldots,e_d(\cdot) \in C^2[0,1]$ and Assumptions 1 and 2 hold, then the regular base curve $\Gamma(t)$ in (6) satisfies*

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

where $\|\cdot\|_{HS}$ denotes Hilbert-Schmidt norm.

Appendix C: Ancillary lemmas

This section lists the ancillary lemmas that support the estimation process.

Lemma 3. *Suppose that V is a symmetric positive-definite matrix and all of its eigenvalues are distinct. Then all the orthogonal matrices that diagonalize V form exactly one element* $[P] \in N$.

Lemma 4. *For* $G, P \in M$ *and* $u, v \in T_P M$, *the tangent maps of* L_G *and* R_G

$$d(L_G)_P : T_P M \to T_{GP} M, \quad \text{and} \quad d(R_G)_P : T_P M \to T_{PG} M$$

are given by $d(L_G)_P(u) = Gu$ *and* $d(R_G)_P(u) = uG$, *respectively. In addition, they preserve the inner product, i.e.,*

$$\langle u, v \rangle = \langle Gu, Gv \rangle = \langle uG, vG \rangle = tr(v^T u).$$

Lemma 5. *Suppose that* $\gamma(t)$ *is a twice differentiable curve on* M *and* $u(0) \in T_{\gamma(0)} M$ *is fixed. The parallel transport of* $u(0)$ *along* $\gamma(t)$ *is*

$$u(t) = v_l(\tfrac{1}{2}R) u(0) \gamma(0)^{-1} v_r(-\tfrac{1}{2}R)\gamma(t),$$

where $R(t) = \gamma'(t)\gamma(t)^{-1}$, *and the functions* $v_l(A)$ *and* $v_r(A)$ *are the solutions to the following ordinary equations,*

$$\frac{d}{dt} v_l(t) = A(t) v_l(t), \quad v_l(0) = I,$$
$$\frac{d}{dt} v_r(t) = v_r(t) A(t), \quad v_r(0) = I,$$

respectively.

Lemma 6. *If* $\gamma : [a, b] \to M$ *is a geodesic curve with* $\gamma(0) = P \in M$ *and* $\gamma'(0) = PS \in T_P M$ *where* $S \in T_I M$, *then we have*

$$\gamma(t) = P \exp(tS) = \exp(tPSP^{-1})P.$$

The Riemannian exponential map at P, denoted by Exp_P, *is given by*

$$Exp_P(tPS) = P \exp(tS).$$

In addition, the length of the geodesic curve γ *from* $t = a$ *to* $t = b$ *is*

$$Len(a, b, \gamma(t)) = (b - a)\|S\|_{HS}.$$

Lemma 7. *If* $P_1, P_2 \in M$ *are in the same connected component satisfying* $P_2^T P_1 \in B(I, 1) \cap M$, *then the distance between them is*

$$d_M(P_1, P_2) = d_M(P_2^T P_1, I) = \|\log(P_2^T P_1)\|_{HS}.$$

Lemma 8. *The distance between* $[P_1]$ *and* $[P_2]$ *on N is*

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

Lemma 9. *The action of* $\mathcal{G}$ *on M is globally discrete, i.e., if there exists a bound* R_1 *relying only on the dimension d such that for any* $P \in M$ *and* $G_1, G_2 \in \mathcal{G}$, $(G_1 \neq G_2)$, *the following inequalities hold,*

$$\|PG_1 - PG_2\|_{HS} > R_1 \quad \text{and} \quad d_M(PG_1, PG_2) > R_1,$$

then the local isometric map ϕ_{PG} can be defined on the ball for any $G \in \mathcal{G}$

$$B_M(PG; R_1) = \{Q \in M \mid d_M(PG, Q) < \tfrac{1}{2}R_1\}.$$

Lemma 10. *If $r(t)$ is the curve on N and $\gamma_1(t), \gamma_2(t)$ are the two lifting curve of $r(t)$, satisfying $\phi(\gamma_1(t)) = \phi(\gamma_2(t)) = r(t)$, then*

