



# Comments on: shape-based functional data analysis

Almond Stöcker<sup>1</sup> · Lisa Steyer<sup>2</sup> · Sonja Greven<sup>2</sup> 

Received: 16 October 2023 / Accepted: 21 October 2023 / Published online: 28 November 2023  
© The Author(s) 2023

**Mathematics Subject Classification** 62R10 · 62R20 · 62R30

## 1 Introduction

We would first like to thank the authors, Yuexuan Wu, Chao Huang and Anuj Srivastava, for an interesting article, as well as for the relevant work, in particular on the square-root-velocity (SRV) framework, that it summarizes, and which we have found very useful in our own work. We consider the SRV framework a milestone in object data analysis.

While many aspects in the article are worthy of attention, we focus our discussion in the following on several aspects and challenges we have encountered in our work in this field. In particular, we will discuss the univariate versus the multivariate case in Sect. 2, the problem of sparsely sampled curves in Sect. 3, and regression respecting invariances in Sect. 4, before concluding with a discussion in Sect. 5.

## 2 The uni- and the multivariate case

### 2.1 How much invariance?

For an introduction to the SRV framework, it is understandable that Wu, Huang and Srivastava restrict their discussion to the univariate case, considering functions that

---

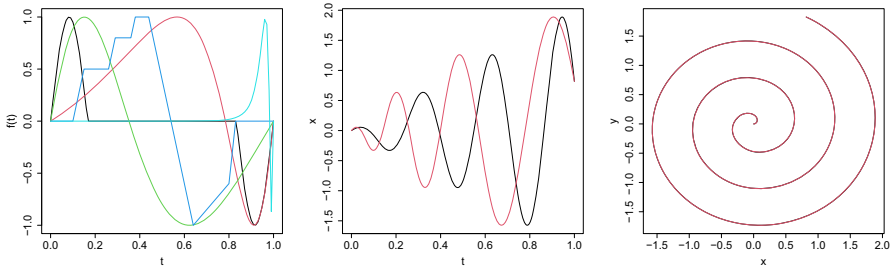
✉ Sonja Greven  
sonja.greven@hu-berlin.de

Almond Stöcker  
almond.stoecker@epfl.ch

Lisa Steyer  
lisa.steyer@hu-berlin.de

<sup>1</sup> Department of Mathematics, EPFL, 1015 Lausanne, Switzerland

<sup>2</sup> Chair of Statistics, School of Business and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany



**Fig. 1** Different representatives of the same equivalence class  $[f]$  under warping: for a univariate function  $f$  (left) and for a bivariate function  $f$ , depicted are  $x$ -coordinate (middle) and  $x$ - versus  $y$ -coordinate (right, identical functions overlaid)

map from an interval  $I$  to the  $\mathbb{R}^1$ . Nevertheless, in our experience it is useful to keep the multivariate case also in mind, where functions map into the  $\mathbb{R}^d$  for some  $d > 1$ . The reason is that the uni- and multivariate cases may differ in the kinds and amounts of invariances that are of interest. In the multivariate case, the equivalence class  $[f]$  of a function  $f : I \rightarrow \mathbb{R}^d$  with respect to reparametrization describes the image of that function (cf. Fig. 1, right), which is often of interest. For example, when considering the outline of an object, such as that of a structure in the brain on a scan (e.g., the hippocampus, Steyer et al. 2023a), parametrization of this outline as a parametrized curve is typically arbitrary and not of interest and the image of the curve,  $[f]$ , is the natural object of analysis. Similarly, if for a movement path, such as a human or animal movement, or handwritten letters or symbols (Steyer et al. 2023b), only the image is recorded or the exact timing not of interest,  $[f]$  is considered for the analysis. Note that in the multivariate case, often additional invariances are of interest, in particular, the shape invariances discussed below in Sect. 2.2.

