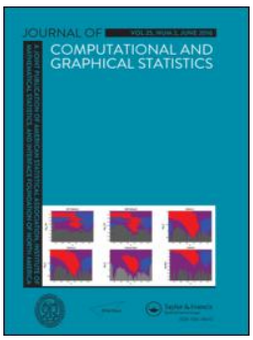

Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: http://www.tandfonline.com/loi/ucgs20

Principal Nested Spheres for Time-Warped Functional Data Analysis

Qunqun Yu, Xiaosun Lu & J. S. Marron

To cite this article: Qunqun Yu, Xiaosun Lu & J. S. Marron (2017) Principal Nested Spheres for Time-Warped Functional Data Analysis, Journal of Computational and Graphical Statistics, 26:1, 144-151, DOI: 10.1080/10618600.2015.1115359

To link to this article: http://dx.doi.org/10.1080/10618600.2015.1115359

Accepted author version posted online: 03 Dec 2015.
Published online: 16 Feb 2017.

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2017, VOL. 26, NO. 1, 144–151
http://dx.doi.org/10.1080/10618600.2015.1115359

Taylor & Francis
Taylor & Francis Group

Principal Nested Spheres for Time-Warped Functional Data Analysis

Qunqun Yu[a,*], Xiaosun Lu[b,*], and J. S. Marron[a,*]

[a]Department of Statistics and Operations Research, University of North Carolina Chapel Hill, Chapel Hill, North Carolina; [b]Department of Biostatistics, Quintiles, Durham, North Carolina

ABSTRACT

There are often two important types of variation in functional data: the horizontal (or phase) variation and the vertical (or amplitude) variation. These two types of variation have been appropriately separated and modeled through a domain warping method (or curve registration) based on the Fisher–Rao metric. This article focuses on the analysis of the horizontal variation, captured by the domain warping functions. The square-root velocity function representation transforms the manifold of the warping functions to a Hilbert sphere. Motivated by recent results on manifold analogs of principal component analysis, we propose to analyze the horizontal variation via a principal nested spheres approach. Compared with earlier approaches, such as approximating tangent plane principal component analysis, this is seen to be an efficient and interpretable approach to decompose the horizontal variation in both simulated and real data examples.

ARTICLE HISTORY
Received October 2015
Accepted October 2015

KEYWORDS
Functional data variability;
Principal nested spheres;
Time warping

1. Introduction

A common phenomenon in functional data is that prominent features in the functions vary in position from one sample to another, such as the timing variation of the adolescent growth spurt in human growth curves. This positional, or phase, variation is called the *horizontal variation*. Another important component of variability in functional data is the amplitude variation, or the *vertical variation*, such as the height differences among the individuals.

There is a large literature on statistical analysis of functions, such as Ramsay (1982), Kneip and Gasser (1992), and Locantore et al. (1999). A general overview of functional data analysis is provided by Ramsay and Silverman (2002, 2005) and a curve registration overview article is provided by Marron et al. (2015). Plenty of useful tools and methods are available, such as functional principal component analysis (FPCA), with many important applications in a wide variety of scientific fields. One open problem in functional data analysis is that in those traditional approaches, the functional data are analyzed under the $\mathbb{L}^2$ metric that tends to strongly focus on the vertical variation. The horizontal variation cannot be easily understood in these vertical analyses. Section 2.1 gives examples illustrating the shortcoming of conventional FPCA.

The main purpose of this article is to find an improved method for the *horizontal analysis*, that is, the analysis of the horizontal variation. Considering the special spherical structure of the horizontal variation (see Section 1.2 for details), we propose to use an approach involving Principal Nested Spheres (PNS) introduced by Jung, Dryden, and Marron (2012). Comparison with several other popular approaches, such as the FPCA, suggests improved efficiency of PNS for horizontal analysis. A toy example (Section 2) and a real data example of blood glucose time series (Section 3) are used to illustrate the advantages of the PNS approach.

1.1 Function Alignment Based on the Fisher–Rao Metric

A useful approach to horizontal analysis is through the idea of *elastic functions*. Some pioneering work in this area includes Cameron (1983), Hardle and Marron (1990), Ramsay and Li (1998), Gervini and Gasser (2004), Liu and Mueller (2004), Kneip and Ramsay (2008), and Tong and Mueller (2008). The basic idea is to first separate the vertical and the horizontal variation through *function alignment*, or curve registration. In particular, consider a collection of functions $f_i(t)$, $i = 1, 2, \ldots, n$ in $\mathcal{F} = \{f \mid f \text{ is absolutely continuous on} [0, 1]\}$ (if the domain is not $[0, 1]$, consider a linear transformation that maps the domain to $[0, 1]$), having both vertical and horizontal variation, such as the bimodal functions shown in Figure 1 (left). Let Γ be the set of orientation-preserving diffeomorphisms of the unit interval $[0, 1]$: $\Gamma = \{\gamma : [0, 1] \to [0, 1] | \gamma(0) = 0, \gamma(1) = 1$ and γ is a diffeomorphism$\}$, where a *diffeomorphism* refers to a bijective differentiable function whose inverse is also differentiable. If functions $f_i(t)$, $i = 1, 2, \ldots, n$ are well aligned by warping the domain properly, then the horizontal and the vertical variation can be separately captured by the domain warping functions $\gamma_i(t) \in \Gamma$ and the resulting aligned functions $f_i(\gamma_i(t))$, respectively. For this toy example, such a set of warping functions and the corresponding aligned functions are shown in the middle and right panels, respectively (details about finding those warping functions are discussed later). Then, the horizontal analysis can be done by studying those warping functions.

A crucial step in the function alignment is to find appropriate domain warping functions. Consider two functions f_1 and

CONTACT J. S. Marron marron@unc.edu Department of Statistics and Operations Research, University of North Carolina Chapel Hill, Chapel Hill, NC 27599.
*All authors equally contributed to this article.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JCGS.

