Functional Additive Models on Manifolds of Planar Shapes and Forms

Almond Stöcker, Lisa Steyer, and Sonja Greven
School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

ABSTRACT
The “shape” of a planar curve and/or landmark configuration is considered its equivalence class under translation, rotation, and scaling, its “form” its equivalence class under translation and rotation while scale is preserved. We extend generalized additive regression to models for such shapes/forms as responses respecting the resulting quotient geometry by employing the squared geodesic distance as loss function and a geodesic response function to map the additive predictor to the shape/form space. For fitting the model, we propose a Riemannian $L_2$-Boosting algorithm well suited for a potentially large number of possibly parameter-intensive model terms, which also yields automated model selection. We provide novel intuitively interpretable visualizations for (even nonlinear) covariate effects in the shape/form space via suitable tensor-product factorization. The usefulness of the proposed framework is illustrated in an analysis of (a) astragalus shapes of wild and domesticated sheep and (b) cell forms generated in a biophysical model, as well as (c) in a realistic simulation study with response shapes and forms motivated from a dataset on bottle outlines. Supplementary materials for this article are available online.

1. Introduction

In many imaging data problems, the coordinate system of recorded objects is arbitrary or explicitly not of interest. Statistical shape analysis (Dryden and Mardia 2016) addresses this point by identifying the ultimate object of analysis as the shape of an observation, reflecting its geometric properties invariant under translation, rotation and rescaling, or as its form (or size-and-shape) invariant under translation and rotation. This article establishes a flexible additive regression framework for modeling the shape or form of planar (potentially irregularly sampled) curves and/or landmark configurations in dependence on scalar covariates. A rich shape analysis literature has been developed for 2D or 3D landmark configurations—presenting for instance selected points of a bone or face—which are considered elements of Kendall’s shape space (see, e.g., Dryden and Mardia 2016). In many 2D scenarios, however, observed points describe a curve reflecting the outline of an object rather than dedicated landmarks (Adams, Rohlf, and Slice 2013). Considering outlines as images of (parameterized) curves shows a direct link to functional data analysis (FDA, Ramsay and Silverman 2005) and, in this context, we speak of functional shape/form data analysis. As in FDA, functional shape/form data can be observed on a common and often dense grid (regular/dense design) or on curve-specific often sparse grids (irregular/sparse design). While in the regular case, analysis often simplifies by treating curve evaluations as multivariate data, more general irregular designs gave rise to further developments in sparse FDA (e.g., Yao, Müller, and Wang 2005; Greven and Scheipl 2017), explicitly considering irregular measurements instead of pre-smoothing curves. To the best of our knowledge, we are the first to consider irregular/sparse designs in the context of functional shape/form analysis.

Shapes and forms are examples of manifold data. Petersen and Müller (2019) propose “Fréchet regression” for random elements in general metric spaces, which requires estimation of a (potentially negatively) weighted Fréchet mean for each covariate combination. Their implicit rather than explicit model formulation renders model interpretation difficult. More explicit model formulations have been developed for the special case of a Riemannian geometry. Besides tangent space models (Kent et al. 2001), extrinsic models (Lin et al. 2017) and models based on unwrapping (Jupp and Kent 1987; Mallasto and Feragen 2018), a variety of manifold regression models have been designed based on the intrinsic Riemannian geometry. Starting from geodesic regression (Fletcher 2013), which extends linear regression to curved spaces, these include MANOVA ( Huckemann, Hotz, and Munk 2010), polynomial regression (Hinkle, Fletcher, and Joshi 2014), smoothing splines (Kume, Dryden, and Le 2007), regression along geodesic paths with nonconstant speed (Hong et al. 2014), or kernel regression (Davis et al. 2010) and Krüging (Pigoli, Menafoglio, and Secchi 2016). However, mostly only one metric covariate or categorical covariates are considered, possibly in hierarchical model extensions for longitudinal data (Muralidharan and Fletcher 2012; Schiratti et al. 2017). By contrast, Zhu et al. (2009), Shi et al. (2009), and Kim et al. (2014) generalize geodesic regression to regression with multiple covariates focusing on Symmetric Positive-Definite (SPD) matrix responses. Cornea et al. (2017) develop a general Generalized Linear Model (GLM) analogue regression framework for responses in a symmetric manifold and apply it to shape analysis.

CONTACT Almond Stöcker almond.stoecker@hu-berlin.de School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/v/JCGS.
Recently, Lin, Müller, and Park (2020) proposed a Lie group additive regression model for Riemannian manifolds focusing on SPD matrices rather than shapes.

In FDA, there is a much wider range of developed regression methods (see overviews in Morris 2015; Greven and Scheipl 2017). Among the most flexible models are Functional Additive Models (FAMs) for (univariate) functional responses (in contrast to FAMs with functional covariates (Ferraty et al. 2011)) with either semi- or nonparametric approaches to model (a) response functions and (b) smooth covariate effects. For (a), nonparametric approaches formulate estimation problems in infinite-dimensional model spaces to motivate finite-dimensional representations or effectively evaluate curves on grids (e.g., Jeon and Park 2020). Semi-parametric approaches directly employ finite expansions in spline bases (Brockhaus, Scheipl, and Greven 2015), Functional Principal Component (FPC) bases (Morris and Carroll 2006) or both (Scheipl, Staicu, and Greven 2015), as well as wavelets (Meyer et al. 2015), sometimes directly expanding functions to model on coefficients and sometimes expanding only predictions while keeping the raw measurements. Nonparametric approaches are formulated in infinite-dimensional model spaces and effectively evaluate curves on grids or apply pre-smoothing techniques (e.g., Jeon and Park 2020). For (b), again semiparametric penalized spline basis approaches are employed (Scheipl, Staicu, and Greven 2015; Brockhaus, Scheipl, and Greven 2015), or local linear/polynomial (Müller and Yao 2008; Jeon et al. 2022) or other nonparametric kernel-based approaches (Jeon and Park 2020; Jeon, Park, and Van Keilegom 2021). Semi- and nonparametric approaches come with different theoretical and practical advantages, but similarities such as regarding asymptotic behavior are also known from scalar nonparametric regression (Li and Ruppert 2008). Advantages of the semi-parametric approach summarized in Greven and Scheipl (2017) include its appropriateness for sparse irregular functional data and its modular extensibility to functional mixed models (Scheipl, Staicu, and Greven 2015; Meyer et al. 2015) and nonstandard response distributions (Brockhaus, Scheipl, and Greven 2015; Stöcker et al. 2021). For bivariate or multivariate functional responses, which are closest to functional shapes/forms but without invariances, Rosen and Thompson (2009), Zhu, Li, and Kong (2012), Olsen, Markussen, and Raket (2018) consider linear fixed effects of scalar covariates, the latter also allowing for warping. Zhu et al. (2017), Backenroth et al. (2018) consider one or more random effects for one grouping variable, linear fixed effects and common dense grids for all functions. Volkmann et al. (2021) combine the FAM model class of Greven and Scheipl (2017) with multivariate FPC analysis (Happ and Greven 2018) to model multivariate (sparse) functional responses.

This article establishes an interpretable FAM framework for modeling the shape or form of planar (potentially irregularly sampled) curves and/or landmark configurations in dependence on scalar covariates, extending $L^2$-Boosting (Bühlmann and Yu 2003; Brockhaus, Scheipl, and Greven 2015) to Riemannian manifolds for model estimation. The three major contributions of our regression framework are: (i) We introduce additive regression with shapes/forms of planar curves and/or landmarks as response, extending FAMs to nonlinear response spaces or, vice versa, extending GLM-type regression on manifolds for landmark shapes both to functional shape manifolds and to include (nonlinear) additive model effects. (ii) We propose a novel Riemannian $L^2$-Boosting algorithm for estimating regression models for this type of manifold response, and (iii) a visualization technique based on tensor-product factorization yielding intuitive interpretations even of multi-dimensional smooth covariate effects for practitioners. Although related tensor-product model transformations based on higher-order SVD have been used, e.g., in control engineering (Baranyi, Yam, and Várkai 2013), we are not aware of any comparable application for visualization in FAMs or other statistical models for object data. Despite our focus on shapes and forms, transfer of the model, Riemannian $L^2$-Boosting, and factorized visualization to other Riemannian manifold responses is intended in the generality of the formulation and the design of the provided R package manifoldboost (developer version on github.com/Almond-S/manifoldboost). The versatile applicability of the approach is illustrated in three different scenarios: an analysis of the shape of sheep astragali (ankle bones) represented by both regularly sampled curves and landmarks in dependence on categorical “demographic” variables; an analysis of the effects of different metric biophysical model parameters (including smooth interactions) on the form of (irregularly sampled) cell outlines generated from a cellular Potts model; and a simulation study with irregularly sampled functional shape and form responses generated from a dataset of different bottle outlines and including metric and categorical covariates.