For the univariate case, on the other hand, considering the equivalence class  $[f]$  leaves only information on the minima and maxima of  $f$ , as also pointed out by the authors of the paper. That is, each such equivalence class can be chosen to be represented by a series of triangles with straight lines joining the extrema (piecewise linear function) (Lahiri et al. 2015) or by a merge tree (Pegoraro and Secchi 2021). Or in other words, almost all information is considered to be phase variation and removed, leaving little in terms of amplitude information, only the sequence and values of minima and maxima. See Fig. 1 for an example of different functions that are considered equivalent in this framework, contrasted with a bivariate example. We believe that in the univariate case it needs to be carefully considered whether removing all phase information is what is wanted. Often despite potential misalignment, some information about timing is still of interest. For instance, in the example of COVID waves in different countries used as a running example by Wu, Huang and Srivastava, how are countries to be compared that had 2, 3 and 4 waves of different heights during the same time span? For instance, if country A skipped the second wave in contrast to country B, but the third wave was of similar height to the second wave in country B, should that wave be matched to the second wave of country B or to the third wave in country B that took place roughly simultaneously? Also, is the length of waves not relevant in terms of, e.g., number of deaths? In such cases timing may not

be completely irrelevant and a decomposition of variation into phase and amplitude variation and subsequent separate or joint analysis (e.g., Hadjipantelis et al. 2015; Happ et al. 2019) may be more natural than completely removing phase variation. For this, of course, the SRV functional data framework is still very useful, even though no alignment method is a panacea in settings with different characteristics such as the different numbers of waves in the COVID data. Also note that while consideration of  $[f]$  removes all phase variation, in practice aligned representatives (such as of all functions aligned to an overall mean) are typically chosen such that warping functions are the identity on average, still keeping some average phase information in the data.

Finally, we would like to point out that an analysis of  $[f]$  in the univariate case is very hard to generalize to functions that are only sparsely observed and possibly with error, as discussed in more detail in Sects. 3 and 5. In this setting, alignment is already difficult, but considering only minima and maxima of a function is especially hard both due to the error, which leads to noisy estimates of extrema, as well as due to the sparsity, which can lead to extrema not being observed at all.

## 2.2 Reparametrization and shape invariances

As the authors explain in their paper, it might be of interest to consider curves (in particular in 2D or 3D) not only invariant under reparametrization, but also under other invariances: In *elastic shape analysis* (Srivastava and Klassen 2016) reparametrization invariant, so-called *elastic* analysis is combined with ideas of *statistical shape analysis* (Dryden and Mardia 2016) where the *shape* of a geometric object is traditionally defined as its equivalence class under rotation, translation and scale. For instance, if a curve in 2D describes the outline of an object such as the hippocampus (e.g., Steyer et al. 2023a), the parametrization along the outline is arbitrary and an elastic analysis is in order. In addition, (parts of) the coordinate system (e.g., based on the orientation of the scanner) may also be arbitrary and invariances with respect to translation, rotation and potentially scale should then be taken into account as well. Depending on the data problem at hand, also different subsets of these invariances might be relevant. For instance, for movement trajectories in 2D or 3D (e.g., Steyer et al. 2023b) the temporal information may actually be relevant, or for an outline a natural parametrization may be given (e.g., Stöcker et al. 2023) and thus only the shape invariances rotation, translation and possibly scale should be taken into account. Depending on the invariances, different metrics and structures arise on the resulting quotient spaces. Invariance modulo scaling will yield a Riemannian manifold, which is also the case for rotation in 2D, while invariance with respect to 3D rotation produces singularities. For invariance under reparametrization no manifold structure is available either.

Accordingly, reparametrization invariance and/or other invariances may need to be considered and the corresponding choice carefully discussed. While the authors of the paper refer to the equivalence class  $[f]$  under warping as ‘shape’ of  $f$ , we thus think it useful to distinguish the terminology ‘elastic’ and ‘warping/reparametrization invariances’ on the one hand from ‘shape invariances’ (scale, rotation, translation) on the other hand to avoid confusion, in particular in the multivariate setting. Otherwise, we believe it helpful to make very explicit which invariances are meant in case of any

doubt, such as done by Huckemann et al. (2010), distinguishing ‘similarity shapes’, ‘affine shapes’ and ‘projective shapes’.

### 3 Elastic analysis of sparsely sampled curves

In practice, a function  $f : [0, 1] \rightarrow \mathbb{R}^d$  is usually not directly observed but only evaluated on a (potentially curve-specific) grid  $t_0 < \dots < t_m$  as  $\mathbf{f}_j = f(t_j)$ ,  $j = 0, \dots, m$ . In some scenarios, functions are sampled densely enough that one might, nonetheless, approximately think of  $f$  as fully observed. Other scenarios, by contrast, demand for explicit consideration of *sparsely* sampled functions. This issue, well known in functional data analysis, is similarly relevant in elastic analysis of curves but particularly challenging.