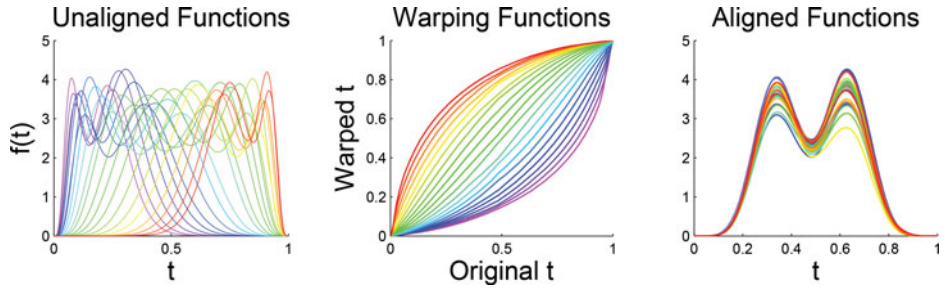

Figure 1. Left: A toy example of bimodal functions with big horizontal variation. The color reflects the order of the horizontal positions of the peaks. Middle: The domain warping functions to align the functions based on the Fisher–Rao metric. Right: The aligned functions. A common color scheme is used in each panel using rainbow color in order of amount of warp.

f_2. Most of the past approaches involve solving $\inf_{\gamma\in\Gamma}\|f_1 - (f_2\circ\gamma)\|$ to align f_2 to f_1, where $\|\cdot\|$ is the standard $\mathbb{L}^2$ metric, that is, $\|f\| = (\int_0^1 |f(t)|^2 dt)^{1/2}$. However, this criterion is problematic, since the objective function is not symmetric in the sense that aligning f_1 to f_2 leads to a different optimal minimum. To illustrate this point, Figure 2, similar to fig. 8 in Marron et al. (2015), shows a simple example of aligning two step functions. It is seen that aligning f_2 to f_1 (middle) and aligning f_1 to f_2 (right) are different under the $\mathbb{L}^2$ metric. The difference between the horizontally hatched blue area in Panel (2, 2) and the vertically hatched pink area in Panel (2, 3) indicates that the two corresponding objective functions $\|f_1 - (f_2\circ\gamma)\|$ and $\|(f_1\circ\gamma) - f_2\|$ are not equal. This is because the $\mathbb{L}^2$ metric is not invariant under reparameterization, or domain warping. In particular, $\|f_1 - f_2\| \neq \|f_1\circ\gamma - f_2\circ\gamma\|$. Srivastava et al. (2011) considered a more appropriate metric in the quotient space $\mathcal{F}/\Gamma$. The rest of this section is a review of the curve registration method proposed in that article. The quotient space $\mathcal{F}/\Gamma$ is the set of equivalence classes (or orbits) of the types $[f] = \{(f\circ\gamma)|\gamma\in\Gamma\}$. The metric on this quotient space motivates us to change our data object from f in $\mathcal{F}$ to the equivalence class $[f]$ in this quotient space. A good understanding of choosing the proper data object can be obtained using the terminology of *object-oriented data analysis* (OODA), introduced by Wang and Marron (2007) and recently surveyed in Marron and Alonso (2014). The data objects are understood as the *atoms* of the analysis. In this section, the data objects are the equivalence classes and a useful metric defined on the quotient space is the Fisher–Rao metric. See Srivastava et al. (2011) for definition and relevant theory. This metric is derived from a Riemannian metric first introduced by Rao (1945). A nice property of the Fisher–Rao metric is that it is warping-invariant. In fact, Cencov (1982, chap. 2) proved that it is the only metric in the tangent space of $\mathcal{F}$ that has this property. Thus, we propose to use the Fisher–Rao metric to align functions for the purpose of the horizontal analysis.

Direct calculations based on the Fisher–Rao metric are challenging. In practice, a convenient square-root velocity function (SRVF) representation, that is, transforming the function $f(t)$ to $\psi(t) = \frac{\dot{f}(t)}{\sqrt{|\dot{f}(t)|}}$, simplifies the Fisher–Rao framework and the SRVF of $f\circ\gamma$, denoted as (ψ,γ), is given by $\tilde{\psi}(t) = \frac{\frac{d}{dt}(f\circ\gamma)(t)}{\sqrt{|\frac{d}{dt}(f\circ\gamma)(t)|}} = (\psi\circ\gamma)(t)\sqrt{\dot{\gamma}(t)}$. The orbit of an SRVF $\psi\in\mathbb{L}^2$ is given by $[\psi] = \text{closure}\{(\psi,\gamma)|\gamma\in\Gamma\} =$

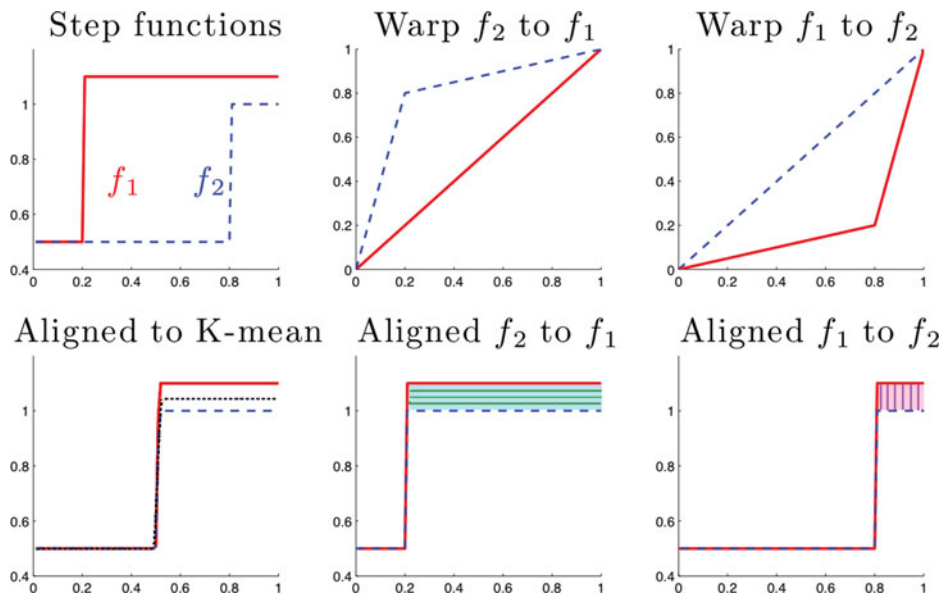

Figure 2. The problem with $\mathbb{L}^2$ metric alignment. The top left panel shows two step functions f_1 (solid red) and f_2 (dashed blue). The four right panels show that warping f_2 to f_1 (middle) is different from warping f_1 to f_2 (right) under the $\mathbb{L}^2$ metric. The top two panels show the warping functions, while the bottom two panels show the aligned functions. Better than either is the Fisher–Rao alignment shown in the bottom left panel, where the black dotted line indicates the Karcher mean function.

