Communications for Statistical Applications and Methods
2024, Vol. 31, No. 5, 601–612

https://doi.org/10.29220/CSAM.2024.31.5.601
Print ISSN 2287-7843 / Online ISSN 2383-4757

Functional hierarchical clustering using shape distance

Kyungmin Ahn[1,a]

[a]Department of Statistics, Keimyung University, Korea

Abstract

A functional clustering analysis is a crucial machine learning technique in functional data analysis. Many functional clustering methods have been developed to enhance clustering performance. Moreover, due to the phase variability between functions, elastic functional clustering methods, such as applying the Fisher-Rao metric, which can manage phase variation during clustering, have been developed to improve model performance. However, aligning functions without considering the phase variation can distort functional information because phase variation can be a natural characteristic of functions. Hence, we propose a state-of-the-art functional hierarchical clustering that can manage phase and amplitude variations of functional data. This approach is based on the phase and amplitude separation method using the norm-preserving time warping of functions. Due to its invariance property, this representation provides robust variability for phase and amplitude components of functions and improves clustering performance compared to conventional functional hierarchical clustering models. We demonstrate this framework using simulated and real data.

Keywords: amplitude variation, functional data analysis, functional hierarchical clustering, phase variation, shape distance, SRVF

1. Introduction

Functional clustering analysis is a crucial machine learning technique that assigns functional observations to several groups in functional data analysis (FDA) (Ramsay and Silverman, 2005). The literature has proposed many methods of functional clustering (Bouveyron and Jacques, 2011; Bouveyron *et al.*, 2015; Febrero-Bande and Oviedo de la Fuente, 2012; Marron *et al.*, 2014). A function comprises two components: The phase (horizontal or x-axis) component and amplitude (vertical or y-axis) component. Therefore, verifying whether phase or amplitude variations derive from the natural characteristics of functional objectives or are unknown noise is critical for FDA. If variations are noise, then the noise must be removed using statistical methods. For example, amplitude variations can be minimized using a smoothing technique employing a basis representation, such as the B-spline basis function in a parametric way or the kernel function in a nonparametric way. For phase variation, noise can be removed using a functional statistical method, called a *registration* or *alignment* method. Many registration methods have been proposed, and currently, the Fisher-Rao registration by Srivastava *et al.* (2011) is one of the best alignment methods.

Therefore, many researchers are interested in combining registration and statistical methods to improve the performance of the statistical results. Some papers have been developing joint methods

This research was supported by the Bisa Research Grant of Keimyung University and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2022-00165598).

[1]Corresponding author: Department of Statistics, Keimyung University, 1095 Dalgubeol-daero, Dalseo-gu, Daegu 42601, Korea. E-mail: ahn@kmu.ac.kr

Published 30 September 2024 / journal homepage: http://csam.or.kr

to address phase variation between curves while implementing a statistical method. For example, Ahn *et al.* (2018, 2020) proposed an elastic functional predictor model by reducing phase variation while estimating regression coefficients. Xu *et al.* (2022) proposed a Fisher-Rao registration algorithm to manage phase variation effectively for neural spike data. However, for the statistical analysis, these methods focus on minimizing or removing phase variation assumed to be noise. This strong assumption can be beneficial if the phase variation is indeed noise; if not, this strong conjecture could lead to incorrect statistical results by distorting the natural characteristics of functional data.

To solve this problem, we propose a state-of-the-art method, a functional hierarchical agglomerative clustering method using shape distances (FHACS), which can address phase and amplitude variations for functional hierarchical clustering analysis. This approach is motivated by the phase-amplitude separation algorithm (Srivastava *et al.*, 2011), where invariant metrics, such as the Fisher-Rao metric and the elastic Riemannian metric for functional data in FDA, are employed. One of the substantial advantages of this separation method is that it avoids the pinching effect, where the function can be pinched while aligning in the $\mathbb{L}^2$ space.

The remainder of this paper is organized as follows. Section 2 describes the FHACS under the Fisher-Rao invariant metric and summarizes the concepts used for the clustering method. We demonstrate this method using several simulated datasets and compare its performance against three alternative methods: 1) the standard functional hierarchical clustering model, 2) the pre-aligned functional hierarchical clustering method, and 3) the elastic functional hierarchical clustering model that functions are aligned while computing the distance between two functions in Section 3. Then, we apply the proposed and three candidate methods to six real datasets to compare the effectiveness and clustering performance in Section 4. Last, Section 5 presents the concluding remarks and limitations of this study.

2. Methods

To introduce FHACS, we first summarized the standard functional hierarchical clustering model and some critical concepts for developing the functional clustering model.