In Section 2, we introduce the manifold geometry of irregular curves modulo translation, rotation and potentially rescaling, which underlies the intrinsic additive regression model formulated in Section 3. The Riemannian $L^2$-Boosting algorithm is introduced in Section 4. Section 5 analyzes different data problems, modeling sheep bone shape responses (Section 5.1) and cell outlines (Section 5.2). Section 5.3 summarizes the results of simulation studies with functional shape and form responses. We conclude with a discussion in Section 6.

2. Geometry of Functional Shapes and Forms

Riemannian manifolds of planar shapes (and forms) are discussed in various textbooks at different levels of generality, in finite (Kendall et al. 1999; Dryden and Mardia 2016) or potentially infinite dimensions (Srivastava and Klassen 2016; Klingenberg 1995). Starting from the Hilbert space $Y$ of curve representatives $y$ of a single shape or form observation, we successively characterize its quotient space geometry under translation, rotation and rescaling including the respective tangent spaces. Building on that, we introduce Riemannian exponential and logarithmic maps and parallel transports needed for model formulation and fitting, and the sample space of (irregularly observed) functional shapes/forms.

To make use of complex arithmetic, we identify the two-dimensional plane with the complex numbers, $\mathbb{R}^2 \cong \mathbb{C}$, and consider a planar curve to be a function $y : \mathbb{R} \to \mathbb{C}$, element of a separable complex Hilbert space $\mathcal{Y}$ with a complex inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$. This allows simple scalar expressions for the group actions of translation $\mathbb{R}^2 \ni (y \mapsto y + \gamma f : \gamma \in \mathbb{C})$ with $f \in \mathcal{Y}$ canonically

$\mathbb{R}^2 = \{y \mapsto y + \gamma f : \gamma \in \mathbb{C}\}$ with $f \in \mathcal{Y}$ canonically
Figure 1. Left: Quotient space geometry: assuming $p$ and $y$ centered, translation invariance is not further considered in the plot; given pole representative $p$, we express $y = \frac{\text{Re}(p'y)}{|p|} p + \frac{\text{Im}(p'y)}{|p|} p + (y - \frac{p'y}{|p|} p) \in \mathcal{Y}$ in its coordinates in $p$ and $p'$ direction, subsuming all orthogonal directions in the third dimension. In this coordinate system, the rotation orbit $\mathcal{Y}_{\text{Rot}}$ corresponds to the dotted horizontal circle, and is identified with the aligned $\tilde{y} := \mathcal{Y}_{\text{Rot}}$ in the half-plane of $p$; $\mathcal{Y}_{\text{Rot}} \times S^d = \frac{\text{Im}(p'y)}{|p|} p$ projecting $y$ onto the hemisphere depicted by the vertical semicircle. Form and shape distances between $[p]$ and $[y]$ correspond to the length of the geodesics $\zeta(t)$ (thick lines) on the plane and sphere, respectively. Right: Geodesic line $c(t)$ between $p = c(0)$ and $p' = c(1)$, Log-map projecting $y$ to $c \in T_p \mathcal{M}$, parallel transport $\text{Trans}_{pp'}$ forwarding $c$ to $\tilde{c} \in T_{p'} \mathcal{M}$, and Exp-map projecting $\tilde{c}$ onto $\mathcal{M}$ visualized for a sphere. Tangent spaces, identified with subspaces of the ambient space, are depicted as gray planes above the respective poles. The parallel transport preserves all angles between tangent vectors and identifies $\zeta(0) \cong \tilde{c}(1)$.

The rotation orbit $\mathcal{Y}_{\text{Rot}}$ of $\mathcal{Y}$ refers to the rotation group $\mathcal{G}$ acting on the underlying real vector space of planar curves $\mathcal{Y}$, under the rotation group $\mathcal{G} := \text{Rot}(\mathcal{Y})$. The rotation orbit $\mathcal{Y}_{\text{Rot}}$ corresponds to the dotted horizontal circle, and is identified with the aligned $\tilde{y} := \mathcal{Y}_{\text{Rot}}$ in the half-plane of $p$; $\mathcal{Y}_{\text{Rot}} \times S^d = \frac{\text{Im}(p'y)}{|p|} p$ projecting $y$ onto the hemisphere depicted by the vertical semicircle. Form and shape distances between $[p]$ and $[y]$ correspond to the length of the geodesics $\zeta(t)$ (thick lines) on the plane and sphere, respectively. Right: Geodesic line $c(t)$ between $p = c(0)$ and $p' = c(1)$, Log-map projecting $y$ to $c \in T_p \mathcal{M}$, parallel transport $\text{Trans}_{pp'}$ forwarding $c$ to $\tilde{c} \in T_{p'} \mathcal{M}$, and Exp-map projecting $\tilde{c}$ onto $\mathcal{M}$ visualized for a sphere. Tangent spaces, identified with subspaces of the ambient space, are depicted as gray planes above the respective poles. The parallel transport preserves all angles between tangent vectors and identifies $\zeta(0) \cong \tilde{c}(1)$.

The ultimate response object is given by the orbit $[y]_G = \{g(y) : g \in G\}$ (or short $[y]$) of $y \in \mathcal{Y}$, the equivalence class under the respective combined group actions $G$ with $G = \text{Tri} \times \text{Rot} \times S^d$, $y = [y]_{\text{Tri} \times \text{Rot} \times S^d} = \{ux + y + f : \lambda \in \mathbb{R}^+, u \in \mathbb{R}^1, y \in \mathbb{C}\}$ is referred to as the shape of $y$ and, for $G = \text{Tri} \times \text{Rot}$, $y = [y]_{\text{Tri} \times \text{Rot}} = \{ux + y + f : u \in \mathbb{R}^1, y \in \mathbb{C}\}$ as its form or size-and-shape. $\mathcal{Y}/G = \{[y]_G : y \in \mathcal{Y}\}$ denotes the quotient space of $\mathcal{Y}$ with respect to $G$. The description of the Riemannian geometry of $\mathcal{Y}/G$ involves, in particular, a description of the tangent spaces $T_{[y]} \mathcal{Y}/G$ at points $[y] \in \mathcal{Y}/G$, which can be considered local vector space approximations to $\mathcal{Y}/G$ in a neighborhood of $[y]$. For a point $q$ in a manifold $\mathcal{M}$ the tangent vectors $\beta \in T_q \mathcal{M}$ can, i.e., be thought of as gradients $\zeta(0)$ of paths $c : \mathbb{R} \supset (-\delta, \delta) \rightarrow \mathcal{M}$ at $0$ where they pass through $c(0) = q$. Besides their geometric meaning, they will also play an important role in the regression model, as additive model effects are formulated on tangent space level. Choosing suitable representatives $\tilde{y} \in [y]_G \subset \mathcal{Y}$ (or short $y$) of orbits $[y]_G$, we use an identification of tangent spaces with suitable linear subspaces $T_{[y]} \mathcal{Y}/G \subset \mathcal{Y}$.

Form geometry: Starting with translation as the simplest invariance, an orbit $[y]_{\text{Tri} \times \text{Rot}}$ can be one-to-one identified with its centered representative $\tilde{y}_{\text{Tri}} = y - \langle y, t \rangle t$ yielding an identification $\mathcal{Y}_{\text{Tri}} \cong \{ y \in \mathcal{Y} : \langle y, t \rangle t = 0 \}$ with a linear subspace of $\mathcal{Y}$. Hence, also $T_{[y]} \mathcal{Y}_{\text{Tri}} = \{ y \in \mathcal{Y} : \langle y, t \rangle t = 0 \}$. For rotation, by contrast, we can only find local identifications with Hilbert subspaces (i.e., charts) around reference points $[p]_{\text{Tri} \times \text{Rot}}$ we refer to as “poles”. Moreover, we restrict to $y, p \in \mathcal{Y}^{*} = \mathcal{Y} \setminus \{0\}_{\text{Tri}}$ eliminating constant functions as degenerate special cases in the translation orbit of zero. For each $[y]_{\text{Tri} \times \text{Rot}}$ in an open neighborhood around $[p]_{\text{Tri} \times \text{Rot}}$ which can be chosen with $\langle \tilde{y}_{\text{Tri}}^{\text{Tri}}, \tilde{y}_{\text{Tri}}^{\text{Tri}} \rangle \neq 0$, $y$ can be uniquely rotation aligned to $p$, yielding a one-to-one identification of the form $[y]_{\text{Tri} \times \text{Rot}} \times \text{Rot} = \langle \tilde{y}_{\text{Tri}}^{\text{Tri}}, \tilde{y}_{\text{Tri}}^{\text{Tri}} \rangle \tilde{y}_{\text{Tri}}^{\text{Tri}} = \arg \min \|y' - p\|$ (compare Figure 1).