closure$\{(\psi \circ \gamma)\sqrt{\dot{\gamma}} | \gamma \in \Gamma\}$. Let $\mathcal{S}$ denote the set of all such orbits. Under the SRVF representation, the Fisher–Rao metric becomes the standard $\mathbb{L}^2$ metric, that is, $d_{\text{FR}}(f_1, f_2) = ||\psi_1 - \psi_2||$, where ψ_1 and ψ_2 are the SRVFs corresponding to f_1 and f_2. Thus, standard statistical tools for the $\mathbb{L}^2$ space, such as mean, covariance, and principal components, can be used. The Karcher mean of the given SRVF orbits $\{[\psi_i]\}$ in the space $\mathcal{S}$ is defined as a local minimum of the sum of squares of elastic distance: $[u]_n = \text{argmin}_{[\psi]\in\mathcal{S}} \sum_{i=1}^n d([\psi], [\psi_i])^2 = \text{argmin}_{[\psi]\in\mathcal{S}} \sum_{i=1}^n inf_{\gamma\in\Gamma} ||\psi - (\psi_i, \gamma)||$. Then, we get a representative u in the orbit $[u]_n$ with respect to the SRVFs ψ_i, $i = 1, 2, \ldots, n$. Warp ψ_i to u by $\gamma_i = \text{argmin}_{\gamma\in\Gamma} ||u - (\psi_i, \gamma)||$. The aligned functions of f_i, $i = 1, 2, \ldots, n$ are $\tilde{f}_i = f_i \circ \gamma_i$ and the Karcher mean of f_i, $i = 1, 2, \ldots, n$ is the cross-sectional mean of the aligned functions. Please refer to Srivastava et al. (2011) for more details about the computation of Fisher–Rao curve registration. As an example of function alignment based on the Fisher–Rao metric, the warping functions (middle) for the toy data (left) in Figure 1 are found by an automatic and unsupervised approach based on this metric, proposed by Srivastava et al. (2011).

The differentiation part of Fisher–Rao alignment means that it is more strongly impacted by high noise levels than other approaches. But the trade-off is that we achieve much better peak alignment than these other methods can. In situations where the noise is so high that many randomly located spurious peaks appear, peak alignment may not be a useful thing to do, so other methods are expected to be more effective. However, the results in Section 3 indicate good performance even in a moderately high noise, real data context, and Srivastava et al. (2011) showed many more.

1.2 PNS for Spherical Structure of Horizontal SRVFs

In the above section, we found the SRVF representation to be very useful for replacing the Fisher–Rao metric in the quotient space $\mathcal{F}/\Gamma$ with the $\mathbb{L}^2$ metric in the space $\mathcal{S}$. In this section, we find that SRVF presents a major challenge, which is that it transforms the manifold of the warping functions to a Hilbert sphere. As in Section 1.1, consider a set of functions f_i, $i = 1, 2, \ldots, n$ and let $\gamma_i \in \Gamma$ be a warping function for f_i. Then the SRVF ψ_i of the warping function γ_i, referred to later as the *horizontal SRVF*, can be written as $\sqrt{\dot{\gamma}_i}$. Noting that $\|\psi_i\|^2 = \int_0^1 \psi_i(t)^2 dt = \int_0^1 \dot{\gamma}_i(t)dt = \gamma_i(1) - \gamma_i(0) = 1$, these horizontal SRVFs naturally lie on the surface of a Hilbert unit sphere. More precisely, since γ_i is a diffeomorphism, the SRVFs stay in the positive orthant of a Hilbert unit sphere.

Inspired by the spherical structure of the horizontal SRVFs, we propose to use the PNS method for horizontal analysis. Introduced by Jung, Dryden, and Marron (2012), PNS is an extension of PCA for curved manifolds, especially for high-dimensional spheres. To understand PNS, it is useful to think of PCA in terms of a nested series of approximating hyperplanes. For $k = 1, \ldots, d$, the hyperplane of dimension k is the best approximating hyperplane of the data, and is given by the plane through the mean, in the direction of the first k eigenvectors. From dimension k, the plane of dimension $k-1$ can be found by simply removing the kth eigenvector. It can also be defined as the minimizer, over hyperplanes of dimension $k-1$, of the sum of squared residuals of the k-dimensional projections onto the $k-1$-dimensional plane. Note that the rank $k-1$ PC scores are the resulting signed residuals. This characterization of PCA gives a clear view of PNS, which is a nested series of subspheres of decreasing dimension, which similarly are each good approximations of the data. Now this idea is extended to data on the unit d-sphere, which is the set of unit vectors in R^d. Given projections onto the subsphere of dimension k, a best-fitting subsphere of dimension $k-1$ is found to minimize the sum of squared arc lengths to the projections onto the lower dimensional sphere. At each step, the signed arc lengths give the corresponding PNS scores. This method uses a *backward approach*, which starts with the high-dimensional sphere and iteratively finds the best-fitting subsphere of one dimension lower. See Marron et al. (2010) and Damon and Marron (2014) for more discussion of backward PCA. It has been shown in a number of cases that PNS can provide more effective analysis of manifold data than many other analogous approaches. See Pizer et al. (2013) for a very effective use of PNS in the study of three-dimensional (3D) shapes.

For comparison purposes, another popular approach to data lying in curved manifolds, principal geodesic analysis (PGA) proposed by Fletcher et al. (2004), is also investigated in this article. Unlike PNS, it is a forward approach, starting with the *Karcher mean*. In the following horizontal analyses, the Karcher mean refers to the representer defined in Definition 3 of Srivastava et al. (2011). PGA approximates the spherical surface by a tangent hyperplane centered at the Karcher mean. By performing PCA on this tangent plane, PGA finds the principal geodesics (i.e., great spheres) passing through the mean that best fit the data.