2.1. Standard functional hierarchical clustering

The functional hierarchical clustering analysis is the most basic, simple clustering algorithm, which determines the distances between a pair of functional objectives and verifies which cluster should be combined or divided depending on the strategies. We use an agglomerative strategy for the bottom-up approach and a divisive strategy for the top-down approach. There are two critical points to consider when implementing this hierarchical clustering analysis: distance and linkage. Distance is nearly fixed for the FDA. We defined observations as functions in the $\mathbb{L}^2$ space, where the inner product of two functions is defined as $\langle f, g\rangle = \int fg\, dt$ in the Hilbert space, $\mathcal{H}$. Thus, we used the $\mathbb{L}^2$ distance to compute the distance between the functions on the time interval, T, expressed as follows:

$$d(f_1, f_2) = \sqrt{\int_T (f_1(t) - f_2(t))^2\ dt}, \tag{2.1}$$

where f_1 and f_2 are the functional observations in the $\mathbb{L}^2$ space.

For the linkage method, we employed several selection methods for the hierarchical clustering analysis. There are many possible linkages, and some examples are listed below.

- Single linkage: The distance between two clusters is the minimum distance between members of the two clusters.
- Centroid linkage: The distance between two clusters is the distance between their centroids.
- Complete linkage: The distance between two clusters is the maximum distance between members of the two clusters.
- Average linkage: The distance between two clusters is the average of all distances between members of the two clusters.

Each linkage method has advantages and disadvantages; thus, no best universal linkage exists for all situations. For example, the single linkage method can manage nonelliptical shapes and is best for capturing clusters of assorted sizes, but it has a problem with chaining. The complete, centroid, and average linkage methods could group clusters when noise exists between clusters, but crowding, dendrogram inversion, and sensitivity to a monotone transformation can occur, respectively.

2.2. Shape distance

Srivastava *et al.* (2011) proposed a method to decompose functions into phase and amplitude components, which involves the functional registration algorithm in which phase variations can be removed by aligning functions to any function template. The template function can be a random sample function (generally, the pairwise case) in the dataset or the Karcher or Fréchet mean of the functions (generally, the groupwise case). After registration, the functions are split into phase and amplitude components.

To implement this method, we defined functional observation f as a real-valued function on the interval $[0, 1]$ and let $\mathcal{F}$ denote the set of all such functions. Then, we defined the time-warping function, γ, used for computing the phase distance. We let Γ is the set of all boundary-preserving map between manifolds which is differentiable and has a differentiable inverse of the unit interval $[0, 1]$ (i.e., $\Gamma = \{\gamma : [0, 1] \to [0, 1] \mid \gamma(0) = 0, \gamma(1) = 1, \gamma$ is a diffeomorphism$\}$). For alignment or registration, Srivastava *et al.* (2011) established a mathematical expression representing a function using its square root velocity function (SRVF), where q denotes the SRVF of the original function, $q : [0, 1] \to \mathbb{R}$, expressed as follows:

$$q(t) = \text{sign}\left(\dot{f}(t)\right) \sqrt{\left|\dot{f}(t)\right|}, \tag{2.2}$$

where $\dot{f}$ denotes the first derivative of the function f. The greatest advantage is that this framework streamlines the complicated Fisher-Rao Riemannian metric and Fisher-Rao distance into the $\mathbb{L}^2$ inner product and $\mathbb{L}^2$ norm, respectively. Then, a function f can be warped by a time-warping function, γ: $f \longmapsto (f \circ \gamma)$. Hence, its SRVF changes by $q \longmapsto (q \circ \gamma)\sqrt{\dot{\gamma}}$. This transformation is denoted by $(q * \gamma) = (q \circ \gamma)\sqrt{\dot{\gamma}}$ (Ahn *et al.*, 2020), which implies the invariance property of the Fisher-Rao metric, that for any $q_1, q_2 \in \mathbb{L}^2$ and $\gamma \in \Gamma$, we have $\|(q_1 * \gamma) - (q_2 * \gamma)\| = \|q_1 - q_2\|$. Hence, the action of Γ on $\mathbb{L}^2$ is by isometrics. A special case of this equation is $\|(q * \gamma)\| = \|q\|$ for all q and γ. Therefore, this action preserves the $\mathbb{L}^2$ norm of the SRVF and avoids pinching effects in the registration process.

Two registration problems depend on the number of functional observations: The pairwise alignment problem for two observations and the groupwise or multiple alignment problem for multiple observations (Srivastava and Klassen, 2016). Functional hierarchical clustering analysis requires only

the distance between two functions, so only pairwise alignment is necessary. Hence, we summarize the pairwise registration problem.

