Comments on: Shape-based functional data analysis

J. E. Borgert[1] · J. S. Marron[1]

Abstract

This discussion paper applauds the authors for their impactful contribution to functional data analysis (FDA). Their primary insight lies in a formal mathematical definition of the "shape" of a curve, which they connect to familiar intuitive notions through a number of examples. Notably, the paper highlights the pitfalls of less well-thought-out curve registration approaches. The authors' application of COVID-19 data enriches the discussion, highlighting the work's practical relevance. We discuss connections of this work with object-oriented data analysis and propose enhancements to the authors' shape-based functional principal component analysis. Additionally, we illustrate the practical significance of adaptive alignment with an example from our own research.

1 Overview

The authors are congratulated on a well-crafted and thought-provoking contribution to the field of functional data analysis (FDA). The fundamental contribution, relative to strong past work by the same group, is the formal definition of the "shape" of a curve in FDA. A number of interesting examples convey how the mathematical definition corresponds well with familiar intuitive notions. These examples also further demonstrate the shortcomings of less carefully devised approaches to curve registration.

Of particular relevance is the authors' application of various forms of COVID-19 data to exemplify key points. This timely use of real-world data adds a layer of impact to their work, emphasizing its relevance and practical utility. In essence, this paper

---

✉ J. E. Borgert
elyseb@live.unc.edu

J. S. Marron
marron@unc.edu

[1] Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, USA

represents a crucial addition to the literature, comprising significant strides in the realm of FDA.

In Section 2, we delve into the connections between this work and concepts from object-oriented data analysis. Section 3 presents a proposal for enhancing the authors' shape-based functional principal component analysis (fPCA). In the final section, we emphasize the practical significance of adaptive alignment, illustrating its application within our own work.

2 Connection to OODA

The authors formalize a novel perspective of functional data analysis (FDA) centered on the idea that the shape of a curve is often the fundamental object of interest in an analysis of functional data. This perspective aligns with object-oriented data analysis (OODA) (Marron and Dryden 2021), a framework for interdisciplinary research that provides new terminology to discuss the often many possible approaches to the analysis of complex data. OODA emphasizes the significance of selecting appropriate *data objects* relevant to a specific scientific inquiry and utilizing methods intrinsic to the corresponding object space.

From an OODA standpoint, the authors' advocacy for shape-based approaches appeals to the concept of *object choice*. As demonstrated in their COVID-19 data example in Section 3.1, the shape of a function often proves to be a more suitable choice in many applications. Their precise definition of shape relies on an equivalence class induced by time warpings of a function. These warpings capture horizontal variation within the set of functions, while warp-equivalent classes distinguish vertical variation. Such variation is discussed in Sections 2.1 and 9.1 of Marron and Dryden (2021) where the terminology *phase* and *amplitude* variation are used, respectively. This shape-based approach not only yields more meaningful data objects but the definitions of amplitude and phase provide collections of curves that are notably richer data objects.

Another aspect to consider regarding data objects is their *representation*. This concept is mirrored in Section 3.5, where the authors explore the use of mode count as an insightful summary of shape. The elastic shape analysis framework defines shape as an equivalence class of functions that are within time warpings of each other. However, an alternative mathematical representation of this class can be extended to a more abstract notion of shape, which focuses on the number of modes in a function while ignoring their heights. In this context, the shape of a function can be characterized by its extreme values in various ways. One definition includes geometric features, such as the heights and counts of peaks and valleys, while ignoring their placements. In another definition, only the extremal points are counted, with heights disregarded. This aligns with OODA insights, highlighting the numerous potential choices for data objects in a shape analysis of functional data.

While the authors primarily treat phase as a nuisance component, this choice depends on the specific context. In some scenarios, phase may be the primary data object of interest, with amplitude as the nuisance component. In other cases, neither amplitude nor phase is treated as a nuisance, and both may be of interest. These various scenarios are detailed in Section 2.1 of Marron et al. (2015).
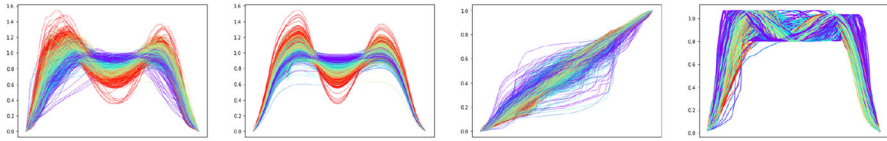
Fig. 1 Left: original data curves depicting the vertical ground reaction force. Middle-left: amplitude objects resulting from fully elastic alignment of the data shown in the left panel. Middle-right: phase objects, represented as warping functions, corresponding to the alignment of the curves in the leftmost panel. Right: an intuitive visualization of phase variation achieved by applying the warping functions from the middle-left panel to the Karcher mean of the amplitude objects. All panels are color-coded based on subjects' walking speed (color figure online)