In contrast to PGA, the PNS method finds the best-fitting subsphere regardless of whether it is a great sphere or not. As noted by Jung, Dryden, and Marron (2012), a great subsphere of a high-dimensional sphere is the intersection of the high-dimensional sphere and a hyperplane that passes through the center point of the sphere and is a natural extension of the concept of great circle in high-dimensional space. When the major variance is nongeodesic, PNS tends to find the best-fitting small spheres instead of only great spheres. Thus, when the data variability on the sphere is big enough, the PNS can give a much more effective decomposition of this variability than PGA. On the other hand, if the data variability is small, the PNS method does not improve much over the PGA method. This is because in this case the data do not have much curvature and can be approximated by a tangent plane well enough. In the following discussion, we focus on examples with big horizontal variation.

2. Horizontal Analyses

This section compares different horizontal analyses of the toy example in Figure 1 (left), where the functions have big horizontal variation and the PNS method gives very useful improvement over PGA.

Figure 3 (left) visualizes the pure horizontal shifts of the peaks for this toy example, which shows warps of the Karcher mean function (red curve in the right panel) by the Fisher–Rao warping functions in Figure 1 (middle). These functions will be

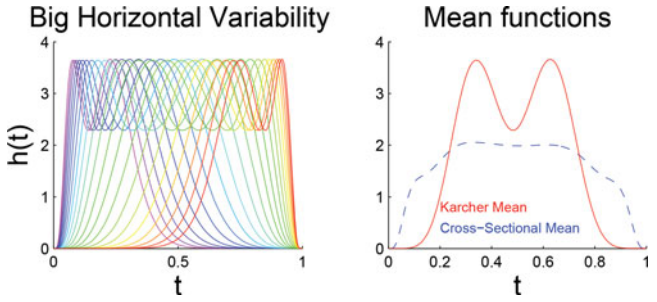

Figure 3. Left: Horizontal variation of the toy example in Figure 1 (left). The color reflects the order of the horizontal positions of the peaks. Right: The Karcher mean function (red solid line) and the cross-sectional mean (blue dashed line). The same rainbow color is used here as in Figure 1.

referred to later as the *horizontally shifted functions*, denoted by $h_i, i = 1, 2, \ldots, n$.

OODA provides a useful framework for studying and comparing the several options that are available as representatives of horizontal variation. In this study, the original data objects are functions. There are three potential candidate data objects for horizontal analysis that are studied in this section: the horizontally shifted functions h_i, the warping functions γ_i, and the horizontal SRVFs ψ_i. Oriented by the need to choose among these candidate data objects, four different horizontal analyses have been performed on the toy data:

1. H-PCA: FPCA of the horizontally shifted functions h_i;
2. E-PCA: FPCA of the warping functions γ_i;
3. S-PGA: PGA of the horizontal ψ_i; and
4. S-PNS: PNS of the horizontal ψ_i.

The first two approaches, using the conventional FPCA, are discussed in Section 2.1. The latter two manifold approaches, motivated by the spherical structure of the horizontal SRVFs, are discussed in Section 2.2. Section 3 compares some of these approaches on a blood glucose dataset. Section 4 summarizes the comparison of these four approaches.

2.1 Conventional FPCA

An intuitive way to understand the horizontal variation of the toy data is to analyze either the horizontally shifted functions h_i or the warping functions γ_i. As FPCA is one of the most widely used statistical tools for functional data analysis, this section discusses both H-PCA and E-PCA. It is seen that H-PCA is rarely a good option for horizontal analysis.

2.1.1 H-PCA

When there is a large amount of horizontal variation, H-PCA is strongly impacted in two ways. First, as shown in the right panel of Figure 3, H-PCA is centered at the cross-sectional mean that is a poor notion of centerpoint. In particular, that blue dashed line does not show bimodal structure, which is an important characteristic of each member of the dataset. Second, H-PCA gives a hard to interpret impression of the major modes of variation in the dataset, as shown in the left two panels of Figure 4. In particular, the natural strong (intuitively low-dimensional) horizontal variation in the data is spread across the PCA spectrum, because this type of variation is strongly nonlinear in the PCA sense. Therefore, H-PCA is not an appropriate approach for horizontal analysis.

2.1.2 E-PCA

E-PCA gives an eigenanalysis of the warping functions $\gamma_i, i = 1, 2, \ldots, n$. Each component gives a mode of variation in that space. It can be challenging to interpret such plots in terms of their implications about the horizontal variation. Better interpretation comes from transforming the decomposition of the warping functions into the original function space, that is, warping the Karcher mean function by the E-PC projections. The second column of Figure 4 shows the first two transformed E-PC projections for the toy example. These two components provide a much more useful summary of the apparent horizontal variation in the raw data than the previous ones from the H-PCA (first column). The first component reflects the horizontal shifts of the peaks, while the second one is about the horizontal distance between the two peaks.

However, this approach has a serious weakness. That is, the PC projection of a warping function is not necessarily bijective, and thus, not a warping function. In other words, the E-PCA can leave the space Γ of warping functions. To illustrate this,

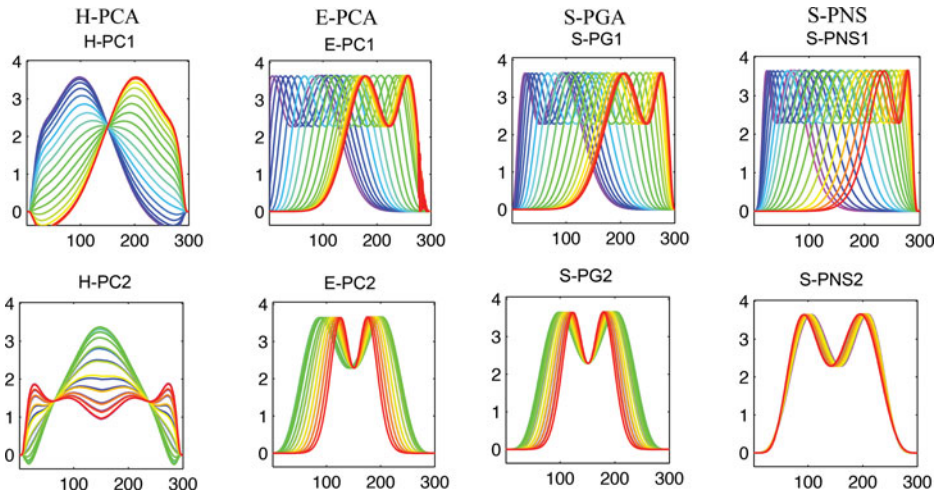

Figure 4. Horizontal analyses of the toy data. The color is consistent with that in the first panel in Figure 3. From the left column to the right: H-PCA, E-PCA, S-PGA, S-PNS. Each column shows the first two components of each analysis. Note that successive improvement in quality of data representation and signal compression is shown.