In the context of elastic analysis of sparsely observed curves, there are two bad news and two good news. The bad news are:  $\mathbf{A}^-$  SRV representation crucially relies on derivatives  $\dot{f}(t_j)$ , which are not observed and for sparse data not approximated well by finite differences  $\Delta_j = (\mathbf{f}_j - \mathbf{f}_{j-1})/(t_j - t_{j-1})$ , say.  $\mathbf{B}^-$  even given the  $\dot{f}(t_j)$ , an optimal warping function  $\gamma$  between  $f$  and another function  $g : [0, 1] \rightarrow \mathbb{R}^d$  can only be approximated by some  $\hat{\gamma} \approx \gamma$  based on discrete evaluations, since optimal warping is not identifiable on  $t_0, \dots, t_m$ . Hence, also the elastic distance  $d_s(f, g)$  of the curves modulo warping can only be approximated. For a sample of curves  $f_1, \dots, f_n$ , the miss-match between  $f_i \circ \hat{\gamma}_i$  and  $f_i \circ \gamma_i$  aligned to some  $g$  is specific to each curve and will not improve or ‘average out’ when the number of samples  $n$  increases, in general. Hence, in any statistical method based on the alignment, such as Fréchet mean computation, it seems inevitable that some bias is induced by sparsity. This is in contrast to the case of classic functional data, which allows a point-wise perspective that is helpful in the case of sparse observations, as it allows borrowing of strength across curves, such that bias can vanish with increasing sample size. The good news are that:  $\mathbf{A}^+$  at least for  $d = 1$  and  $d = 2$  dimensional curves, the two problems can, in fact, be reduced to the second issue  $\mathbf{B}^-$ . For  $d = 1$  and if  $f$  is differentiable between the  $t_j$ , there is an  $f^* \in [f]$  such that  $\dot{f}^*(t_j^*) = \Delta_j$ ,  $j = 1, \dots, m$ , for any chosen points  $0 < t_1^* < \dots < t_m^* < 1$ , and we can simply assume that the  $\Delta_j$  directly present observations of the derivative. This directly follows from the mean value theorem, which yields  $t_{j-1} < \xi_j < t_j$  such that  $\dot{f}(\xi_j) = \Delta_j$  for  $j = 1, \dots, m$  and lets us choose  $f^* = f \circ \gamma$  with a warping function  $\gamma$  satisfying  $\xi_j = \gamma(t_j^*)$  and  $\dot{\gamma}(t_j^*) = 1$ . For  $d = 2$ , Stöcker et al. (2022) show that the same holds under the additional assumption that all loops and corners of the curve described by  $f$  are contained in the sample points  $\mathbf{f}_0, \dots, \mathbf{f}_m$ . While similarly intuitive, the proof for  $d = 2$  is more involved. For  $d = 3$ , there is no corresponding result and it is, in fact, easy to find a counterexample (say, a helix).  $\mathbf{B}^+$  While a certain bias has to be expected due to imperfect alignment, it quickly becomes very small if the sample points  $\mathbf{f}_j$  reflect the shape of the curve reasonably well and points  $t_1^*, \dots, t_m^*$  in the initial parametrization of the curves are selected reasonably (e.g., with respect to arc length of the sample polygon with corners  $\mathbf{f}_j$ ) (Steyer et al. 2023a, b; Stöcker et al. 2022). In fact, we were repeatedly positively surprised about the quality of the results even for small  $m$ . While Steyer et al. (2023a, b) (widely) treat observed curves as polygons, and propose estimators for Fréchet means,

elastic distances and regression models in this context, Stöcker et al. (2022) also use the covariance structure for prediction of distances and inner products. However, more work is required in this direction, and our results indicating fast bias decay and good practical behavior are purely empirical so far.

## 4 Regression

The authors address the problem of regression with equivalence classes  $[f]$  of curves modulo warping as data objects, an important and interesting topic, which is, however, particularly challenging when working with  $[f]$  explicitly instead of fixing pre-determined representatives  $f$ . They consider three cases: the case where  $[f]$  has the role of the response in a model with scalar covariates, the case where  $[f]$  takes the role of a covariate in a regression model with scalar response, and the case where both response and covariates might be functional object data. We address the first two cases here.