While $\tilde{y}_{\text{Tri}}^{\text{Tri} \times \text{Rot}}$ depends on $p$, we omit this in the notation for simplicity. All $\tilde{y}_{\text{Tri}}^{\text{Tri} \times \text{Rot}}$ rotation aligned to $\tilde{y}_{\text{Tri}}^{\text{Tri}}$ lie on the hyper-plane determined by $\text{Im}(\langle \tilde{y}_{\text{Tri}}^{\text{Tri}}, \tilde{y}_{\text{Tri}}^{\text{Tri}} \rangle) = 0$ (Figure 1), which yields $T_{[y]} \mathcal{Y}_{\text{Tri} \times \text{Rot}} = \{ y \in \mathcal{Y} : \langle y, t \rangle t = 0, \text{Im}(\langle y, p \rangle) = 0 \}$ with normal vectors $\tilde{\zeta}(1) = t, \tilde{\zeta}(2) = \sqrt{-1} t, \tilde{\zeta}(3) = \sqrt{-1} p$. Note that, despite the use of complex arithmetic, $T_{[y]} \mathcal{Y}_{\text{Tri} \times \text{Rot}}^{*}$ is a real
vector space not closed under complex scalar multiplication. The geodesic distance of $[y]_{Trl \times Rot}$ to the pole $[p]_{Trl \times Rot}$ is given by $d([y]_{Trl \times Rot}, [p]_{Trl \times Rot}) = \|\tilde{y}^{Trl \times Rot} - \tilde{p}^{Trl \times Rot}\|$. It reflects the length of the shortest path (i.e., the geodesic) between the forms and the minimum distance between the orbits as sets.

**Shape geometry:** To account for scale invariance in shapes $[y]_{Trl \times Rot \times Scl}$, they are identified with normalized representatives $\tilde{y}^{Trl \times Rot \times Scl} = \tilde{y}^{Trl \times Rot} \cdot \text{Vol}([p]_{Trl \times Rot \times Scl})^{-1}$. Motivated by the normalization, we borrow the well-known geometry of the sphere $\mathcal{S} = \{ y \in \mathcal{Y} : \|y\| = 1 \}$, where $T_y \mathcal{S} = \{ y \in \mathcal{Y} : R_e(y, p) = 0 \}$ is the tangent space at a point $p$ and geodesics are great circles. Together with translation and rotation invariance, the shape tangent space is then given by $T_{[p]}\mathcal{Y}^*_{Trl \times Rot \times Scl} = T_{[p]}\mathcal{Y}^*_{Trl \times Rot} \cap T_{[p]}\mathcal{S} = \{ y \in \mathcal{Y} : (y, r) = 0, (y, p) = 0 \}$ with normal vector $\tilde{\epsilon} = \tilde{p}/\|\tilde{p}\|$ in addition to $\tilde{\epsilon}(1), \tilde{\epsilon}(2), \tilde{\epsilon}(3)$ above. The geodesic distance $d([p]_{Trl \times Rot \times Scl}, [y]_{Trl \times Rot \times Scl}) = \arccos(\tilde{y}^{Trl \times Rot \times Scl} \cdot \tilde{p}^{Trl \times Rot \times Scl})$ corresponds to the arc-length between the representatives. This distance is often referred to as *Procrustes distance* in statistical shape analysis.

We may now define the maps needed for the regression model formulation. Let $\tilde{y}$ and $\tilde{p}$ be shape/form representatives of $y$ and $p$ rotation aligned to the shape/form pole representative $p$. Generalizing straight lines to a Riemannian manifold $\mathcal{M}$, geodesics $c : (-\delta, \delta) \to \mathcal{M}$ can be characterized by their “intercept” $c(0) \in \mathcal{M}$ and “slope” $\dot{c}(0) \in T_{c(0)}\mathcal{M}$. The exponential map $\text{Exp}_p : T_p \mathcal{M} \to \mathcal{M}$ at a point $q \in \mathcal{M}$ is defined to map $\beta \mapsto c(1)$ for $c$ the geodesic with $c(0) = q$ and $\dot{c}(0) = \beta$. It maps $\beta \in T_q \mathcal{M}$ to a point $\text{Exp}_q(\beta)$ in $\mathcal{M}$ located $d(q, \text{Exp}_q(\beta)) = \|\beta\|$ apart of the pole in the direction of $\beta$. On the form space $[y]_{Trl \times Rot}$, the exponential map is simply given by $\text{Exp}_{[\tilde{p}]_{Trl \times Rot}}(\beta) = \tilde{y}^{Trl \times Rot} + \tilde{p}^{Trl \times Rot}$ above.

Consider a data scenario with $n$ observations of a random response covariate tuple $(Y, X)$, where the realizations of $Y$ are planar curves $y_i : \mathcal{T} \to \mathbb{R}^2$, $i = 1, \ldots, n$, belonging to a Hilbert space $\mathcal{Y}$ defined as above and potentially irregularly measured on individual grids $t_1 \leq \cdots \leq t_{i,k} \in \mathcal{T}$. The response object $Y$ is the equivalence class of $Y$ with respect to translation, rotation and possibly scale and the sample $\{y_1, \ldots, y_n\}$ is equipped with the respective Riemannian manifold geometry introduced in the previous section. For $i = 1, \ldots, n$, realizations $x_i \in \mathcal{X}$ of a covariate vector $X$ in a covariate space $\mathcal{X}$ are observed. $X$ can contain several categorical and/or metric covariates.

For regressing the mean of $[Y]$ on $X = x$, we model the shape/form $[\mu]$ of $[y] \in \mathcal{Y}$ as

$$[\mu] = \text{Exp}_{[\tilde{p}]}(h(x)) = \text{Exp}_{[\tilde{p}]}\left(\sum_{j=1}^{J} h_j(x)\right),$$

with an additive predictor $h : \mathcal{X} \to T_{[\tilde{p}]}\mathcal{Y}^*_{\mathcal{G}}$ acting at the tangent space at an “intercept” $[p] \in \mathcal{Y}^*_{\mathcal{G}}$. Generalizing an additive model $Y = \mu + \epsilon = p + h(x) + \epsilon$ in a linear space, we implicitly define $[\mu]$ as the conditional mean of $[Y]$ given $X = x$ by assuming zero-mean “residuals” $\epsilon$. In their definition, we follow Cornea et al. (2017) but extend to the functional response variational approach of the form case.

Judge and Mardia (2016), Hakemmann, Hotz, and Munk (2010), a corresponding derivation for the form case is given in Section S.1.2, supplementary materials including a discussion of the quotient space geometry in differential geometric terms.

Based on this understanding of the response space, we may now proceed to consider a sample of curves $y_1, \ldots, y_n \in \mathcal{Y}$ representing orbits $[y_1], \ldots, [y_n]$ with respect to group actions $G$. In the functional case, with the domain $\mathcal{T} = [0, 1]$, these curves are usually observed as evaluations $y_i = (y_i(t_1), \ldots, y_i(t_{i,k}))$ on a finite grid $t_1 < \cdots < t_{i,k} \in \mathcal{T}$ which may differ between observations. In contrast to the regular case common grids, this more general data structure is referred to as *irregular functional shape/form data.* To handle this setting, we replace the original inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{Y}$ by individual $\langle y_i, y_j \rangle = y_i^\top W y_j$ providing inner products on the $k_i$-dimensional space $Y_i = \mathbb{C}^{k_i}$ of evaluations $y_i, y_j$ on the same grid. The symmetric positive-definite weight matrix $W_i$ can be chosen to implement an approximation to integration w.r.t. the original measure $v$ with a numerical integration measure $\nu_i$ such as given by the trapezoidal rule. Alternatively, $W_i = \frac{1}{k_i} I_{k_i}$ with $k_i \times k_i$ identity matrix $I_{k_i}$ presents a canonical choice that is analog to the landmark case for $k_i \equiv k$. Moreover, data-driven $W_i$ could also be motivated from the covariance structure estimated for (potentially sparse) $y_1, \ldots, y_n$ along the lines of Yao, Müller, and Wang (2005), Stöcker et al. (2022). While this is beyond the scope of this article, potential procedures are sketched in Section S.7, supplementary materials.
which corresponds to $E(\epsilon_{[\mu]}(t)) = 0$ for ($v$-almost) all $t \in \mathcal{T}$. Here, we assume $[Y]$ is sufficiently close to $[\mu]$ with probability 1 such that $\log(\hat{[\mu]})$ is well-defined, which is the case whenever $\tilde{Y}, \tilde{\mu} \neq 0$ for centered shape/form representatives $\tilde{Y}$ and $\tilde{\mu}$, an unrestricted and common assumption (compare also Cornea et al. 2017). However, residuals $\epsilon_{[\mu]}$ for different $[\mu]$ belong to separate tangent spaces. To obtain a formulation in a common linear space instead, local residuals are mapped to residuals $\epsilon = \text{Transf}_{[\mu]}[p]\left(\epsilon_{[\mu]}\right)$ by parallel transporting them from $[\mu]$ to the common covariate independent pole $[p]$. After this isometric mapping into $T_\mu Y_{\mathcal{G}}^\vee$, we can equivalently define the conditional mean $[\mu]$ via $E(\epsilon) = 0$ for the transported residuals $\epsilon$.