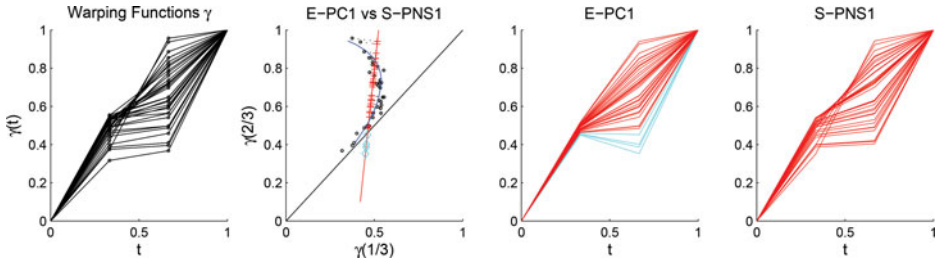

Figure 5. A toy example to illustrate the potential problem of E-PCA. The first panel: A set of two-dimensional warping functions γ_i, each determined by $\gamma_i(1/3)$ and $\gamma_i(2/3)$. The second panel: The scatterplot of $\gamma_i(1/3)$ and $\gamma_i(2/3)$ (black circles), the E-PC1 direction (red line) of these points and the S-PNS1 curve (blue curve). The red crosses indicate the E-PC1 projections above the black diagonal line, and the cyan diamonds indicate the E-PC1 projections below the line. The blue curve of S-PNS1 projections remains in the positive orthant. The third panel: The projected curve visualization of those E-PC1 projections. Note that the cyan curves are not bijective, that is, not valid warping functions. The fourth panel: The projected curve visualization of the S-PNS1 projections. These projections stay in the space Γ.

Figure 5 shows the FPCA of a set of simple two-dimensional warping functions γ_i (left panel), each of which is determined by two values, $\gamma_i(1/3)$ and $\gamma_i(2/3)$. It is seen in the middle two panels that some of the E-PC1 (the first E-PCA component) projections (cyan) have a decreasing part, that is, $\gamma_i(1/3) > \gamma_i(2/3)$. Warping the Karcher mean function with these nonwarping E-PC projections is problematic. Computationally, this causes the wiggly right end of the yellow and the red functions in Panel (1, 2) of Figure 4. This problem can be avoided using the S-PNS method. The blue curve in the second panel in Figure 5 shows the nonlinear path of the S-PNS1 projections. Note that this curve remains entirely in the positive orthant. This is also seen in the fourth panel. In particular, these S-PNS1 projections are all proper warps.

2.2 Analyses on SRVF Manifold

The following analyses avoid the problem shown in Section 2.1.2, by appropriately using the spherical structure of the horizontal SRVFs ψ_i. The idea is to first decompose the variability of the spherical SRVFs and then transform the projections of the SRVF components back to the warping function space Γ using the formula $\gamma_\xi(t) = \int_0^t \xi(t)^2 dt$, where ξ is a point on the SRVF sphere. It can be easily checked that $\gamma_\xi \in \Gamma$. Finally, the decomposition of the horizontal variation of the original functions can be obtained via warping the Karcher mean function with the transformed SRVF projections. The following discussion shows how manifold approaches, especially the S-PNS approach, can provide a more efficient decomposition for horizontal analysis than the conventional FPCA.

2.2.1 S-PGA

The third column in Figure 4 shows the first two components of the horizontal variation in the toy data, based on the S-PGA analysis. Compared with the H-PCA results (first column), this approach gives a much better decomposition of the horizontal variation. At first glance, the S-PGA decomposition (third column) looks similar to E-PCA (second column). But a close look shows some distortion in the lower right of the red curves, and the purple curves moving out of the plot window because the corresponding warps have left the positive orthant.

2.2.2 S-PNS

The first two components of the horizontal variation in the toy data based on S-PNS are shown in the fourth column in Figure 4. Results from this decomposition give more signal compression than those from the previous analyses. The first component simultaneously captures both the mode of peak location and the mode of distance between peaks. Thus, the two components previously needed are now reduced to essentially one. Among the four panels in the first row, these S-PNS1 projections explain the horizontal variation of the original bimodal functions best, as they are almost identical to the raw horizontal warps of the Karcher mean, shown in the left panel of Figure 3. Very little variability is left for the second S-PNS component to explain. This suggests that the horizontal variability is almost one dimensional in some nonlinear sense, which is consistent with the fact that the warping function γ_i in this toy example can be summarized by a single parameter a_i. In particular, these were generated as $\gamma_i(t) = \frac{e^{a_i t}-1}{e^{a_i}-1}$, for $a_i \in [-5, 5]$.

For further insight of this type, Figure 6 compares S-PNS1 and E-PC1 scores in the 3D space generated by the E-PC1, E-PC2, and E-PC3 directions. The stars are the scores of the warping functions in Figure 1. The squares are the projections of the stars onto E-PC1, corresponding to the curves shown in the Panel (1, 2) in Figure 4, in this 3D space. The circles are the

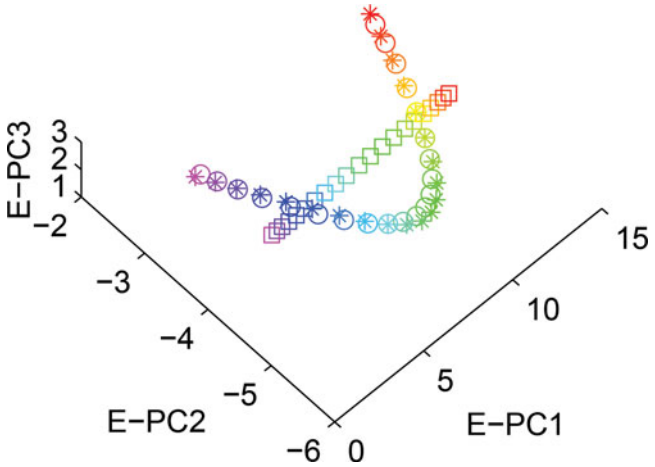

Figure 6. Visualization of S-PNS1 and E-PC1 scores in the 3D space generated by the E-PC1, E-PC2, and E-PC3 directions for the toy example. Stars are the E-PC1, E-PC2, E-PC3 scores, squares are the E-PC1 projections, and circles are the S-PNS1 projections. Same rainbow color scheme as in Figure 1 is used. This plot shows S-PNS1 gives a much better one-dimensional representation than E-PC1, of the population of warps in this 3D space.