### 4.1 The case of equivalence classes as responses

As presented by the authors, the search for suitable elastic model formulations for curve responses naturally leads to consider extensions of manifold regression models already available for other quotient space structures. In the following, we discuss in how far *geodesic regression* (Fletcher 2013) as prototype of intrinsic regression on manifolds can be transferred to the space  $\mathcal{S}$  of curves modulo warping. Subsequently, we briefly describe an approach with a somewhat larger model class, which we refer to as ‘quotient linear regression’ (Steyer et al. 2023a), and which is so far the closest to a geodesic regression model on  $\mathcal{S}$  and also the only approach we are aware of that takes warping invariance intrinsically into account. *Intrinsic regression* respects the intrinsic geometry of the space in two aspects: A) the criterion with respect to which the model is fit to data (or the assumed error model), B) the model for the mean in dependence on covariates. In a more narrow sense of intrinsic regression, the criterion in A) is typically chosen to minimize the squared intrinsic distance, i.e., the distance corresponding to the length of the shortest connecting path, and B) the model propagates along geodesics, or is at least inspired by them.

We focus here on geodesic regression and briefly repeat its definition for a Riemannian manifold  $\mathcal{M}$  (such as arises, e.g., under shape invariances in 2D): assume the conditional (Fréchet) mean  $\kappa(x) \in \mathcal{M}$  follows a geodesic  $\kappa$  in dependence on a scalar covariate, say  $x \in [-1, 1]$ . Then we may write  $\kappa(x) = \exp_{\mathfrak{p}}(\beta x)$  for some ‘intercept’  $\mathfrak{p} \in \mathcal{M}$  and ‘slope’  $\beta \in T_{\mathfrak{p}}\mathcal{M}$ , a tangent vector in the tangent space at  $\mathfrak{p}$ , where  $\exp_{\mathfrak{p}}$  is the Riemannian exponential. Locally, a geodesic  $\kappa$  describes the shortest path between points  $\kappa(x)$  and  $\kappa(x')$  with  $x, x' \in [-\epsilon, \epsilon]$ ,  $\epsilon > 0$ . Here, we assume that this holds for the entire domain  $[-1, 1]$ .

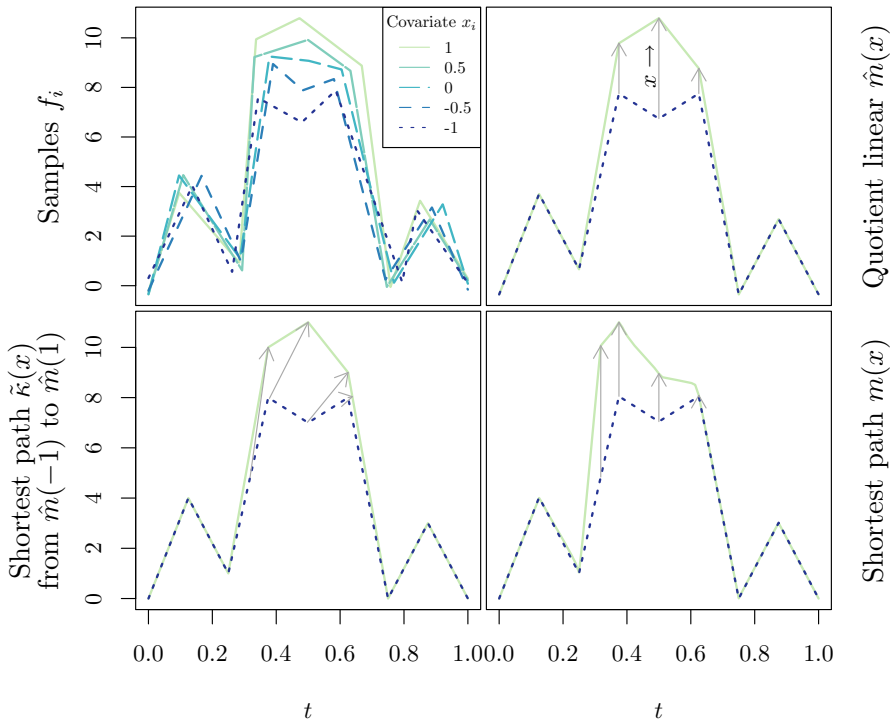
Attempting to formulate an analogous model on  $\mathcal{S}$ , where we have no Riemannian manifold structure available, we may still consider ‘shortest-path regression’ assuming that the conditional mean lies on a shortest path  $\tilde{\kappa} : [-1, 1] \rightarrow \mathcal{S}$  between some