The careful selection of data objects in an object-oriented data analysis necessitates the use of statistical methodologies that respect the underlying geometry of the chosen object space. A characteristic example of such intrinsic methodology is the *elastic shape analysis* approach that the authors introduce in Section 3.2. This approach provides an elegant solution to the well-known challenges presented by L^2-based methods detailed in Marron et al. (2015).

In Section 4.1, the authors mention a recently developed geometric representation known as the peak persistence diagram (PPD) for estimating the number of peaks in functional data. This diagram is based on an adaptive approach aimed at mitigating the issue of aligning noise artifacts often introduced by fully elastic alignment. The approach involves adjusting a roughness penalty (or elasticity) parameter, denoted as λ, which induces shape changes in the average function. As λ increases, the method smooths geometric features. The optimal λ is determined by analyzing the persistence of internal peaks in the estimated signal function using the PPD. This adaptive peak detection technique follows a *scale space* approach.

As detailed by Lindeberg (1994), scale space was originally a concept explored in the field of computer vision. An approach related to PPDs, developing statistical inference in scale space, is the SiZer (SIgnificance of ZERo crossings) method for smoothing (Chaudhuri and Marron 1999). SiZer overcomes bandwidth selection challenges in kernel density estimation by studying the family of Gaussian kernel smooths that is the scale space. This method visualizes both coarse-scale macroscopic features and fine-scale details, rendering the choice of a single bandwidth irrelevant. In particular, SiZer incorporates statistical inference identifying peaks as regions where the density estimates exhibit increasing followed by decreasing trends, assessed through zero crossings in the derivative of the density estimate. Both the PPD approach and SiZer incorporate scale space visualization tools that graphically represent data features against the resolution.

3 Improving shape fPCA

An important concept of this work is *shape fPCA*, involving a tangent plane approach to the SRVF data objects lying on a unit sphere. In the case of warping functions, these SRVFs must also lie on the positive orthant. Yu et al. (2017) have shown that
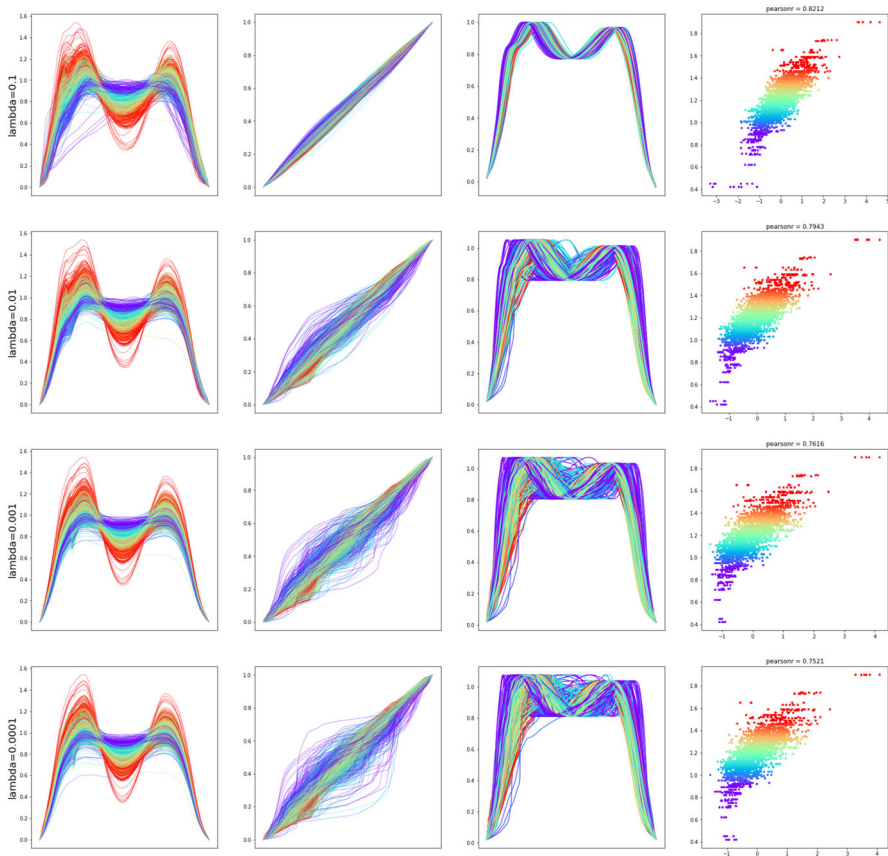
Fig. 2 Alignment results over various values of the elasticity parameter (lambda). Left column: amplitude objects. Middle-left column: warping functions. Middle-right column: warpings of the Karcher mean. Right column: scatter plot of the first principal component score from PCA of amplitude objects versus walking speed. Again, color scheme corresponds to walking speed (color figure online)