To align functions, the optimal time-warping function γ^* is determined for any two given functions $f_1, f_2 \in \mathcal{F}$, such that the equation can be expressed as follows:

$$\gamma^* = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \|f_1 - f_2 \circ \gamma\|^2 . \tag{2.3}$$

Then, for any $t \in [0, 1]$, the value $f_1(t)$ is registered to $f_2(\gamma^*(t))$. In here, many potential problems can be raised such as the invariance problem and pinching effect of using equation 2.3 in the $\mathbb{L}^2$ space. Thus, we applied SRVF, q, rather than f to align the functions successfully to solve any potential problems. Then, equation 2.3 can be changed as follows:

$$\gamma^* = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \|q_1 - (q_2 * \gamma)\|^2 , \tag{2.4}$$

where q_1 and q_2 are the SRVFs of function f_1 and f_2, respectively. We employed equation 2.4 to compute the shape distances for FHACS.

2.2.1. Amplitude distance

Next, we defined the amplitude distance. For any two functions $f_1, f_2 \in \mathcal{F}$ and the corresponding SRVFs, $q_1, q_2 \in \mathbb{L}^2$, we defined the amplitude distance d_a (Srivastava and Klassen, 2016; Tucker *et al.*, 2013; Ahn *et al.*, 2020) as follows:

$$d_a(f_1, f_2) = \inf_{\gamma \in \Gamma} (\|q_1 - (q_2 * \gamma)\|) . \tag{2.5}$$

In here, the distance d_a is a proper distance with three properties: symmetry, positive definiteness, and triangle inequality (for more details, see Srivastava and Klassen (2016)).

2.2.2. Phase distance

We let $\gamma_1, \gamma_2, \ldots, \gamma_n \in \Gamma$ represent the set of observed warping functions. However, computing the phase distance in a standard way has two central problems : 1) the infinite dimensionality and 2) Γ is a nonlinear manifold (Tucker *et al.*, 2013). To solve these issues, similar to the definition of SRVF, we need to represent an element of $\gamma \in \Gamma$ using the square root of its derivative, $\sqrt{\dot{\gamma}}$ that this representation simplifies the complicated geometry of Γ to a unit sphere. That is, $\| \sqrt{\dot{\gamma}}\|^2 = \int_0^1 \dot{\gamma}(t)\, dt = \gamma(1) - \gamma(0)$. Hence, the phase distance for the two functions is as follows:

$$d_p(f_1, f_2) = \cos^{-1} \int_0^1 \sqrt{\dot{\gamma}^*(t)}\, dt, \tag{2.6}$$

where $\gamma^*(t)$ is the optimal time-warping function from equation 2.4. This definition is based on the Fisher–Rao distance using SRVFs (Srivastava and Klassen, 2016; Tucker *et al.*, 2013; Ahn *et al.*, 2020), following the properties below:

1. $d_p(f_1, f_2) = d_p(f_2, f_1)$;

2. $d_p(f_1, f_2) = d_p(f_1, af_2 + b)$ for any $a \in \mathbb{R}^+$ and $b \in \mathbb{R}$;

3. $d_p(f_1, f_3) \geq d_p(f_1, f_2) + d_p(f_2, f_3)$ for any $f_1, f_2, f_3 \in \mathcal{F}$.

2.2.3. Shape distance, combining the phase and amplitude distances

Based on two distances, as discussed, we defined the shape distance d_s by combining two distances, the phase and amplitude distances, as presented below:

$$d_s(f_1, f_2) = \lambda d_a(f_1, f_2) + (1 - \lambda) d_p(f_1, f_2), \tag{2.7}$$

where $\lambda \in [0, 1]$ is a proportion tuning parameter. We can control both distances by weighting each distance; for example, if $\lambda = 0$, the shape distance is focused on the phase distance $d_p(\cdot)$. If $\lambda = 1$, the shape distance can be simplified as the amplitude distance $d_a(\cdot)$. Moreover, if $\lambda = 0.5$, the shape distance is the average of two distances: $(1/2)(d_a(\cdot) + d_p(\cdot))$. For FHACS, this shape distance is applied instead of the $\mathbb{L}^2$ distance to increase the efficiency and performance of the clustering analysis by addressing the phase and amplitude variations.