$p_- = \tilde{\kappa}(-1)$  and  $p_+ = \tilde{\kappa}(1)$  in  $\mathcal{S}$ . (While in the context of metric spaces, where in general no exp-map is available, shortest paths are typically directly referred to as ‘geodesics’, we will not do this here to stress the difference.) Using that  $\mathcal{S}$  is a quotient metric space over  $\mathcal{Q} = \mathbb{L}^2$ , the space of SRV-transforms of curves, we can further refine the model formulation: assuming aligned representatives  $p_-, p_+ \in \mathcal{Q}$ ,  $[p_-] = p_-, [p_+] = p_+$ , we know that convex combinations of  $p_-$  and  $p_+$  are also aligned, and that  $\tilde{\kappa}$  corresponds to a line in  $\mathcal{Q}$  and can be written as  $\tilde{\kappa}(x) = [\mu(x)] = [p + \tilde{\beta}x]$  setting  $p = (p_- + p_+)/2$  and  $\tilde{\beta} = (p_+ - p_-)/2$ , as, e.g., implied by Steyer et al. (2023a) Corollary 2.12, providing us with an intercept and slope.

We have arrived at a model of the desired form, but before proceeding with a discussion of how to fit it, we make a short side remark to point out a connection to ‘generalized geodesics’ discussed by Huckemann et al. (2010) for geodesic PCA in certain quotient spaces not endowed with a full Riemannian manifold structure but exhibiting singularities: Determining a shortest path  $[\mu]$  on  $\mathcal{S}$ ,  $\mu : x \mapsto p + \tilde{\beta}x$  corresponds to a ‘horizontal geodesic’ in  $\mathcal{Q}$  pointing into a direction orthogonal to the invariance group action. Unlike the situation in Huckemann et al. (2010), where the ‘horizontal space’ of such directions forms a vector space, we know, however, little about the corresponding set  $\mathcal{H}_p = \mathcal{H}_p^- \cap \mathcal{H}_p^+$ , with  $\mathcal{H}_p^\pm = \{\tilde{\beta} \mid p \pm \epsilon \tilde{\beta} \text{ aligned to } p \text{ for some } \epsilon > 0\}$  in our case. We only have that  $\mathcal{H}_p^-$  and  $\mathcal{H}_p^+$  are convex cones (again Steyer et al. (2023a) Corollary 2.12). Hence, while our shortest-path regression model is similar to a ‘generalized geodesic model’, which could have presented a next step in mimicking geodesic regression, we cannot rely on the same useful vector space structure as in Huckemann et al. (2010) for model fitting.

To fit the model is to minimize the least-squares criterion  $LS(p_-, p_+) = \sum_{i=1}^n d_s([q_i], \tilde{\kappa}(x_i))^2$  for data  $(q_i, x_i) \in \mathcal{Q} \times [-1, 1]$ ,  $i = 1, \dots, n$  with respect to  $p_-, p_+ \in \mathcal{S}$ . A straightforward iterative algorithm would start with initial representatives  $q_i^{[0]} = q_i$ , and fit, in iteration  $j, 1$ . a functional linear model  $\hat{\mu}^{[j]}(x) = \hat{p}^{[j]} + \hat{\beta}^{[j]}x$  to covariate tuples  $(q_1^{[j]}, x_1), \dots, (q_n^{[j]}, x_n)$  before 2. aligning  $\hat{\mu}^{[j]}(1)$  to  $p_-^{[j]} = \hat{\mu}^{[j]}(-1)$  to obtain  $p_+^{[j]}$ , yielding an estimator  $\tilde{\kappa}^{[j]} : x \mapsto [\mu^{[j]}(x)]$  for  $\tilde{\kappa}$ , where  $\mu^{[j]}$  is the line between  $p_-^{[j]}$  and  $p_+^{[j]}$  as described above. In a 3. step, the  $q_i$  are then aligned to the  $\mu^{[j]}(x_i)$  to obtain  $q_i^{[j]}$  and we return to step 1. It is, however, not clear if the algorithm converges to a (local) minimum. In particular, step 2, ensuring  $\tilde{\kappa}$  is a shortest path, is not oriented toward risk reduction. By contrast, e.g., in a similar model for forms (size-and-shapes) of curves modulo rotation instead of modulo warping (both presenting isometric group actions),  $\hat{\beta}$  can be linearly constrained to the horizontal space, such that the optimization in step 1 already yields a geodesic line without further alignment (Stöcker et al. 2023). In conclusion, the described model presents a natural generalization of a geodesic model, which lacks, however, a promising fitting algorithm.