Exp$_{[p]}$ maps the additive predictor $h(x) = \sum_{j=1}^{j} h_j(x) \in T_{[p]} Y_{\mathcal{G}}^\vee$ to the response space. It is analogous to a response function in GLMs but depends on $[p]$. Although covariate effects $h_j(x)$ often only depend on an individual covariate in $x$ for each $j$, they might also depend on covariate combinations in general to allow (smooth) interactions. While other response functions could be used, we restrict to the exponential map here, such that the model contains a geodesic model (Fletcher 2013)—the direct generalization of simple linear regression—as a special case for $h(x) = \beta x_1$ with a single covariate $x_1$ and tangent vector $\beta$. Typically, it is assumed that $h$ is centered such that $E(h(X)) = 0$, and the pole $[p]$ is the overall mean of $[Y]$ defined, like the conditional mean, via residuals of mean zero.

### 3.1. Tensor-Product Effect Functions $h_j$

Scheipl, Staicu, and Greven (2015) and other authors employ Tensor-Product (TP) bases for functional additive model terms. This naturally extends to tangent space effects, which we model as

$$h_j(x) = \sum_{r,l} \theta_{j}^{(r,l)} b_{j}^{(l)}(x) \partial_r,$$

with the TP basis given by the pair-wise products of $m$ linearly independent tangent vectors $\partial_r \in T_{[p]} Y_{\mathcal{G}}^\vee$, $r = 1, \ldots, m$, and $m_l$ basis functions $b_{j}^{(l)} : \mathcal{X} \to \mathbb{R}$, $l = 1, \ldots, m_j$, for the $j$th covariate effect depending on one or more covariates. The real coefficients can be arranged as a matrix $\{\theta_{j}^{(r,l)}\}_{r,l} = \Theta_{j} \in \mathbb{R}^{m \times m_j}$. Also for infinite-dimensional $T_{[p]} Y_{\mathcal{G}}^\vee$ and a general nonlinear dependence on $x$, a basis representation approach requires truncation to finite dimensions $m$ and $m_j$ in practice. Choosing the bases to capture the essential variability in the data, their size can be extended with increasing data size and computational resources.

While, in principle, the basis $[\partial_r]$, $r$ could also vary across effects $j = 1, \ldots, J$, we assume a common basis for notational simplicity, which presents the typical choice. Due to the identification of $T_{[p]} Y_{\mathcal{G}}^\vee$ with a subspace of the function space $\mathcal{Y}$, the $[\partial_r]$, may be specified using a function basis commonly used in additive models: Let $b_{0}^{(l)} : \mathcal{T} \to \mathbb{R}$, $l = 1, \ldots, m_0$ be a basis of real functions, say a B-spline basis (other typical bases used in the literature include wavelet (Meyer et al. 2015) or FPC bases (Müller and Yao 2008)). Then we construct the tangent space basis as $\partial_r = \sum_{l=1}^{m_0} (\tilde{e}_p^{(l,r)} + z_{p}^{(m_0,l,r)} \sqrt{-1}) b_{0}^{(l)}$,

employing the same basis for the 1- and $\sqrt{-1}$-dimension before transforming it with a basis transformation matrix $Z_p = [\tilde{e}_p^{(l,r)}]_{l,r} \in \mathbb{R}^{2m_0 \times m}$ implementing the linear tangent space constraints $R(e(\partial_r, \zeta^{(r)})) = 0$ (or the empirical version) for all $\partial_r$ and normal vectors $\zeta^{(1)}, \zeta^{(2)}, \zeta^{(3)}$ for forms and additionally $\zeta^{(4)}$ for shapes defining $T_{[p]} Y_{\mathcal{G}}^\vee$ as described in Section 2. Thus, the tangent space basis dimension is $m = 2m_0 - 3$ for forms or $m = 2m_0 - 4$ for shapes (or could, in principle, be larger if the original basis already meets the constraints). For details on the construction of $Z_p$ see Section S.1.3, supplementary materials. For closed curves, we additionally choose $Z_p$ to enforce periodicity, that is, $\partial_r(t) = \partial_r(t + t_0)$ for some $t_0 \in \mathbb{R}$ (compare Hofner, Kneib, and Hothorn 2016).

Given the tangent space basis, we may now modularly specify the usual additive model basis functions $b_{j}^{(l)} : \mathcal{X} \to \mathbb{R}$, $l = 1, \ldots, m_j$, for the $j$th covariate effect to obtain the full functional additive model "tool box" offered by, for example, Brockhaus, Scheipl, and Greven (2015). Typically, $b_{j}^{(l)}(x) = b_{j}^{(l)}(z)$ depending on an individual covariate, say on $z$, in $x = (\ldots, z, \ldots)^T$. But for a single covariate also multiple different effects can be specified and a single interaction effect depends on multiple covariates. A linear effect—linear in the tangent space—of the form $h_j(x) = \beta z$ of a scalar (typically centered) covariate $z$ and $\beta \in T_{[p]} Y_{\mathcal{G}}^\vee$ is simply implemented by a single function $b_j^{(l)}(x) = z$. A smooth effect of the generic form $h_j(x)(t) = f(z, t)$ can be implemented by choosing, for example, a B-spline basis $b_j^{(1)}(z), \ldots, b_j^{(m_j)}(z)$ (asymptotic properties of penalized B-splines and connections to kernel estimators are discussed, for example, by Wood, Pya, and Säfken 2016, Li and Ruppert (2008)). For a categorical covariate $\kappa$ in $x$, with effect $h_j(x) : [1, \ldots, K] \to T_{[p]} Y_{\mathcal{G}}^\vee$, $\kappa \mapsto \beta_\kappa$, the basis $b_j(x) = b_j^{(1)}(\kappa), \ldots, b_j^{(m_j)}(\kappa))^T$ maps $\kappa \mapsto e_\kappa$ to a usual contrast vector $e_\kappa$ with the basis being of dimension $m_j = K - 1$ just as in standard linear models. Here, we typically use effect-encoding to obtain centered effects. Moreover, TP interactions of the model terms described above, as well as group-specific effects and smooth effects with additional constraints (Hofner, Kneib, and Hothorn 2016) can be specified in the model formula, relying on the mboost framework introduced by Hothorn et al. (2010), which also allows to define custom effect designs. For identification of an overall mean intercept $[p]$, sum-to-zero constraints yielding $\sum_{j=1}^{J} h_j(x_i) = 0$ for observed covariates $x_i$ can be specified, and similar constraints can be used to distinguish linear from nonlinear effects and interactions from their marginal effects (Kneib, Hothorn, and Tutz 2009). Different quadratic penalties can be specified for the coefficients $\Theta_{j}$, allowing to regularize high-dimensional effect bases and to balance effects of different complexity in the model fit (see, Section 4).

### 3.2. Tensor-Product Factorization

The multidimensional structure of the response objects makes it challenging to graphically illustrate and interpret additive model terms, in particular when it comes to nonlinear (interaction)
effects, or when effect sizes are visually small. To solve this problem, we suggest to rewrite estimated TP effects $\hat{h}_j$ with estimated coefficient matrix $\hat{\Theta}_j$ as

$$\hat{h}_j(\mathbf{x}) = \sum_{r=1}^{m'_r} \xi^{(r)}_j \hat{h}^{(r)}_j(\mathbf{x})$$