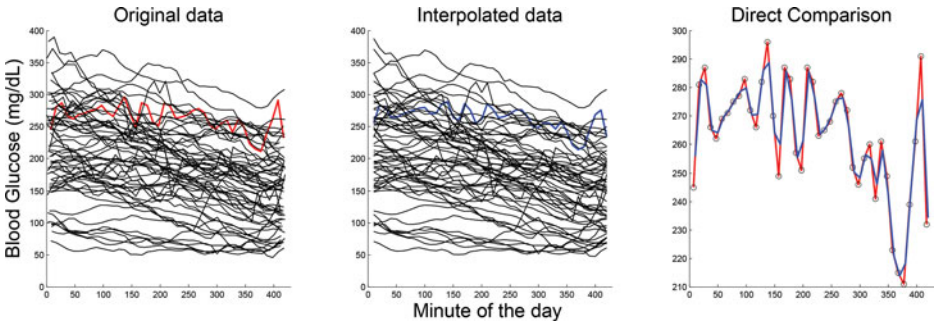

Figure 7. Study of effect of interpolation. Left panel shows the original blood glucose curves over 59 days for one subject. Middle panel shows the application of linear interpolation to the data in the left panel. The right panel is a close-up of the interpolation of the red curve (chosen as a challenging case) on the left by the blue curve in the middle showing that the interpolation captures major shape aspects of the curve.

S-PNS1 projections, corresponding to the curves shown in the Panel (1, 4) in Figure 4. In this 3D space, we can clearly see that S-PNS1 gives a much better one-dimensional representation of the original warps than E-PC1. Matlab scripts for generating this toy example, and for analysis using the S-PNS method are provided on Marron's webpage *http://www.unc.edu/depts/stator/miscellaneous/marron/Matlab7Software/SRVF/*.

3. Blood Glucose Example

We now apply E-PCA and S-PNS to study phase and amplitude variation in Continuous Glucose Monitoring (CGM) data in the management of Type 1 diabetes in young children aged 4–9 years (Mauras et al. 2012). H-PCA is not shown because of its inferior performance in Section 2.1.1 and S-PGA is not also explicitly shown here because it is similar to E-PCA. A CGM device makes glucose measurements every few minutes (usually 5 or 10) every day. Our data objects are the curves formed by the measurements for each day for each patient. The number of days per patient ranges from 12 to 391. Both the height and timing of the glucose peaks and valleys (amplitude and phase variation) are vital for diabetics. In this section, we compare the performances of E-PCA and S-PNS on the horizontal phase variation and our results will show that S-PNS1 explains much more variation than E-PC1, consistent with the previous toy example.

There were 146 children in this study. We looked at multiple children at multiple time windows and we chose the CGM curves for one representative child (in the sense of median PC1 score) in the night time window 00:00:00–07:00:00 over days as our data objects, shown in the left panel of Figure 7. Missing values in the CGM data are an important challenge. First, to avoid boundary interpolation problems, we excluded the days that had no measurement within 10 min of the start or the end. To give the data curves a common set of time points, we applied linear interpolation to the regular grid points (00:10:00, 00:20:00, …, 07:00:00) in the time window 00:00:00–07:00:00 (a time window of particular interest) over days. To reduce the gap between the true blood glucose levels and the estimated blood glucose levels introduced by the linear interpolation, we excluded the days having any pairwise empty interval bigger than 25 min. The middle panel in Figure 7 shows the interpolated curves. The red curve in the left panel is one original curve, visually chosen as a challenging case to interpolate. The blue curve in the middle panel is the corresponding interpolated version. The right panel focuses on the quality of the interpolation by zooming in on those two curves. The blue curve reduces the sharp corners of the red one, but does not distort the overall shape of the data, which can also be seen by the fact that the left panel and middle panel look similar to each other. Hence, the following analysis is based on the interpolated data in Figure 7.

3.1 Phase and Amplitude Separation

To compare the performance of E-PCA and S-PNS on horizontal variation of the blood glucose example, we first applied Fisher–Rao curve registration to separate horizontal and vertical variation on the interpolated curves in Figure 7. The first panel of Figure 8 shows the warping functions. The second panel shows the aligned functions that represent a large amount of vertical variation. Note most curves have been aligned with three common peaks. The third panel is the Karcher mean of the interpolated curves showing the same three peaks. The last one gives a more interpretable view of the horizontal variation obtained via warping the Karcher mean by the inverse of the warping functions. This highlights the horizontal variation in the three peaks. There is interesting variation in the timing of these peaks, which is studied more deeply in next section.

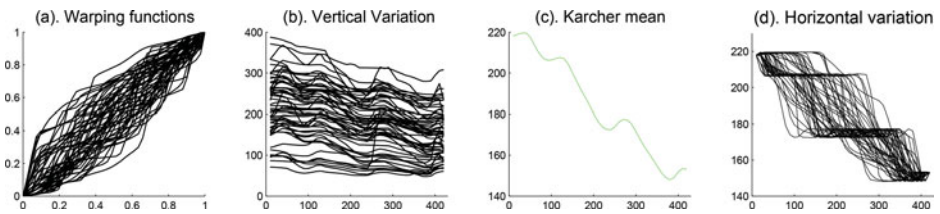

Figure 8. Phase and amplitude separation. (a) Warping functions. (b) Vertical variation. (c) Karcher mean derived by Fisher–Rao curve registration showing three distinct peaks. (d) Horizontal variation obtained via warping the Karcher mean by the inverse of the warping functions. These show both strong vertical (b) and horizontal (d) variations.