In contrast to the generalization of a geodesic model above, the ‘quotient linear model’ described in Steyer et al. (2023a) for responses in quotient metric spaces (focusing on curves modulo warping in  $\mathbb{R}^d$  with  $d \geq 2$  rather than  $d = 1$ ), drops the shortest path condition and directly considers models of the form  $\hat{\kappa}(x) = [\hat{\mu}(x)] = [\hat{p} + \hat{\beta}x]$  with parameters  $\hat{p}, \hat{\beta} \in \mathcal{Q}$ , defined by a, in this case linear, model  $\hat{\mu}$  on the



**Fig. 2** Artificial data scenario sketching a situation where quotient linear regression might be preferable over shortest path regression: *Top-left*: observed curve representatives  $f_i, i = 1, \dots, 5$ , with  $x_i \in [-1, 1]$ , (piecewise linear for simplicity). Phase variability seems present but limited. Schematics of a quotient linear model (*top-right*), depicting the back-transform  $\hat{m}(x)$  of the SRV-level mean  $\hat{\mu}(x)$ , are likely considered more natural than corresponding schematics for the shortest-path model  $\tilde{\kappa}$ , depicted with the same representatives  $\hat{m}(\pm 1) \in \mathfrak{p}_{\pm}$  (*bottom-left*) and aligned representatives  $m(\pm 1) \in \mathfrak{p}_{\pm}$  (*bottom-right*)

ambient space. Removing step 2 in the fitting approach described above, we obtain good fits of the quotient linear model.

Although motivated as feasible simplification of shortest-path regression here, the additional flexibility of a quotient linear model might in fact be desirable: it can describe effects where a “valley”, i.e., a minimum of  $p_-$  at some  $t_0 \in [0, 1]$ , transitions into a peak, i.e., a maximum of  $p_+(t_0)$ , with increasing  $x$  (or vice versa). The line between such  $p_-$  and  $p_+$  does not correspond to a shortest path between  $\mathfrak{p}_- = [p_-]$  and  $\mathfrak{p}_+ = [p_+]$ , since optimal alignment—not generally aligning minima with maxima—would yield a warping function  $\gamma$  with  $\gamma(t_0) \neq t_0$  in this case. A corresponding model can, however, be natural in certain data scenarios, as illustrated in Fig. 2. Conversely, if the true model is in fact a shortest path  $\tilde{\kappa}$ , the quotient linear model still yields consistent estimators using the larger model class (in the sense of Steyer et al. (2023a) Corollary 2.7), or an approximation of it if  $\hat{\beta}$  is restricted to, e.g., be a spline function.

While we focused our discussion on the case of one scalar covariate  $x \in [-1, 1]$ , the extension to a multiple model with covariate vector  $x = (x_1, \dots, x_J)^T$  brings

additional challenges for intrinsic models. While for Riemannian manifolds, these can take the form  $\kappa(x) = \exp_p(\sum_{j=1}^J \beta_j x_j)$  as in Cornea et al. (2017); Stöcker et al. (2023), the generalization to  $\mathcal{S}$  modulo warping is even more challenging. The proposed quotient linear models, by contrast, have the additional advantage of naturally generalizing to multiple regression.