factorized into $m'_r = \min(m_j, m_0)$ components consisting of covariate effects $\hat{h}^{(r)}_j : \mathcal{X} \to \mathbb{R}$, $r = 1, \ldots, m'_r$, in corresponding orthonormal directions $\xi^{(r)}_j \in T_{[p]}\mathcal{Y}^*_G$ with $\langle \xi^{(r)}_j, \hat{h}^{(r)}_j \rangle = 1 (r = l)$, that is, $1$ if $r = l$ and $0$ otherwise. Assuming $E \left( \hat{h}^{(l)}_j(\mathbf{x})^2 \right) < \infty, l = 1, \ldots, m_j$, for the underlying effect basis, the $\hat{h}^{(r)}_j$ are specified to achieve decreasing covariance variances $v^{(r)}_j \geq \cdots \geq v^{(m'_r)}_j \geq 0$ given by $v^{(r)}_j = E \left( \hat{h}^{(r)}_j(\mathbf{x})^2 \right)$. In practice, the expectation over the covariates $\mathbf{X}$ and the inner product $\langle \cdot, \cdot \rangle$ are replaced by empirical analogs (compare Corollary 3, supplementary materials). Due to orthonormality of the $\xi^{(r)}_j$, the component variances add up to the total predictor variance $\sum_{r=1}^{m'_j} v^{(r)}_j = v_j = E \left( \hat{h}_j(\mathbf{x}, \hat{h}_j(\mathbf{x})) \right)$. Moreover, the TP factorization is optimally concentrated in the first components in the sense that for any $l \leq m'_j$ there is no sequence of $\xi^{(r)}_j \in \mathcal{Y}$ and $\hat{h}^{(r)}_j : \mathcal{X} \to \mathbb{R}$, such that $E \left( \| \hat{h}_j(\mathbf{x}) - \sum_{r=1}^{l} \xi^{(r)}_j \hat{h}^{(r)}_j(\mathbf{x}) \|^2 \right) < E \left( \| \hat{h}_j(\mathbf{x}) - \sum_{r=1}^{l} \xi^{(r)}_j \hat{h}^{(r)}_j(\mathbf{x}) \|^2 \right)$, that is, the series of the first $l$ components yields the best rank $l$ approximation of $\hat{h}_j$. The factorization relies on SVD of a (transformed version of) the coefficient matrix $\hat{\Theta}_j$ and the fact that it is well-defined is a variant of the Eckart-Young-Mirsky theorem (proof in Section S.2, supplementary materials).

Particularly when large shares of the predictor variance are explained by the first component(s), the decomposition facilitates graphical illustration and interpretation: choosing a suitable constant $\tau \neq 0$, an effect direction $\xi^{(r)}_j$ can be visualized by plotting the pole representative $p$ together with $\text{Exp}_p(\tau \xi^{(r)}_j)$ on the level of curves, while accordingly rescaled $\frac{1}{\tau} \hat{h}^{(r)}_j(\mathbf{x})$ is displayed separately in a standard scalar effect plot. Adjusting $\tau$ offers an important degree of freedom for visualizing $\xi^{(r)}_j$ on an intuitively accessible scale while faithfully depicting $\xi^{(r)}_j \hat{h}^{(r)}_j(\mathbf{x})$. When based on the same $\tau$, different covariate effects can be compared across the plots sharing the same scale. We suggest $\tau = \max_{\mathbf{x}} \sqrt{v_j}$, the maximum total predictor standard deviation of an effect, as a good first choice.

Besides factorizing effects separately, it can also be helpful to apply TP factorization to the joint additive predictor, yielding

$$h(\mathbf{x}) = \sum_{r=1}^{m'_r} \xi^{(r)} \hat{h}^{(r)}_j(\mathbf{x}) = \sum_{r=1}^{m'_r} \xi^{(r)} \left( \hat{h}^{(r)}_1(\mathbf{x}) + \cdots + \hat{h}^{(r)}_{j}(\mathbf{x}) \right),$$

with $m'_r = \min\{m_j, m\}$ and again $\xi^{(r)} \in T_{[p]}\mathcal{Y}^*_G$ orthonormal and the corresponding variance concentration in the first components, but now determined w.r.t. entire additive predictors $\hat{h}^{(r)}_j = \sum_{j=1}^{J} \hat{h}^{(r)}_j$ spanned by all covariate basis functions in the predictor. In this representation, the first component yields a geodesic additive model approximation where the predictor moves along a geodesic line $c(\tau) = \text{Exp}_{[p]}(\xi^{(1)} \tau)$ with the signed distance $\tau \in \mathbb{R}$ from $[p]$, modeled by a scalar additive predictor $\hat{h}^{(1)}(\mathbf{x})$ composed of covariate effects analogous to the original model predictor. In Section 5, we illustrate its potential in three different scenarios.

### 4. Component-Wise Riemannian $L^2$-Boosting

Component-wise gradient boosting (e.g., Hothorn et al. 2010) is a step-wise model fitting procedure accumulating predictors from smaller models, so called base-learners, to built an ensemble predictor aiming at minimizing a mean loss function. To this end, the base-learners are fit (via least squares) to the negative gradient of the loss function in each step and the best fitting base-learner is added to the current ensemble predictor. Due to its versatile applicability, inherent model selection, and slow over-fitting behavior, boosting has proven useful in various contexts (Mayr et al. 2014). Boosting with respect to the least squares loss function $\ell(y, \mu) = \frac{1}{2} (y - \mu)^2$, $y, \mu \in \mathbb{R}$, is typically referred to as $L^2$-Boosting and simplifies to repeated refitting of residuals $\epsilon = y - \mu = -P \ell(y, \mu)$ corresponding to the negative gradient of the loss function. For $L^2$-Boosting with a single learner, Bühlmann and Yu (2003) show how fast bias decay and slow variance increase over the boosting iterations suggest stopping the algorithm early before approaching the ordinary (penalized) least squares estimator. Lutz and Bühlmann (2006) prove consistence of component-wise $L^2$-Boosting in a high-dimensional multivariate response linear regression setting and Stöcker et al. (2021) illustrate in extensive simulation studies how stopping the boosting algorithm early based on curve-wise cross-validation applies desired regularization when fitting (even highly autocorrelated) functional responses with parameter-intense additive model base-learners and, thus, leads to good estimates even in challenging scenarios.

When generalizing to least squares on Riemannian manifolds with the loss $\frac{1}{2} d^2((y), [\mu])$ given by the squared geodesic distance, the negative gradient $-\nabla_{[\mu]} \frac{1}{2} d^2((y), [\mu]) = \text{Log}_{[\mu]}(y) = \epsilon_{[\mu]}$ (compare e.g., Pennec 2006) corresponds to the local residuals $\epsilon_{[\mu]}$ defined in Section 3. This analogy to $L^2$-Boosting motivates the presented generalization where local residuals are further transported to residuals $\epsilon$ in a common linear space.

Consider the pole $[p]$ known and fixed for now. Assuming its existence, we aim to minimize the population mean loss

$$\sigma^2(h) = E \left( \frac{1}{\gamma(Y)} \right) \text{Exp}_{[p]} \left( h(\mathbf{x}) \right)$$

with the point-wise minimizer $h^*(\mathbf{x}) = \text{argmin}_{h} E \left( \frac{1}{\gamma(Y)} \right)$.

Exp$_{[p]} \left( h(\mathbf{x}) \right) | \mathbf{X} = \mathbf{x}$ minimizing the conditional expected squared distance. Fixing a covariate constellation $\mathbf{x} \in \mathcal{X}$, the prediction $[\mu] = \text{Exp}_{[p]}(h^*(\mathbf{x}))$ corresponds to the Fréchet mean (Karcher 1977) of $[Y]$ conditional on $\mathbf{X} = \mathbf{x}$. In a finite-dimensional context, Pennec (2006) show that $E (\epsilon_{[\mu]}^*) = 0$ for a Fréchet mean $[\mu]$ if residuals $\epsilon_{[\mu]}$ are uniquely defined with probability one. This indicates the connection to our residual based model formulation in Section 3. We fit the model by
reducing the empirical mean loss \( \hat{\sigma}^2(h) = \frac{1}{n} \sum_{i=1}^n d_i^2(\hat{y}_i) \), where we replace the population mean by the sample mean and compute the geodesic distances \( d_i \) with respect to the inner products \( \langle \cdot, \cdot \rangle \) defined for the respective evaluations of \( y_i \).

A base-learner corresponds to a covariate effect \( h_j(x) = \sum_{r,l} b_{j(r)}^l(x) \hat{\theta}_{j(r)}^l \), \( \hat{\Theta}_j = \{ \hat{\theta}_{j(r)}^l \}_{r,l} \) which is repeatedly fit to the transported residuals \( \epsilon_i, \ldots, \epsilon_n \) by penalized least-squares (PLS) minimizing \( \sum_{i=1}^n \| \epsilon_i - h_j(x_i) \|^2 + \lambda_j tr(\hat{\Theta}_j^\top \hat{\Theta}_j) + \lambda tr(\hat{\Theta}_j^\top P \hat{\Theta}_j) \). Via the penalty parameters \( \lambda_j, \lambda \geq 0 \) the effective degrees of freedom of the base-learners are controlled (Hofner et al. 2011) to achieve a balanced “fair” base-learner selection despite the typically large and varying number of coefficients involved in the TP effects. The symmetric penalty matrices \( P_j \in \mathbb{R}^{m_j \times m_j} \) and \( P \in \mathbb{R}^{m \times m} \) (imposing, e.g., a second-order difference penalty for B-splines in either direction) can equivalently be arranged as a \( m_j \times m_j \) penalty matrix \( R_j = \lambda_j (P_j \otimes I_{m_j}) + (\lambda (m_j \otimes P) \) for the vectorized coefficients \( (\Theta_j) = (\theta_{j(1,1)}^1, \ldots, \theta_{j(m,1)}^1, \ldots, \theta_{j(m,m)}^1)^\top \), \( \otimes \) denotes the Kronecker product. The standard PLS estimator is then given by vec \( (\Theta_j) = (\Psi_j + R_j)^{-1} \psi_j \) with \( \psi_j = \sum_{i=1}^n \left\{ \text{Re} \left( b_j^{(l)}(x_i) \hat{\partial}_r \psi_l^{(r)}(x_i) \right) \delta_{i,l} \right\} \) \((r,l) = (1, 1), \ldots, (m, 1), \ldots, (m, m) \in \mathbb{R}^{m \times m} \) and \( \Psi_j = \sum_{i=1}^n \left\{ \text{Re} \left( b_j^{(l)}(x_i) \hat{\partial}_r \psi_l^{(r)}(x_i) \right) \right\} \). In a regular design, using the functional linear array model (Brockhaus, Scheipl, and Greven 2015) can save memory and computation time by avoiding construction of the complete matrices. The basis construction of \( \hat{\partial}_r \) via a transformation matrix \( Z_p \) (Section 3.1) is reflected in the penalty by setting \( P = Z_p^\top (I_2 \otimes P_0) Z_p \) with \( P_0 \) the penalty matrix for the un-transformed basis \( b_j^{(l)}(r) \).