3. Simulation studies

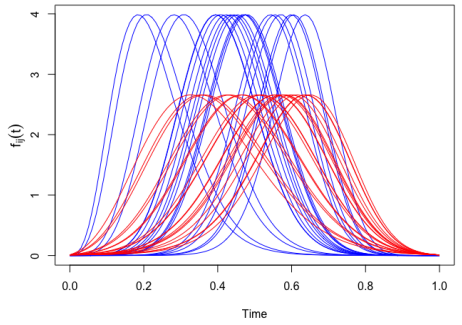
3.1. Simulation setting

To assess the performance of the functional hierarchical clustering algorithm using the shape distance, we evaluated the model on simulated data constructed using the following:

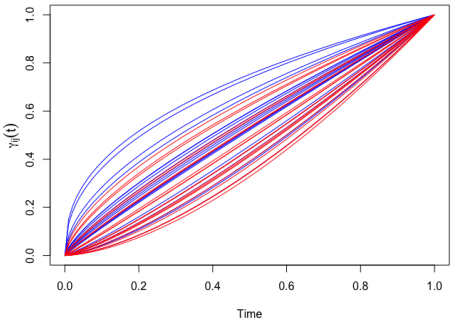
$$f_{ij}^0(t) = a_{ij} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(t - \mu_j)^2}{2\sigma_j^2}\right), \quad i = 1, \ldots, 20, \;\; j = 1, \ldots, K, \tag{3.1}$$

where μ_j and σ_j denote the mean and standard deviation for each j group, respectively. In addition, a_{ij} represents the parameter for the amplitude variation and $t \in [0, 1]$. Using equation 3.1, we generated three simulated datasets, where each data point reflects the situation of generating clusters: 1) two groups/clusters ($K = 2$) are based on the amplitude ($\mu_j \in \{0.5, 0.5\}$, $\sigma_j \in \{0.1, 0.11\}$, and $a_{ij} = 1$ for all i, j), 2) two groups ($K = 2$) are based on the phase ($\mu_j \in \{0.35, 0.65\}$, $\sigma_j \in \{0.1, 0.1\}$, and $a_{ij} = 1$ for all i, j), and 3) three groups ($K = 3$) are based on both the phase and amplitude ($\mu_j \in \{0.4, 0.5, 0.6\}$ and $\sigma_j \in \{0.1, 0.11, 0.13\}$ with $a_{ij} \sim N(1, 0.1^2)$). Then, 20 functions are generated for each group j. Therefore, the 40, 40, and 60 functions are generated for simulated Datasets 1, 2, and 3, respectively. These given functions are perturbed using random time-warping functions $\{\gamma_{ij}\}$ to obtain the observed functions ($f_{ij} = f_{ij}^0 \circ \gamma_{ij}$).

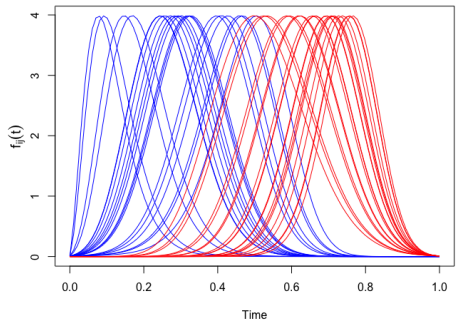
Figure 1 presents the resulting simulated data using equation 3.1. Figure 1(a), Figure 1(c), and Figure 1(e) show the simulated data that were perturbed by corresponding time-warping functions in Figure 1(b), Figure 1(d), and Fig. 1f, respectively. We note that the warped functions of simulated datasets have values of zero at $t = 0$ and $t = 1$. Since the functions are warped by the equation, $f_{ij}(t) = f_{ij}^0(\gamma_{ij}(t))$ where $\gamma_{ij}(0) = 0$ and $\gamma_{ij}(1) = 1$, the warped functions also should be $f_{ij}(0) = 0$ and $f_{ij}(1) = 0$. Figure 1(a) displays simulated Dataset 1 where the true clusters are based on an amplitude with two clusters. The functions have phase variations that are noise. In addition, Figure 1(c) displays simulated Dataset 2, where the true clusters are based on a phase with two clusters. Unlike simulated Dataset 1, functions have phase variation as noise, but two clusters are also distinguished by various phases.



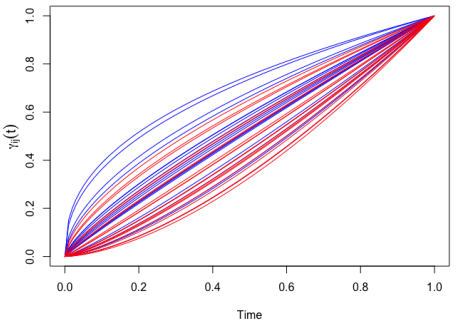
(a) *Simulated Dataset 1 (amplitude-based clusters)*



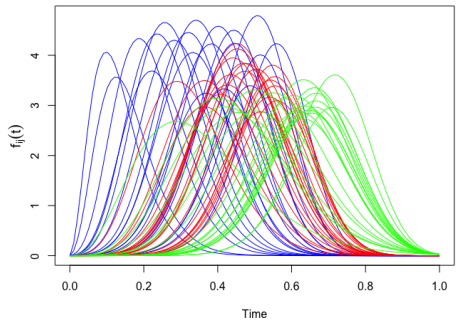
(b) *Corresponding time-warping functions, $\gamma_{ij}(t)$*