### 4.2 The case of elastic functional covariates

Just as for elastic functions as responses, the reverse case, constructing an intrinsic functional regression model with elastic functions as covariates, seems very challenging. A first proposal is given in Ahn et al. (2020), also referred to by the authors, but we believe that many open problems remain, as discussed in the following. For better readability, we will here focus on the case of the 'linear' model, where the index function  $g$  is the identity. Thus, we consider the model

$$y_i = \sup_{\gamma_i \in \Gamma} \langle \beta, (q_i \circ \gamma_i) \sqrt{\dot{\gamma}_i} \rangle + \epsilon_i, \quad \epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2),$$

where  $q_i$  denotes the SRV transformation of the covariate function  $f_i$  for all  $i = 1, \dots, n$ .

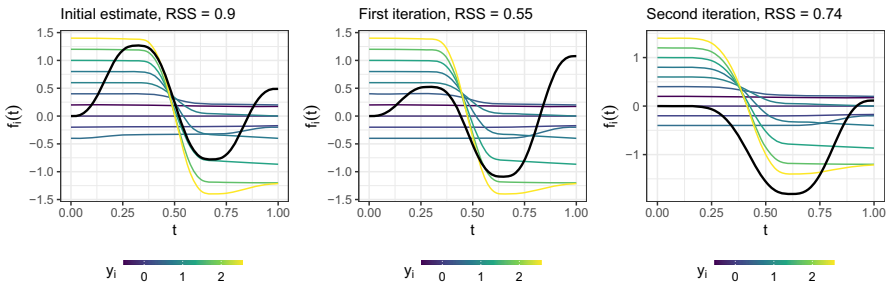
A somewhat unexpected and possibly undesirable behavior of this 'linear' model is that the expected value of the response is always positive, i.e.,  $\mathbb{E}(y_i) \geq 0$  for all  $i = 1, \dots, n$ , since  $\sup_{\gamma_i \in \Gamma} \langle \beta, (q_i \circ \gamma_i) \sqrt{\dot{\gamma}_i} \rangle \geq 0$  for all  $i = 1, \dots, n$ . A proof for this statement can be constructed similarly to the proof in the case of piecewise linear functions given in the online supplement of Steyer et al. (2023b). In addition to this unusual behavior, we see the main difficulty of the proposed model in the estimation of the model parameter  $\beta \in \mathbb{L}_2$ . Least squares (or equivalently, maximum likelihood) leads to the optimization problem

$$\operatorname{argmin}_{\beta} \sum_{i=1}^n \left( y_i - \sup_{\gamma_i \in \Gamma} \langle \beta, (q_i \circ \gamma_i) \sqrt{\dot{\gamma}_i} \rangle \right)^2,$$

which cannot be solved analytically because of the supremum within the minimization problem. The authors in Ahn et al. (2020) propose to iterate between minimizing with respect to  $\beta$  and maximizing with respect to the warping functions  $\gamma_i$ . Although this procedure shows promising results in simulations, it remains unclear under which conditions this algorithm will converge and find an optimal  $\beta$ , since in general the residual sum of squares (RSS) cannot be expected to decrease with each iteration. A counterexample, where the RSS is larger after the second iteration than after the first (and continues to non-systematically increase and decrease in the following iterations), is shown in Fig. 3.

Additionally to these points, issues with non-identifiability in non-elastic models with functional covariates (Scheipl and Greven 2016) are expected to become even more challenging in the elastic setting. Thus, we consider the case of elastic functional





**Fig. 3** Example of an attempt to iteratively estimate the effect function (black, thick line) for elastic covariate functions  $f_i$  and response  $y_i$ ,  $i = 1, \dots, 10$ . The initial estimate is chosen as the true parameter function  $\beta$ , which is known in this simulated setting

covariates as an interesting area of open research problems where we are looking forward to further developments in future.