In each iteration of the proposed Algorithm 1, the best-performing base-learner is added to the current ensemble additive predictor \( h(x) \) after multiplying it with a step-length parameter \( \eta \in (0, 1] \). Due to the additive model structure this corresponds to a coefficient update of the selected covariate effect. Accordingly, after repeated selection, the effective degrees of freedom of a covariate effect, in general, exceed the degrees specified for the base-learner. They are successively adjusted to the data. To avoid over-fitting, the algorithm is typically stopped early before reaching a minimum of the empirical mean loss. The stopping iteration is determined, for example, by resampling strategies such as bootstrapping or cross-validation on the level of shapes/forms.

The pole \( p \) is, in fact, usually not a priori available. Instead we typically assume \( p = \text{argmin}_p E \left( d^2([Y], [q]) \right) \) is the overall Fréchet mean, also often referred to as Riemannian center of mass for Riemannian manifolds or as Procrustes mean in shape analysis (Dryden and Mardia 2016). Here, we estimate it as \( p = \text{Exp}_{[p]}(h_0) \) in a preceding Riemannian \( L^2 \)-Boosting routine. The constant effect \( h_0 \in T_{[p]} \mathcal{Y}_G \), in the intercept-only special case of our model is estimated with Algorithm 1 based on a preliminary pole \( p_0 \in \mathcal{Y}_G \). For shapes and forms, a good candidate for \( p_0 \) can be obtained as the standard functional mean of a reasonably well aligned sample \( y_1, \ldots, y_n \in \mathcal{Y} \) of representatives.

The proposed Riemannian \( L^2 \)-Boosting algorithm is available in the R (R Core Team 2018) package manifoldboost (github.com/Almond-S/manifoldboost). The implementation is based on the package FDBOOST (Brockhaus, Rügamer, and Greven 2020), which is in turn based on the model-based boosting package mboost (Hothorn et al. 2010).

## 5. Applications and Simulation

### 5.1. Shape Differences in Astragali of Wild and Domesticated Sheep

In a geometric morphometric study, Pöllath, Schafberg, and Peters (2019) investigate shapes of sheep astragali (ankle bones)
to understand the influence of different living conditions on the
micromorphology of the skeleton. Based on a total of \( n = 163 \)
shapes recorded by Pöllath, Schafberg, and Peters (2019), we
model the astragalus shape in dependence on different vari-
ables, including domestication status (wild/feral/domesticated),
sex (female/male/NA), age (juvenile/subadult/adult/NA), and
mobility (confined/pastured/free) of the animals as categorical
covariates. The sample comprises sheep of four different popu-
lations: Asiatic wild sheep (Field Museum, Chicago; Lay 1967;
Zeder 2006), feral Soay sheep (British Natural History Museum,
London; Clutton-Brock et al. 1990), and domestic sheep of the
Karakul and Marsch breed (Museum of Livestock Sciences,
Halle (Saale); Schafberg and Wussow 2010). Table S1 in Section
S.3, supplementary materials shows the distribution of available
covariates within the populations. Each sheep astragalus shape,
\( i = 1, \ldots, n \), is represented by a configuration composed of 11
selected landmarks in a vector \( Y_{im} \in \mathbb{C}^{11} \) and two vectors of
sliding semi-landmarks \( y_{c1} \in \mathbb{C}^{14} \) and \( y_{c2} \in \mathbb{C}^{18} \) evaluated
along two outline curve segments, marked on a 2D image of the
bone (dorsal view). Several example configurations are displayed in
Figure S1, supplementary materials. In general, we could sepa-
rately specify smooth function bases for the outline segments
\( y_{c1} \) and \( y_{c2} \), respectively. Due to their systematic recording,
we assume, however, that not only landmarks but also semi-
landmarks are regularly observed on a fixed grid, and refrain
from using smooth function bases for simplicity. Accordingly,
shape configurations can directly be identified with their eval-
uation vectors \( y_i = (Y_{im}^T, y_{c1}^T, y_{c2}^T)^T \in \mathbb{C}^{43} = \mathcal{Y} \), and the
geometry of the response space \( \mathcal{Y}^{*} \) widely cor-
responds to the classic Kendall’s shape space geometry, with the
difference that, considering landmarks more descriptive than
single semi-landmarks, we choose a weighted inner product
\( \langle y_i, y'_i \rangle = y_i^T W y'_i \) with diagonal weight matrix \( W \) with diagonal
\((1_{11}^T, 1_{14}^T, 1_{18}^T, 1_{18}^T)^T \) assigning the weight of three landmarks to
each outline segment. We model the astragalus shapes \( y_i \in \mathcal{Y}^{*} \)
\( /\text{Trl} \times \text{Rot} \times \text{Scl} \) as
\[
\mu_i = \exp(\theta) \left( \beta_{\text{status}} + \beta_{\text{pop}}, + \beta_{\text{age}}, + \beta_{\text{sex}}, + \beta_{\text{mobility}}, \right)
\]
with the pole \( \theta \in \mathcal{Y}^{*} \) specified as overall mean and the
conditional mean \( \mu_i \in \mathcal{Y}^{*} /\text{Trl} \times \text{Rot} \times \text{Scl} \) depending on the effect
coded covariate effects \( x_{ij} \mapsto \beta_{x_{ij}} \in \mathcal{Y}^{*} /\text{Trl} \times \text{Rot} \times \text{Scl} \). For
identifiability, the population and mobility effects are centered
around the status effect, as we only have data on different popu-
lations/mobility levels for domesticated sheep. All base-learners
are regularized to one degree of freedom by employing ridge
penalties for the coefficients of the covariate bases \( b^{(i)} \) while the
coefficients of the response basis (the standard basis for \( \mathbb{C}^{43} \) are
left un-penalized. With a step-length of \( \eta = 0.1 \), 10-
fold shape-wise cross-validation suggests early stopping after 89
boosting iterations. Due to the regular design, we can make use of
the functional linear array model (Brockhaus, Scheipl, and
Greven 2015) for saving computation time and memory, which
lead to 8 sec of initial model fit followed by 47 sec of cross-
validation. To interpret the categorical covariate effects, we rely
on TP factorization (Figure 2). The first component of the status
effect explains about 2/3 of the variance of the status effect and
over 50% of the cumulative effect variance in the model. In that
main direction, the effect of feral is not located between wild and
domestic, as might be naively expected. By contrast, the second
component of the effect seems to reflect the expected order
and still explains a considerable amount of variance. Similar to
Pöllath, Schafberg, and Peters (2019), we find little influence of

---

**Figure 2.** Left: Shares of different factorized covariate effects in the total predictor variance. Right: Factorized effect plots showing the two components of the status effect (rows): in the right column, the two first directions \( \xi^{(1)}, \xi^{(2)} \in \mathcal{Y}^{*} /\text{Trl} \times \text{Rot} \times \text{Scl} \) are visualized via line-segments originating at the overall mean shape (empty circles) and ending in the shape resulting from moving 1 unit into the target direction (solid circles; large: landmarks; small: semi-landmarks along the outline); in the left column, the status effect in the respective direction is depicted. As illustrated in the middle plot, an effect of 1 would correspond to the full extend of the direction shown to the right.
age, sex, and mobility on the astragalus shape. Yet, all covariates were selected by the boosting algorithm.

Visually, differences in estimated mean shapes are rather small, which is, in our experience, quite usual for shape data. With differences in size, rotation and translation excluded by definition, only comparably small variance remains in the observed shapes. Nonetheless, TP factorization provides accessible visualization of the effect directions and allows to partially order the effect levels in each direction.