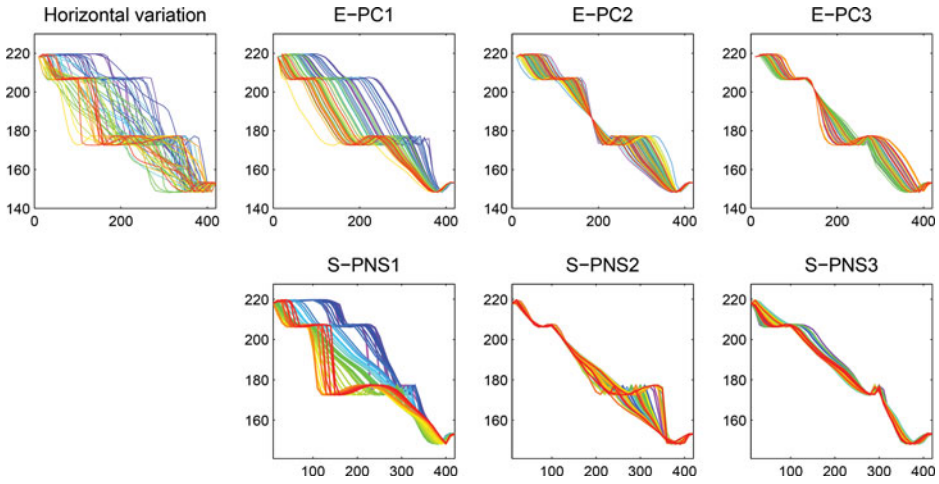

Figure 9. Horizontal variation analyses: The first column is the horizontal variation shown in the original space. These curves are the same as in the right panel in Figure 8, but with colors in the order of S-PNS1 scores, used in all the plots. The upper right three panels are the first three components of the transformed E-PC projections. The lower panels are the S-PNS analysis. These show S-PNS explains more variation with fewer components than E-PCA.

3.2 E-PCA Versus S-PNS of Horizontal Variation

In Figure 9, the upper left panel shows the horizontal variation using a rainbow color scheme (purple, blue, cyan, green, yellow, red) in the order of S-PNS1 scores. The upper second to fourth panels show the transformed E-PC projections using E-PCA with the same color scheme. E-PC1 captures an overall timing effect, E-PC2 feels the slope from second peak to third peak and E-PC3 is driven by the width of the third peak. The lower panels are the first three components of the horizontal variation based on S-PNS. In comparison to E-PC1, S-PNS1 also contains a similar overall timing effect. But it feels additional variation in terms of the slope from the second peak to the third peak, which drove E-PC2, indicated by the different shapes of the curves, and it shows the width of the third peak, seen in E-PC3. S-PNS2 is driven by the location of the third peak, which is split between E-PC2 and E-PC3. S-PNS3 shows relatively less variation because the first two components have already explained most of the interesting variation.

To better understand the mode of variation explained in the S-PNS1 projections, we replace the randomly spaced data scores by an equally spaced grid of coefficients from the smallest S-PNS1 score to the largest and show the resulting S-PNS1 projections in Figure 10. We can clearly see the timing effect, the effect of slope from the second peak to the third peak from E-PC2 and the third peak's width effect from E-PC3 simultaneously. S-PNS2 contains the timing of the third peak, an important part of E-PC2. The main point is that S-PNS explains much more variation with fewer components than E-PCA, that is, gives a more efficient data representation.

Figure 11 provides a 3D plot of S-PNS1 and E-PC1 in the space of E-PC1, E-PC2, E-PC3 in the same format as Figure 6. The stars are the scores of the warping functions shown in the left panel in Figure 8. The squares are the projections of the stars onto the E-PC1 direction, corresponding to the curves shown in the upper second panel of Figure 9, in this 3D space. The black line shows the E-PC1 direction going through the mean, that is, the best 1D linear approximation of the data. The circles are the S-PNS1 projections, corresponding to the curves in the lower second panel in Figure 9. The black curve shows the

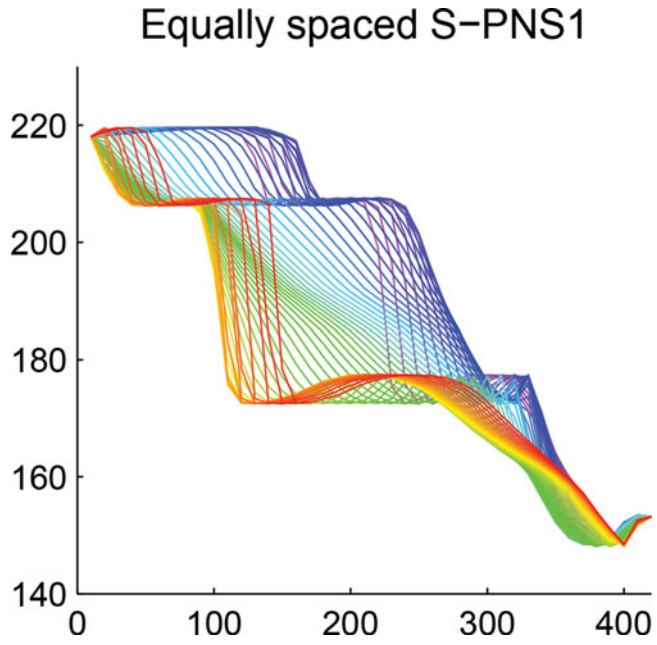

Figure 10. Equally spaced S-PNS1 projections using similar rainbow colors. This equal spacing of coefficients allows better interpretation of this mode of variation.