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

Lemma 11. *If $e_1(\cdot), \ldots, e_d(\cdot) \in C^1[0,1]$ and Assumptions 1 and 2 hold, then for sufficiently large D,*

$$\begin{aligned}
&\lim_{n\to+\infty} pr\{\sup_{1\le i\le D} d_N([Q_{j_i}], [Q_{j_{i+1}}]) < R\} = 1,\\
&\lim_{n\to+\infty} pr\{\sup_{1\le i\le D} d_M(\mathcal{P}_{j_i}, \mathcal{P}_{j_{i+1}}) < R\} = 1,\\
&pr\{\sup_{1\le i\le D} d_M(P(t_{j_i}), P(t_{j_{i+1}})) < R\} = 1,
\end{aligned}$$

where R is the constant from Lemma 1 and pr refers to the probability of a set.

Lemma 12. *Suppose that r is a curve N, then $r'(t) = z(t)r(t)$ where $z(t) \in T_I M$ and $z(t)r(t) \in T_{r(t)}N$. Further more, if r is geodesic, then $z(t)$ is a constant.*

Lemma 13. *If $e_1, \ldots, e_d \in C^2[0,1]$ and Assumptions 1 and 2 hold, then*

$$\sup_{0\le t\le 1}(\|\tfrac{d}{dt}\tilde{P}(t)\|_{HS} + \|\tfrac{d^2}{dt^2}\tilde{P}(t)\|_{HS}) = O_p(1),$$

The items $\tilde{r}(t)$ and $\tilde{P}(t)$ are defined as $\tilde{r}(t) = R^{-1}_{\gamma(0)} \circ d\phi^{-1}_{\gamma(0)}(\dot{r}(t))$ and $\tilde{P}(t) = R^{-1}_{\gamma(0)} \circ d\phi^{-1}_{\gamma(0)}([\dot{P}(t)])$.

Lemma 14. *Suppose that r is a curve on N and Γ is a curve on M with $\phi(\Gamma) = r$. Then the tangent map of ϕ and the parallel transport commutes. To be specific, for any $s,t \in [0,1]$ and $v \in T_{r(t)}N$, we have*

$$\left(d\phi_{\Gamma(s)}\right)^{-1} \circ Par^{r(s)}_{r(t)}(v) = Par^{\Gamma(s)}_{\Gamma(t)} \circ \left(d\phi_{\Gamma(t)}\right)^{-1}(v).$$

Acknowledgments

Fang Yao is the corresponding author. We thank Dr. Zhenhua Lin for helpful discussion. Data were provided by World Bank, OECD National Accounts data files, Stockholm International Peace Research Institute (SIPRI), Yearbook: Armaments, Disarmament and International Security from https://data.worldbank.org/.

Funding

Fang Yao's research is partially supported by the National Key R&D Program of China (No. 2020YFE0204200), the National Natural Science Foundation of China (No. 12292981, 11931001), the LMAM and the Fundamental Research Funds for the Central Universities, Peking University (LMEQF).

Supplementary Material

Supplement to "Dynamic principal component analysis from a global perspective" (DOI: 10.3150/24-BEJ1743SUPP; .pdf). Supplementary material (Shao and Yao, 2024) includes detailed computation algorithm, technical proofs and additional simulations as well as the extension to general Lie group.

References

Bordin, C.J. and Bruno, M.G.S. (2020). Particle filtering on the complex Stiefel manifold with application to subspace tracking. In *ICASSP* 5485–5489.

Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory*, 2nd ed. *Holden-Day Series in Time Series Analysis*. Oakland, CA: Holden-Day, Inc. MR0595684

Cai, T.T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355. MR2906870 https://doi.org/10.1214/11-AOS898

Chen, E.Z. and Li, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics* **32** 2611–2617.

Clark, R.M. and Thompson, R. (1984). Statistical comparison of palaeomagnetic directional records from lake sediments. *Geophys. J. Int.* **76** 337–368.

do Carmo, M.P. (1992). *Riemannian Geometry*. *Mathematics: Theory & Applications*. Boston, MA: Birkhäuser, Inc. MR1138207 https://doi.org/10.1007/978-1-4757-2201-7

Gilmore, R. (2012). *Lie groups, Lie Algebras, and Some of Their Applications*. Courier Corporation.