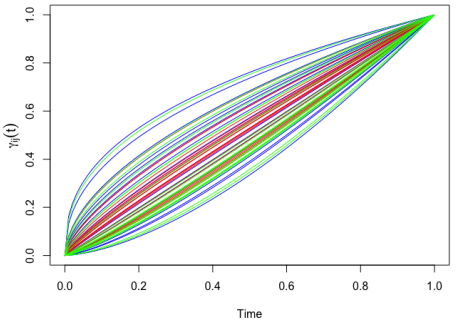
(c) *Simulated Dataset 2 (phase-based clusters)*



(d) *Corresponding time-warping functions, $\gamma_{ij}(t)$*



(e) *Simulated Dataset 3 (amplitude and phase-based clusters)*



(f) *Corresponding time-warping functions, $\gamma_{ij}(t)$*

Figure 1: *Plots for simulated data. The functions are generated using equation 3.1 and then perturbed using random time-warping functions $\{\gamma_{ij}\}$ for each simulated dataset, i.e., (b), (d), (f) for simulated Dataset 1, 2, and 3, respectively. The clusters in each dataset are distinguished by color to mark the true number of clusters. For example, simulated Datasets 1 and 2 have two clusters, whereas simulated Dataset 3 has three clusters.*

Therefore, removing the phase variations without consideration might provide incorrect results. Last, Figure 1(e) presents three groups where the true clusters are based on the phase and amplitude, which is a more complicated case when functions have both phase and amplitude variations as noise.

To demonstrate the effectiveness and clustering performance of the model, we compared FHACS with three natural alternatives. These models are either commonly used in the literature or are modifications of current models addressing phase variability in functional data. These models are the standard functional hierarchical agglomerative clustering model (SFHAC), the pre-aligned functional hierarchical agglomerative clustering model (PAFHAC), and the elastic functional hierarchical agglomerative clustering model (EFHAC). We summarize these three clustering models.

- SFHAC: This standard hierarchical agglomerative clustering (HAC) algorithm uses the $\mathbb{L}^2$ distance instead of the Euclidean distance. The functions are defined in the $\mathbb{L}^2$ space; thus, the $\mathbb{L}^2$ metric is applied for the statistical clustering analysis. Hence, equation 2.1 calculates the distance between functional observations.

- PAFHAC: In this model, functional observations are pre-aligned (using a phase-amplitude separation algorithm) (for more details, refer to Srivastava and Klassen (2016)), and then applied the standard hierarchical clustering model. Phase variations are removed by aligning the functions concerning their Karcher mean; hence, this is a registration problem for the groupwise case.

- EFHAC: In this model, functions are aligned pairwise while computing the distance between two functions. The concept of the model is similar to that of FHACS, but this model only focuses on the amplitude distance between two functions. The difference between PAFHAC and EFHAC is that PAFHAC computes the Karcher mean of the functions, registers all functions concerning the mean, and applies SFHAC. However, EFHAC computes the amplitude distance while removing the phase distance between two functions each time. Hence, the clustering results should be the same as when λ of FHACS equals 1 in equation 2.7 (i.e., $d_s(f_1, f_2) = d_a(f_1, f_2)$).

We applied FHACS and the other three alternative clustering algorithms to compare the accuracy of clustering performance. These models can easily be implemented using the existing `R` packages from `R` software. We used the `fdacluster` (Stamm, 2024) `R` package for SFHAC and EFHAC. For PAFHAC, we aligned multiple functions based on the phase-amplitude separation algorithm using the `fdasrvf` `R` package (Tucker, 2024). We employed a dynamic programming algorithm to determine the optimal time-warping functions, $\{\gamma_{ij}^*\}$. For FHACS, we also used the `fdasrvf` `R` package (Tucker, 2024) to compute the phase and amplitude distances. Then, we used equation 2.7 to compute the shape distance. We fixed all models to the HAC algorithm with complete linkage for consistency.

3.2. Clustering performance accuracy

To compare the clustering performance as a measurement, we calculated the Hubert–Arabie adjusted random index (ARI), which is the corrected-for-chance version of the rand index (Rand, 1971; Hubert and Arabie, 1985; Vinh *et al.*, 2009), and is computed below:

$$\text{ARI} = \frac{\text{Index} - \text{Expected Index}}{\text{Max Index} - \text{Expected Index}},$$

where the independent clusters have an expected index of zero and identical partitions have an ARI equal to 1.