### 5.2. Cellular Potts Model Parameter Effects on Cell Form

The stochastic biophysical model proposed by Thirion et al. (2019), a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian. Different parameters controlling the Hamiltonian have to be calibrated, e.g., a Cellular Potts Model (CPM), simulates migration dynamics of cells (e.g., wound healing or metastasis) in two dimensions. The progression of simulated cells is the result of many consecutive local elementary events sampled with a Metropolis-algorithm according to a Hamiltonian.

To provide additional insights, we model the cell form in dependence on four CPM parameters considered particularly relevant: the bulk stiffness $x_{1\ell}$, membrane stiffness $x_{2\ell}$, substrate adhesion $x_{3\ell}$, and signaling radius $x_{4\ell}$ are subsumed in a vector $x_{\ell}$ of metric covariates for $\ell = 1, \ldots, n$. Corresponding sampled cell outlines $y_{\ell}$ were provided by Sophia Schaffer in the context of Schaffer (2021), who ran underlying CPM simulations and extracted outlines. Deriving the intrinsic orientation of the cells from their movement trajectories, we parameterize $y_{\ell} : [0, 1] \rightarrow \mathbb{C}$, clockwisely relative to arc-length such that $y_{\ell}(0) = y_{\ell}(1)$ points into the movement direction of the barycenter of the cell. Using an average of $k = \frac{1}{n} \sum_{i=1}^{n} k_i \approx 43$ samples per curve (after sub-sampling preserving 95% of their inherent variation, as described in Volkman et al. 2021, supplement), the evaluation vectors $y_{\ell} \in \mathbb{C}^{n}$ are equipped with an inner-product implementing rule integration weights. Example cell outlines are depicted in Figure S4, supplementary materials. The results shown below are based on cell samples obtained from 30 different CPM parameter configurations. For each configuration, 33 out of 10,000 Monte Carlo samples were extracted as approximately independent. This yields a dataset of $n = 990 = 30 \times 33$ cell outlines.

As positioning of their regularly sampled cell outlines $y_{\ell}$, $i = 1, \ldots, n$, in the coordinate system is arbitrary, we model the cell forms $[y_{\ell}] \in Y_{\text{TR} + \text{Rot}}^{*}$. Their estimated overall form mean $[p]$ serves as pole in the additive model

$$[\mu_i] = \exp[p] \left( h(x_i) \right) = \exp[p] \left( \sum_{j} \beta_j x_{ij} + \sum_{j} f_j(x_{ij}) + \sum_{j \neq j'} f_{jj'}(x_{ij}, x_{ij'}) \right)$$

where the conditional form mean $[\mu_i]$ is modeled in dependence on tangent-space linear effects with coefficients $\beta_j \in T(p) \mathbb{V}_{\text{TR}} \times \text{Rot}$ and nonlinear smooth effects $f_j$ for covariate $j = 1, \ldots, 4$ as well as smooth interaction effects $f_{jj'}$ for each pair of covariates $j \neq j'$. All involved (effect) functions are modeled via a cyclic cubic B-spline basis $[b_{0i}^{(r)}]$, with 7 (inner) knots and a ridge penalty, and quadratic P-splines with 4 knots for the covariates $x_{ij}$ equipped with a second-order difference penalty for the $f_j$ and ridge penalties for interactions. Covariate effects are mean centered and interaction effects $f_{jj'}(x_{ij}, x_{ij'})$ are centered around their marginal effects $f_j(x_{ij}), f_{jj'}(x_{ij}, x_{ij'})$, which are in turn centered around the linear effects $\beta_j x_{ij}$ and $\beta_{jj'} x_{ij} x_{ij'}$, respectively. Resulting predictor terms involve 69 (linear effect) to 1173 (interaction) basis coefficients but are penalized to a common degree of freedom of 2 to ensure a fair base-learner selection.

Observing that the most relevant components point into similar directions, we jointly factorize the predictor as $h(x_i) = \sum r \xi^{(r)}(x_i) \tilde{h}^{(r)}(x_i)$ with TP factorization. The first component explains about 93% of the total predictor variance (Figure S3, supplementary materials), indicating that, post-hoc, a good share of the model can be reduced to the geodesic model $[\tilde{\mu}_i] = \exp[p] (\xi^{(1)}(\tilde{h}^{(1)}(x_i)))$ illustrated in Figure 3. A positive effect in the direction $\xi^{(1)}$ makes cells larger and more keratocyte / crossiant shaped, a negative effect—pointing into the opposite direction—makes them smaller and more mesenchymal shaped / elongated. The bulk stiffness $x_{1\ell}$ turns out to present the most important driving factor behind the cell form, explaining over 75% of the cumulative variance of the effects (Figure S2, supplementary materials). Around 80% of its effect are explained by the linear term reflecting gradual shrinkage at the side of the cells with increasing bulk stiffness.

### 5.3. Realistic Shape and form Simulation Studies

To evaluate the proposed approach, we conduct simulation studies for both shape and form regression for irregular curves. We compare sample sizes $n \in \{54, 162\}$ and average grid sizes $k = \frac{1}{n} \sum_{i=1}^{n} k_i \in \{40, 100\}$ as well as an extreme case with $k_i = 3$ for each curve but $n = 720$, that is, where only random triangles are observed (yet, with known parameterization over $[0, 1]$). We additionally investigate the influence of nuisance effects and compare different inner product weights. While important results are summarized in the following, comprehensive visualizations can be found in Section 5.5, supplementary materials.

**Simulation design:** We simulate models of the form $[\mu_i] = \exp[p](\beta_i + f_i(z_1))$ with overall mean $[p]$, a binary effect with levels $\kappa \in \{0, 1\}$ and a smooth effect of $z_1 \in [-60, 60]$. We choose a cyclic cubic B-spline basis with 27 knots for $T(p) \mathbb{V}_{\text{TR}} \mathbb{G}$, placing them irregularly at 1/27-quantiles of unit-speed parameterization time-points of the curves. Cubic B-splines with four regularly placed knots are used for covariates in smooth effects. True models are based on the *bot* dataset from R package *Momocs* (Bonhomme et al. 2014) comprising outlines of 20 beer ($\kappa = 0$) and 20 whiskey ($\kappa = 1$) bottles of different brands.
A smooth effect is induced by the 2D viewing transformations resulting from tilting the planar outlines in a 3D coordinate system along their longitudinal axis by an angle of up to 60 degrees toward the viewer (\(z_1 = 60\)) and away (\(z_1 = -60\)) (i.e., in a way not captured by 2D rotation invariance). Establishing ground truth models based on a fit to the bottle data, we simulate new responses \(y_1, \ldots, y_n\) via residual resampling (Section S.5, supplementary materials) to preserve realistic autocorrelation. Subsequently, we randomly translate, rotate and scale \(y_1, \ldots, y_n\) in \(Y\) somewhat around the aligned shape/form representatives to obtain realistic samples.

The implied residual variance \(\frac{1}{n} \sum_{i=1}^{n} \|\epsilon_i\|^2 = \frac{1}{n} \sum_{i=1}^{n} d_i^2 (\|y_i\|, |\mu_i|)\) on simulated datasets ranges around 105\% of the predictor variance \(\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{h}(x_i)\|^2 = \frac{1}{n} \sum_{i=1}^{n} d_i^2 (\|\mu_i|, |p|)\) in the form scenario and around 65\% in the shape scenario. All simulations were repeated 100 times, fitting models with the model terms specified above and three additional nuisance effects: a linear effect \(\beta z_i\) (orthogonal to \(f_1(z_i)\)), an effect \(f_2\) of the same structure as \(f_1\) but depending on an independently uniformly drawn variable \(z_2\), and a constant effect \(h_0 \in \mathcal{T}|p| \mathcal{Y}$\) to test centering around \(|p|\). Base-learners are regularized to 4 degrees of freedom (step-length \(\eta = 0.1\)). Early-stopping is based on 10-fold cross-validation.