in particularly noisy cases, this tangent plane PCA gives a distorted analysis which actually leaves the positive orthant, resulting in invalid warping functions. Similar phenomena were also observed in the data set discussed in the next section. We propose generally adopting in shape fPCA the methodology of Yu et al. (2017), which is based on an improved PCA analog for spheres, *Principal Nested Spheres* (PNS) (Jung et al. 2012). Chapter 8 of Marron and Dryden (2021) provides further review of PNS and other geodesic-based methods.

4 Elastic shape analysis of biomechanical force data

An interesting application of shape analysis is a functional data analysis of human movement data obtained from a study of patients with knee osteoarthritis (OA). Knee

OA is a leading cause of disability among older adults. Biomechanical factors during walking, such as forces, play a central role in the development and progression of knee OA. Ground reaction force (GRF) data collected continuously during gait analysis are often in the form of curves as a function of time. An important observation of the GRF data is that they exhibit phase variation, which motivated the application of elastic shape analysis to this data.

Figure 1 illustrates *fully* elastic shape analysis applied to a set of GRF curves. The color scheme represents the subject's walking speed (red for faster, purple for slower). In that figure, the leftmost panel displays the original vertical GRF curves, which represent the force normal to the ground. The subsequent panel shows aligned curves from a fully elastic registration, capturing amplitude variation. The third panel displays the warping functions, while the rightmost panel provides an intuitive view of phase variation by applying the warps to the Karcher mean. The staircase-like pattern in the warping functions and the difficulty in interpreting the phase variation in the rightmost panel suggest potential over-alignment, prompting the need for adjustment with a roughness penalty.

This example highlights the usefulness of an adaptive approach in choosing the penalty parameter. While peak persistence diagrams were not employed in this analysis, as our objective was not signal recovery, we adopted a similar adaptive approach by varying the elasticity parameter and selecting an optimal alignment based on the interpretability of the phase variation.

The first three columns of Fig. 2 present alignment results for various values of the elasticity parameter (lambda). As lambda increases, the staircase effect smooths out, and interpretable patterns in the phase variation become discernible. In the rightmost column of the same figure, a scatter plot displays the first principal component score from PCA of amplitude objects versus walking speed. Notably, relaxing the alignment results in a stronger association with walking speed, indicating more clinically interpretable amplitude objects. After a detailed exploration of different lambda values, we selected a parameter value of 0.03 for this analysis.

References

Chaudhuri P, Marron JS (1999) SiZer for exploration of structures in curves. J Am Stat Assoc 94:807–823

Jung S, Dryden IL, Marron JS (2012) Analysis of principal nested spheres. Biometrika 99:551–568

Lindeberg T (1994) Scale-space theory in computer vision. Kluwer, Boston

Marron JS, Dryden IL (2021) Object-oriented data analysis. Chapman and Hall/CRC, New York

Marron JS, Ramsay JO, Sangalli LM, Srivastava A (2015) Functional data analysis of amplitude and phase variation. Stat Sci 30:468–484

Yu Q, Lu X, Marron J (2017) Principal nested spheres for time-warped functional data analysis. J Comput Graph Stat 26(1):144–151