Table 1: FHACS results for each simulated dataset

λ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Sim. Dataset 1	0.03	0.031	0.031	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Sim. Dataset 2	**1**	**1**	0.03	0.03	0.03	0.03	0.15	0.15	0.11	0.11	0.03
Sim. Dataset 3	0.24	0.13	0.15	**0.57**	0.35	0.26	0.26	0.30	0.29	0.26	0.27

ARI is computed for each λ, and bold font marks the highest ARI.

Table 2: Three alternative clustering results for each simulated dataset

	SFHAC	PAFHAC	EFHAC
Sim. Dataset 1	−0.01	**1**	**1**
Sim. Dataset 2	0.03	0.01	0.03
Sim. Dataset 3	0.23	0.11	0.27

Bold font marks the highest ARI.

For the simulation experiments in the clustering analysis, we directly specified the true number of clusters K ($K = 2$ for simulated Datasets 1 and 2, and $K = 3$ for simulated Dataset 3) and evaluated the accuracy of the clustering results. We used the `mclust` (Scrucca *et al.*, 2016) `R` package for the ARI calculation.

3.3. Simulation results

Tables 1 and 2 list the clustering results for the simulation studies. For Table 1, the ARI is computed for each λ to quantify the shape distance in equation 2.7. Based on the results of the tables, λ can measure the weight between the amplitude and phase distances properly for the simulated data. For example, clusters in simulated Dataset 1 were generated based on the amplitude; thus, phase variations are true noise in this case. The model yields the best clustering results when λ is increasing; therefore, the parameter weight is focused on the amplitude distance in equation 2.7.

Similarly, PAFHAC and EFHAC also performed well because these two models are clustering models that focus on the amplitude distance after removing phase variation. The groups from simulated Dataset 2 were generated based on two distinct phases, and the results reveal that parameter $\lambda = 0$ for FHACS has the best clustering performance. Unlike the performance of PAFHAC and EFHAC for simulated Dataset 1, the accuracy of the two models was extremely poor because the model only measures the amplitude distance. When the groups are based on phase and amplitude, as in simulated Dataset 3, the tuning parameter λ was set at 0.3 for the best ARI. These simulation results demonstrate that the parameter λ can adequately adjust the weight of two distances. That is, λ plays a critical role in weighting amplitude and phase distances to improve clustering performance.

4. Application to real data

4.1. Description of real data

Functional data have phase and amplitude variability in several crucial application areas, such as biology, human anatomy, biochemistry, images, and sensors. This section presents the results on multiple real datasets for clustering methods using six examples from the University of California at Riverside time-series classification archive (Chen *et al.*, 2015, July). This database has information, including the actual number of clusters and time points. Table 3 describes the data (for more details, refer to Chen *et al.* (2015, July)).

Figure 2 presents the functional data for each real dataset: Gun Point (GP; Figure 2(a)), Toe

Table 3: Description of the real data

Data	# of obs.,	Time points	Class	Donor
Gun Point (GP)	50	150	2	A. Ratanamahatana & E. Keogh
Toe Segmentation (TS)	40	270	2	Tony Bagnall
Face Four (FF)	24	350	4	A. Ratanamahatana & E. Keogh
Cylinder-Bell-Funnel (CBF)	30	128	3	N. Saito
Italy Power Demand (IPD)	27	64	2	J.J. van Wijk & E. Keogh & L. Wi
Plane	105	144	7	J. Gao

Segmentation (TS; Figure 2(b)), Face Four (FF; Figure 2(c)), Cylinder-Bell-Funnel (CBF; Figure 2(d)), Italy Power Demand (IPD; Figure 2(e)), and Plane (Figure 2(f)). Each figure is colored to distinguish the true number of clusters (2, 2, 4, 3, 2, and 7 for GP, TS, FF, CBF, IPD, and Plane, respectively). From the figures, the phase and amplitude variability values are naturally found in the data, as the observations are at different times for each measurement during collection.

4.2. Numerical results

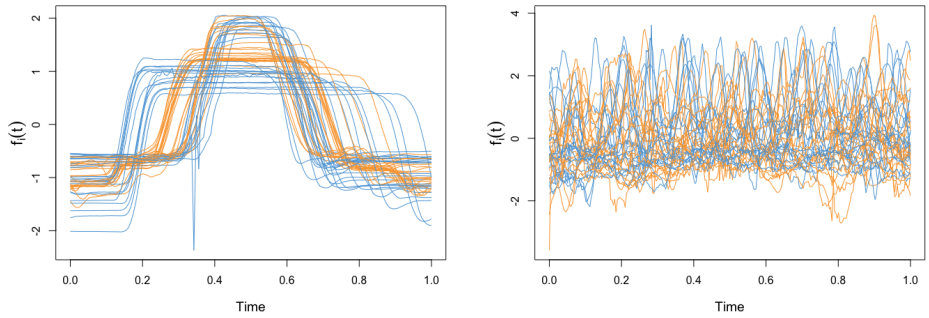
Next, we applied FHACS and the other three alternative hierarchical clustering methods to these real datasets. Similar to the simulation study, for FHACS, we set $\lambda = \{0, 0.1, 0.2, \ldots, 0.9, 1\}$ to determine the optimal tuning parameter for measuring the weight between the phase and amplitude distances. Table 4 lists the clustering performance results for FHACS, and Table 5 lists the results for the three alternative models.