## 5 Discussion

The SRV framework offers many opportunities for the field of functional data analysis, whenever the parametrization is irrelevant, or when alignment or a decomposition into phase and amplitude information is sought. At the same time, a number of interesting open challenges remain. First, while the SRV framework has shown to be superior to simpler approaches for the purpose of alignment, optimal alignment in the elastic distance sense does not always produce the most meaningful or desirable alignment in practice, as no alignment method can, and in fact the question of what kind of alignment is considered the most meaningful may be answered differently in different settings. We discussed some realistic data settings, where elastic alignment may have limitations, including different numbers of maxima such as in the COVID wave data (cf. Sect. 2), or transitions from maxima to minima over the course of a covariate (cf. Fig. 2). Thus, the goal of the alignment should always be considered carefully and objectives chosen correspondingly. The question of which curve to align to needs to be similarly carefully considered. Alignment of all curves to the mean, for instance, does not necessarily yield curves that are aligned to each other, complicating the interpretation of a sample of ‘aligned’ curves. Also, alignment to an overall mean may not always be ideal, in particular if curves do not all have the same (number of) features. For example, we saw that for regression in such settings, alignment to a model prediction (Steyer et al. 2023a) performs superior to pre-alignment to an overall mean due to the closer similarity. The question of which curve to align to is also relevant for the problem of elastic functional covariates discussed in Sect. 4.2 above. Lastly, much work remains in particular in the context of sparse and noisy data. We discussed some extensions to sparse data settings in Sect. 4 above, but while sparse and error-prone data are commonly addressed together in (nonelastic) functional data analysis, the case with error has not yet received much attention in the elastic setting and is in fact much more challenging. Not only can error cause problems with location and

height of minima and maxima, but error also causes problems with any derivative-based approach including SRVs. While in (nonelastic) FDA, denser sampling grids lead to better recovery of the underlying function even in the error-prone case based on smoothing, denser grids lead to an inflation of any error on the derivative level approximated by finite differences. Here, in fact, somewhat sparser sampling can be helpful if error is present. It may be for this reason, that while we see interesting challenges remaining in this theoretically demanding setting, in practice we see good and sometimes surprisingly good results using the SRV framework, e.g., recovering important data structures in an elastic setting even under extreme sparsity (Steyer et al. 2023b; Stöcker et al. 2022).

**Acknowledgements** The authors gratefully acknowledge funding by Grant No. GR 3793/3-1 from the German research foundation (DFG) for the project ‘Flexible regression methods for curve and shape data’.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahn K, Tucker JD, Wu W, Srivastava A (2020) Regression models using shapes of functions as predictors. *Comput Stat Data Anal* 151:107017
- Cornea E, Zhu H, Kim P, Ibrahim JG (2017) Regression models on Riemannian symmetric spaces. *J R Stat Soc Ser B Stat Methodol* 79(2):463–482
- Dryden IL, Mardia KV (2016) *Statistical shape analysis: with applications in R*, 2nd edn. Wiley, Chichester
- Fletcher PT (2013) Geodesic regression and the theory of least squares on Riemannian manifolds. *Int J Comput Vision* 105(2):171–185
- Hadjipantelis PZ, Aston JA, Müller H-G, Evans JP (2015) Unifying amplitude and phase analysis: a compositional data approach to functional multivariate mixed-effects modeling of Mandarin Chinese. *J Am Stat Assoc* 110(510):545–559
- Happ C, Scheipl F, Gabriel A-A, Greven S (2019) A general framework for multivariate functional principal component analysis of amplitude and phase variation. *Stat* 8(1):220
- Huckemann S, Hotz T, Munk A (2010) Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica* 20(1):1–58
- Lahiri S, Robinson D, Klassen E (2015) Precise matching of pl curves in  $\mathbb{R}^n$  in the square root velocity framework. *Geometry, Imag Comput* 2(3):133–186
- Pegoraro M, Secchi P (2021) Functional data representation with merge trees. *arXiv preprint arXiv:2108.13147*
- Scheipl F, Greven S (2016) Identifiability in penalized function-on-function regression models. *Electr J Stat* 10(1):495–526
- Srivastava A, Klassen EP (2016) *Functional and shape data analysis*, vol 1. Springer, New York
- Steyer L, Stöcker A, Greven S (2023a) Regression in quotient metric spaces with a focus on elastic curves. *arXiv preprint arXiv:2305.02075*
- Steyer L, Stöcker A, Greven S (2023b) Elastic analysis of irregularly or sparsely sampled curves. *Biometrics* 79:2103–2115

- Stöcker A, Steyer L, Grevén S (2023) Functional additive models on manifolds of planar shapes and forms. *J Comput Graph Stat.* <https://doi.org/10.1080/10618600.2023.2175687>
- Stöcker A, Pfeuffer M, Steyer L, Grevén S (2022) Elastic full Procrustes analysis of plane curves via Hermitian covariance smoothing. arXiv preprint [arXiv:2203.10522](https://arxiv.org/abs/2203.10522)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.