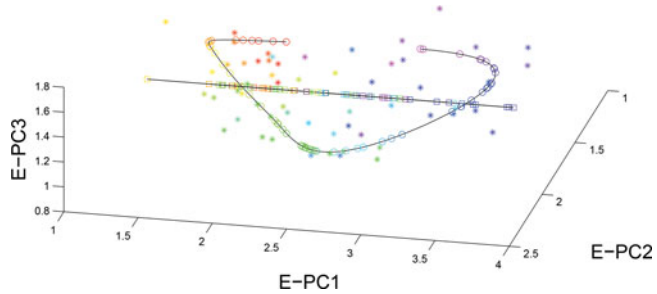

Figure 11. Visualization of S-PNS1 and E-PC1 scores in the 3D space generated by the E-PC1, E-PC2, and E-PC3 directions for the blood glucose data: Stars are the E-PC1, E-PC2, E-PC3 scores, squares are the E-PC1 projections, and circles are the S-PNS1 projections. Rainbow color in the order of S-PNS1 scores is used. This plot shows S-PNS1 gives a better one-dimensional representation of the original data in this 3D space.

best S-PNS one-dimensional approximation of the data, using a linear interpolation through the points corresponding to the curves in Figure 10. The same rainbow color in the order of S-PNS1 scores is used everywhere in this plot. The stars are more close to the S-PNS1 curve than the E-PC1 curve, which indicates S-PNS1 captures more variation and better represents the original data in this 3D space. This shows that S-PNS1 essentially has much greater flexibility to allow modeling with a richer nonlinear one-dimensional mode of variation.

4. Conclusions

This article aimed at finding an appropriate method for horizontal analysis of functional data, where the horizontal variation is separated from the vertical variation using a domain-warping method based on the Fisher–Rao metric. Four different approaches, including two conventional FPCA approaches (H-PCA and E-PCA) and two manifold approaches based on the spherical structure of the horizontal SRVFs (S-PGA and S-PNS), have been applied to a toy example and two of these (E-PCA and S-PNS) have been applied to the blood glucose example for comparison. The manifold approaches are generally better than the FPCA approaches, and S-PNS works the best in terms of both the signal compression and the interpretability of the results. As a conclusion, we propose to use the S-PNS approach for horizontal analysis, especially when the horizontal variability is large.

Acknowledgments

The authors thank the reviewers for their valuable feedback that led to significant improvement of the article. The authors are also grateful to Kai Zhang for helpful comments. This material was based upon work partially supported by the National Science Foundation under Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Cameron, M. A. (1983), "The Comparison of Time Series Recorders," *Technometrics*, 25, 9–22. [144]

Cencov, N. N. (1982), *Statistical Decision Rules and Optimal Inferences* (Vol. 53), Providence, RI: American Mathematical Society. [145]

Damon, J., and Marron, J. S. (2014), "Backwards Principal Component Analysis and Principal Nested Relations," *Journal of Mathematical Imaging and Vision*, 50, 107–114. [146]

Fletcher, P. T., Lu, C., Pizer, S. M., and Joshi, S. (2004), "Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape," *IEEE Transactions on Medical Imaging*, 23, 995–1005. [146]

Gervini, D., and Gasser, T. (2004), "Self-Modeling Warping Functions," *Journal of the Royal Statistical Society*, 66, 959–971. [144]

Hardle, W., and Marron, J. S. (1990), "Semiparametric Comparison of Regression Curves," *Annals of Statistics*, 18, 63–89. [144]

Jung, S., Dryden, I. L., and Marron, J. S. (2012), "Analysis of Principal Nested Spheres," *Biometrika*, 99, 551–568. [144,146]

Kneip, A., and Gasser, T. (1992), "Statistical Tools to Analyze Data Representing a Sample of Curves," *The Annals of Statistics*, 20, 1266–1305. [144]

Kneip, A., and Ramsay, J. O. (2008), "Combining Registration and Fitting for Functional Models, *Journal of American Statistical Association*, 103, 1155–1165. [144]

Liu, X., and Mueller, H. G. (2004), "Functional Convex Averaging and Synchronization for Time-warped Random Curves, *Journal of American Statistical Association*. [144]

Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T., and Cohen, K. L. (1999), "Robust Principal Component Analysis for Functional Data," *Test*, 8, 1–73. [144]

Marron, J. S., and Alonso, A. M. (2014), "Overview of Object Oriented Data Analysis," *Biometrical Journal*, 56, 732–753. [145]

Marron, J. S., Jung, S., and Dryden, I. L. (2010), "Speculation on the Generality of the Backward Stepwise View of PCA," in *Proceedings of MIR 2010: 11th ACM SIGMM International Conference on Multimedia Information Retrieval*, Danvers, MA: Association for Computing Machinery, Inc., pp. 227–230. [146]

Marron, J. S., Ramsay, J. O., Sangalli, L. M., and Srivastava, A. (2015), "Functional Data Analysis of Amplitude and Phase Variation, *Statistical Science*. [144]

Mauras, N., Beck, R., Xing, D., Ruedy, K., Buckingham, B., Tansey, M., White, N. H., Weinzimer, S. A., Tamborlane, W., Kollman, C., and the Diabetes Research in Children Network Study Group. (2012), "A Randomized Clinical Trial to Assess the Efficacy and Safety of Real-time Continuous Glucose Monitoring in the Management of Type 1 Diabetes in Young Children Aged 4 to < 10 Years," *Diabetes Care*, 35, 204–210. [149]

Pizer, S. M., Jung, S., Goswami, D., Vicory, J., Zhao, X., Chaudhuri, R., Damon, J. N., Huckemann, S., and Marron, J. S. (2013), "Nested Sphere Statistics of Skeletal Models," in *Innovations for Shape Analysis*, M. Breuβ, A. Bruckstein, and P. Maragos, eds., Heidelberg: Springer, pp. 93–115. [146]

Ramsay, J. O. (1982), "When the Data are Functions," *Psychometrika*, 47, 379–396. [144]

Ramsay, J. O., and Li, X. (1998), "Curve Registration," *Journal of the Royal Statistical Society*, 60, 351–363. [144]

Ramsay, J. O., and Silverman, B. W. (2002), *Applied Functional Data Analysis: Methods and Case Studies* (Vol. 77), New York: Springer. [144]

——— (2005), *Functional Data Analysis* (2nd ed.), New York: Springer. [144]

Rao, C. R. (1945), "Information and Accuracy Attainable in the Estimation of Statistical Parameters," *Bulletin of Calcutta Mathematical Society*, 37, 81–91. [145]

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2011), "Statistical Analysis and Modeling of Elastic Functions, *arXiv:1103.3817*. [145,146]

Tong, R., and Mueller, H. G. (2008), "Pairwise Curve Synchronization for Functional Data," *Biometrika*, 95, 875–889. [144]

Wang, H., and Marron, J. S. (2007), "Object Oriented Data Analysis: Sets of Trees," *The Annals of Statistics*, 35, 1849–1873. [145]