In Tables 4 and 5, FHACS provides better clustering performance than the other three competing models. The clustering results in tables imply that the phase distance is also critical to classify groups of functional data because most λ values are near 0 instead of 1. The clustering performance of TS, FF, and Plane datasets especially outperformed when λ is equal to 0. Therefore, functional clustering models such as EFHAC and PAFHAC that remove only phase variation without considering functional data perform poorly on real data, implying that in many possible existing real datasets, phase variation is from natural characteristics of functional observations. That is, FHACS performs better when focused on both phase and amplitude distances. Like PAFHAC and EFHAC, focusing only on amplitude distance when removing phase variation produces poor results because this method ignores phase variation while analyzing functional data.

5. Concluding remarks and limitations

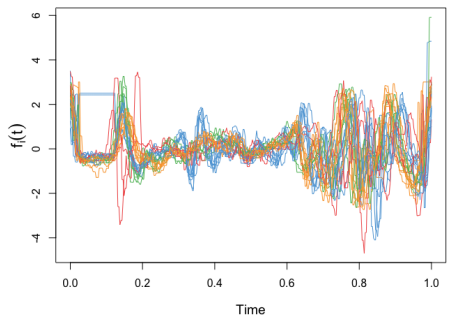
The functional clustering algorithm was developed for the FDA to improve clustering performance. Many functional data have phase variability due to the lack of temporal synchronization across measurements. Therefore, awareness has been rising regarding phase variability removal to improve model clustering performance. However, removing or minimizing only the phase variation can be problematic because functional data can also involve phase characteristics. Hence, we proposed a state-of-the-art functional hierarchical clustering method that can measure both the phase and amplitude distances, defined as the shape distance. We applied this hierarchical clustering model to three simulated datasets to investigate clustering performance in different functional data situations, and we applied it to six real datasets. The proposed method outperforms other alternative three functional hierarchical clustering models. Moreover, this definition of the shape distance can be applied to other clustering or classification methods, such as k-means and k-nearest neighbors, to address the phase and amplitude components of functional data.

This study has some limitations. First, functional or multivariate standard hierarchical clustering

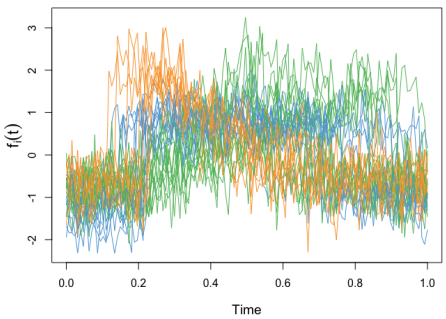


(a) *Gun point (GP)*

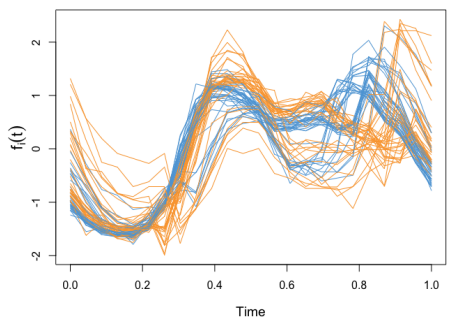
(b) *Toe segmentation (TS)*



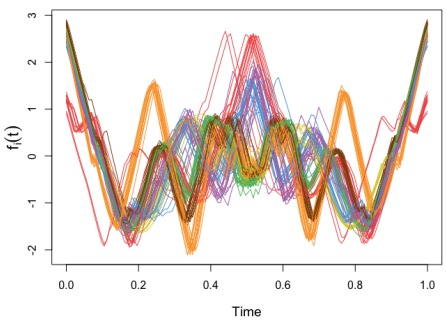
(c) *Face four (FF)*



(d) *Cylinder-bell-funnel (CBF)*



(e) *Italy power demand (IPD)*



(f) *Plane*

Figure 2: *Plots for real data. Each figure is colored differently to distinguish the true number of clusters.*

suffers from choosing the optimal number of clusters, K, unlike other clustering algorithms, such as k-means. Thus, it is challenging to determine which K is the optimal cluster for the hierarchical clustering analysis. Moreover, we added a new tuning parameter, λ, to weight the phase and amplitude distances; therefore, determining the optimal K and λ can be more complicated. However, this can be