Green, P.J. and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. *Monographs on Statistics and Applied Probability* **58**. London: CRC Press. MR1270012 https://doi.org/10.1007/978-1-4899-4473-3

Groisser, D., Jung, S. and Schwartzman, A. (2017). Geometric foundations for scaling-rotation statistics on symmetric positive definite matrices: Minimal smooth scaling-rotation curves in low dimensions. *Electron. J. Stat.* **11** 1092–1159. MR3631822 https://doi.org/10.1214/17-EJS1250

Hatcher, A. (2005). *Algebraic Topology*. Tsinghua University Press.

Hörmann, S., Kidziński, Ł. and Hallin, M. (2015). Dynamic functional principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 319–348. MR3310529 https://doi.org/10.1111/rssb.12076

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR3379106 https://doi.org/10.1002/9781118762547

Hu, X. and Yao, F. (2024). Dynamic principal component analysis in high dimensions. *J. Amer. Statist. Assoc.* **119** 308–319. MR4713894 https://doi.org/10.1080/01621459.2022.2115917

Jakubiak, J., Silva Leite, F. and Rodrigues, R.C. (2006). A two-step algorithm of smooth spline generation on Riemannian manifolds. *J. Comput. Appl. Math.* **194** 177–191. MR2239067 https://doi.org/10.1016/j.cam.2005.07.003

Jolliffe, I.T. (2002). *Principal Component Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2036084

Jupp, P.E. and Kent, J.T. (1987). Fitting smooth paths to spherical data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **36** 34–46. MR0887825 https://doi.org/10.2307/2347843

Kim, K.-R., Dryden, I.L., Le, H. and Severn, K.E. (2021). Smoothing splines on Riemannian manifolds, with applications to 3D shape space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 108–132. MR4220986 https://doi.org/10.1111/rssb.12402

Knapp, A.W. (2013). *Lie groups Beyond an Introduction*. Springer Science & Business Media.

Le, H. (2003). Unrolling shape curves. *J. Lond. Math. Soc. (2)* **68** 511–526. MR1994697 https://doi.org/10.1112/S0024610703004393

Machado, L. and Leite, F.S. (2006). Fitting smooth paths on Riemannian manifolds. *Int. J. Appl. Math. Stat.* **4** 25–53. MR2262660

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **2** 559–572.

Rodrigues, R.C., Silva Leite, F. and Jakubiak, J. (2005). A new geometric algorithm to generate smooth interpolating curves on Riemannian manifolds. *LMS J. Comput. Math.* **8** 251–266. MR2193213 https://doi.org/10.1112/S146115700000098X

Rui, C. and Leite, F.S. (2007). Interpolation problems on Riemannian manifolds – a geometric approach. *IFAC Proc. Vol.* **40** 822–827.

Samir, C., Absil, P.-A., Srivastava, A. and Klassen, E. (2012). A gradient-descent method for curve fitting on Riemannian manifolds. *Found. Comput. Math.* **12** 49–73. MR2886156 https://doi.org/10.1007/s10208-011-9091-7

Shampine, L.F., Gladwell, I. and Thompson, S. (2003). *Solving ODEs with Matlab*. Cambridge: Cambridge Univ. Press. MR1985643 https://doi.org/10.1017/CBO9780511615542

Shao, L. and Yao, F. (2024). Supplement to "Dynamic principal component analysis from a global perspective." https://doi.org/10.3150/24-BEJ1743SUPP

Srivastava, A. and Klassen, E. (2004). Bayesian and geometric subspace tracking. *Adv. in Appl. Probab.* **36** 43–56. MR2035773 https://doi.org/10.1239/aap/1077134463

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745

Yuan, Y., Zhu, H., Lin, W. and Marron, J.S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956 https://doi.org/10.1111/j.1467-9868.2011.01022.x

Zhu, H., Chen, Y., Ibrahim, J.G., Li, Y., Hall, C. and Lin, W. (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Amer. Statist. Assoc.* **104** 1203–1212. MR2750245 https://doi.org/10.1198/jasa.2009.tm08096

Received January 2023 and revised February 2024