Table 4: Clustering performance results for the real data

λ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
GP	0.001	0.179	**0.047**	−0.003	0.003	0.003	0.003	−0.003	−0.003	−0.003	−0.003
TS	**0.109**	0.100	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
FF	**0.659**	0.514	0.514	0.514	0.514	0.444	0.179	0.179	0.213	0.240	0.240
CBF	0.028	0.059	0.020	0.059	0.091	0.059	**0.241**	0.110	0.110	0.145	0.145
IPD	0.110	0.110	0.110	**0.573**	0.110	−0.008	0.110	0.004	0.004	0.004	−0.008
Plane	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0.806	0.806	0.806

Bold represents the highest ARI among the four clustering algorithms.

Table 5: Results for real data using the three alternative clustering algorithms

	SFHAC	PAFHAC	EFHAC
GP	0.011	0.029	−0.003
TS	−0.018	0.015	0.015
FF	0.369	−0.021	0.240
CBF	0.137	0.196	0.145
IPD	0.031	−0.008	−0.008
Plane	0.623	0.701	0.806

solved with domain knowledge for the functional data, for example, with the actual number of clusters. Selecting a specific λ to improve the clustering performance gradually is another computational problem. In this study, we set only one λ for one clustering analysis. The functional hierarchical clustering algorithm computes the distance between two objectives; hence, each pair of functions is likely to have a different optimal λ value. Clustering performance can be improved further if each λ_l is set to $l = \{1, 2, \ldots, \binom{n}{2}\}$ for n functional observations. However, this is nearly impossible due to the computational costs.

References

Ahn K, Tucker JD, Wu W, and Srivastava A (2018). Elastic handling of predictor phase in functional regression models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, UT, 324–331.

Ahn K, Tucker JD, Wu W, and Srivastava A (2020). Regression models using shapes of functions as predictors, *Computational Statistics & Data Analysis*, **151**, 107017.

Bouveyron C, Côme E, and Jacques J (2015). The discriminative functional mixture model for a comparative analysis of bike sharing systems, *The Annals of Applied Statistics*, **9**, 1726–1760.

Bouveyron C and Jacques J (2011). Model-based clustering of time series in group-specific functional subspaces, *Advances in Data Analysis and Classification*, **5**, 281–300.

Chen Y, Keogh E, Hu B, Begum N, Bagnall A, Mueen A, and Batista G (2015, July). The UCR Time Series Classification Archive, Available from: www.cs.ucr.edu/ eamonn/time_series_data/

Febrero-Bande M and M Oviedo de la Fuente (2012). Statistical computing in functional data analysis: The R package fda. usc, *Journal of Statistical Software*, **51**, 1–28.

Hubert L and Arabie P (1985). Comparing partitions, *Journal of Classification*, **2**, 193–218.

Marron JS, Ramsay JO, Sangalli LM, and Srivastava A (2014). Statistics of time warpings and phase variations, *Electronic Journal of Statistics*, **8**, 1697–1702.

Ramsay JO and Silverman BW (2005). *Functional Data Analysis* (2nd ed), Springer, Berlin.

Rand WM (1971). Objective criteria for the evaluation of clustering methods, *Journal of the American*

Statistical Association, **66**, 846–850.

Scrucca L, Fop M, Murphy TB, and Raftery AE (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models, *The R Journal*, **8**, 289.

Srivastava A and Klassen E (2016). *Functional and Shape Data Analysis*, Springer, Berlin.

Srivastava A, Wu W, Kurtek S, Klassen E, and Marron JS (2011). Registration of functional data using Fisher-Rao metric, Available from: arXiv:1103.3817

Stamm A (2024). *fdacluster: Joint Clustering and Alignment of Functional Data*, R package version 0.3.0.9000, Available from: https://github.com/astamm/fdacluster

Tucker JD (2024). *fdasrvf: Elastic Functional Data Analysis*. R package version 2.2.0, Available from: https://github.com/jdtuck/fdasrvf_R

Tucker JD, Wu W, and Srivastava A (2013). Generative models for functional data using phase and amplitude separation, *Computational Statistics & Data Analysis*, **61**, 50–66.

Vinh NX, J Epps, and J Bailey (2009). Information theoretic measures for clusterings comparison: Is a correction for chance necessary?, In *Proceedings of the 26th Annual International Conference on Machine Learning*, 1073–1080.

Xu Z, Zhou X, Xu Y, and Wu W (2022). Removing nonlinear misalignment in neuronal spike trains using the Fisher-Rao registration framework, *Journal of Neuroscience Methods*, **367**, 109436.

Received July 03, 2024; Revised August 02, 2024; Accepted August 03, 2024