**Form scenario:** In the form scenario, the smooth covariate effect \(f_1\) offers a particularly clear interpretation. TP factorization decomposes the true effect into its two relevant components, where the first (major) component corresponds to the bare projection of the tilted outline in 3D into the 2D image plane and the second to additional perspective transformations (Figure 4). For this effect, we observe a median relative mean squared error \(\text{rMSE}(\hat{h}) = \sum_{i=1}^{n} \|\hat{h}_j(x_i) - h_j(x_i)\|^2 / \sum_{i=1}^{n} \|h(x_i)\|^2\) of about 3.7\% of the total predictor variance for small data settings with \(n = 54\) and \(k = 100\) (5.9\% with \(k = 40\)), which reduces to 1.5\% for \(n = 162\) (for both \(k = 40\) and \(k = 100\)). It is typical for functional data that, from a certain point, adding more (highly correlated) evaluations per curve leads to distinctly less improvement in the model fit than adding further observations (compare, e.g., also Stöcker et al. 2021). In the extreme \(k_1 = 3\) scenario, we obtain an rMSE of around 15\%, which is not surprisingly considerably higher than for the moderate settings above. Even in this extreme setting (Figure 4), the effect directions are captured well, while the size of the effect is underestimated. Rotation alignment based on only three points (which are randomly distributed along the curves) might considerably differ from the full curve alignment, and averaging over these sub-optimal alignments masks the full extend of the effect. Still, results are very good given the sparsity of information in this case. Having a simpler form, the binary effect \(\beta_k\) is also estimated more accurately with an rMSE of around 1.5\% for \(n = 54\), \(k = 100\) (1.9\% for \(k = 40\)) and less than 0.8\% for \(n = 162\) (for both \(k = 40\) and \(k = 100\)). The pole estimation accuracy varies on a similar scale.

**Shape scenario:** Qualitatively, the shape scenario shows a similar picture. For \(n = 40\), we observe median rMSEs of 2.8\% \((n = 54)\) and 2.2\% \((n = 162)\) for \(f_1(z_i)\), and 1.5\% and 0.6\% for the binary effect \(\beta_k\). For \(k = 100\), accuracy is again slightly higher.

**Nuisance effects and integration weights:** Nuisance effects in the model where generally rarely selected and, if selected at all, only lead to a marginal loss in accuracy. The constant effect is only selected sometimes in the extreme triangle scenarios, when pole estimation is difficult. We refer to Brockhaus et al. (2017), who perform gradient boosting with functional responses and a large number of covariate effects with stability selection, for simulations with larger numbers of nuisance effects and further discussion in a related context, as variable selection is not our main focus here. Finally, simulations indicate that inner product weights implementing a trapezoidal rule for numerical integration are important for high accuracy.
integration are slightly preferable for typical grid sizes ($k = 40, 100$), whereas weights of $1/k_i$ equal over all grid points within a curve gave slightly better results in the extreme $k_i = 3$ settings.

All in all, the simulations show that Riemannian $L_2$-Boosting can adequately fit both shape and form models in a realistic scenario and captures effects reasonably well even for a comparably small number of sampled outlines or evaluations per outline.

### 6. Discussion and Outlook

Compared to existing (landmark) shape regression models, the presented approach extends linear predictors to more general additive predictors including also, for example, smooth nonlinear model terms and interactions, and yields the first regression approach for functional shape as well as form responses. Moreover, we propose novel visualizations based on TP factorization that, similarto FPC analysis, enable a systematic decomposition of the variability explained by an additive effect on tangent space level. Yielding meaningful coordinates for model effects, its potential for visualization will be useful also for FAMs in linear spaces and also beyond our model framework, such as we exemplarily illustrate for the nonparametric approach of Jeon and Park (2020) in Section S.8, supplementary materials.

Instead of operating on the original evaluations $y_i \in \mathbb{C}^{k_i}$ of response curves $y_i$, as in all applications above, another frequently used approach expands $y_i$, $i = 1, \ldots, n$, in a common basis first, before carrying out statistical analysis on coefficient vectors (compare Ramsay and Silverman (2005), Morris (2015), and Müller and Yao (2008) for smoothing spline, wavelet or FPC representations in FDA or Bonhomme et al. (2014) in shape analysis). Shape/form regression on the coefficients is, in fact, a special case of our approach, where the inner product is evaluated on the coefficients instead of evaluations (Section S.6, supplementary materials).

The proposed model is motivated by geodesic regression. However, in the multiple linear predictor, a linear effect of a single covariate does, in general, not describe a geodesic for fixed nonzero values of other covariate effects. Or put differently, $\text{ExpExp}_p(h_1 + h_2) \neq \text{ExpExp}_p(h_1) \cdot \text{ExpExp}_p(h_2)$ in general. Thus, hierarchical geodesic effects of the form $\text{ExpExp}_p(h_1) \cdot \text{ExpExp}_p(h_2)$, relevant, i.e., in mixed models for hierarchical/longitudinal study designs (Kim et al. 2017), present an interesting future extension of our model. Moreover, an “elastic” extension based on the square-root-velocity framework (Srivastava and Klassen 2016) presents a promising direction for future research, as do other manifold responses.

### Supplementary Materials

Supplementary material with further details is provided in an online supplement.

### Acknowledgments

We sincerely thank Nadja Pöllath for providing carefully recorded sheep astragalus data and important insights and comments, and Sophia Schaffer for running and discussing cell simulations and providing fully processed cell outlines.

### Disclosure Statement

The authors report there are no competing interests to declare.

### Funding

We gratefully acknowledge funding by grant GR 3793/3-1 from the German research foundation (DFG) and support by the Open Access Publication Fund of Humboldt-Universität zu Berlin.
Pennec, X. (2006), "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements," Journal of Mathematical Imaging and Vision, 25, 127–154. [1605]
Petersen, A., and Müller, H.-G. (2019), "Fréchet Regression for Random Objects with Euclidean Predictors," The Annals of Statistics, 47, 691–719. [1600]
Pigoli, D., Menafoglio, A., and Secchi, P. (2016), "Kriging Prediction for Manifold-Valued Random Fields," Journal of Multivariate Analysis, 145, 117–131. [1600]
Pöllath, N., Schaftberg, R., and Peters, J. (2019), "Astragalar Morphology: Approaching the Cultural Trajectories of Wild and Domestic Sheep Applying Geometric Morphometrics," Journal of Archaeological Science: Reports, 23, 810–821. [1606,1607]
R Core Team (2018), R: A Language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing. [1606]
Ramsay, J. O., and Silverman, B. W. (2005), Functional Data Analysis, New York: Springer. [1600,1610]
Rosen, O., and Thompson, W. K. (2009), "A Bayesian Regression Model for Multivariate Functional Data," Computational Statistics & Data Analysis, 53, 3773–3786. [1601]
Schaftberg, R., and Wussow, J. (2010), "Julius Kühn. Das Lebenswerk eines agrarwissenschaftlichen Visionärs," Züchtungskunde, 82, 468–484. [1607]
Scheipl, F., Staicu, A.-M., and Greven, S. (2015), "Functional Additive Mixed Models," Journal of Computational and Graphical Statistics, 24, 477–501. [1600,1604]
Schiratti, J.-B., Allassonnière, S., Colliot, O., and Durrleman, S. (2017), "A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations," The Journal of Machine Learning Research, 18, 4840–4872. [1600]
Shi, X., Styner, M., Lieberman, J., Ibrahim, J. G., Lin, W., and Zhu, H. (2009), "Intrinsic Regression Models for Manifold-Valued Data," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 192–199, Springer. [1600]
Srivastava, A., and Klassen, E. P. (2016), Functional and Shape Data Analysis, New York: Springer-Verlag. [1601,1610]
Stöcker, A., Brockhaus, S., Schaffer, S. A., Bronk, B. v., Opitz, M., and Greven, S. (2021), "Boosting Functional Response Models for Location, Scale and Shape with an Application to Bacterial Competition," Statistical Modelling, 21, 385–404. [1601,1605,1609]
Stöcker, A., Pfeuffer, M., Steyer, L., and Greven, S. (2022), "Elastic Full Procrustes Analysis of Plane Curves via Hermitian Covariance Smoothing," https://doi.org/10.48550/arXiv.2203.10522 [1603]
Thüroff, F., Guychak, A., Reiter, M., and Frey, E. (2019), "Bridging the Gap between Single-Cell Migration and Collective Dynamics," eLife, 8, e46842. [1608]
Volkmann, A., Stöcker, A., Scheipl, F., and Greven, S. (2021), "Multivariate Functional Additive Mixed Models," Statistical Modelling. https://doi.org/10.1177/1471082X211056158 [1601,1608]
Wood, S. N., Pya, N., and Säfken, B. (2016), "Smoothing Parameter and Model Selection for General Smooth Models," Journal of the American Statistical Association, 111, 1548–1563. [1604]
Zhu, H., Chen, Y., Ibrahim, J. G., Li, Y., Hall, C., and Lin, W. (2009), "Intrinsic Regression Models for Positive-Definite Matrices with Applications to Diffusion Tensor Imaging," Journal of the American Statistical Association, 104, 1203–1212. [1600]
Zhu, H., Li, R., and Kong, L. (2012), "Multivariate Varying Coefficient Model for Functional Responses," Annals of Statistics, 40, 2634–2666. [1601]
Zhu, H., Morris, J. S., Wei, F., and Cox, D. D. (2017), "Multivariate Functional Response Regression, with Application to Fluorescence Spectroscopy in a Cervical Pre-cancer Study," Computational Statistics and Data Analysis, 111, 88–101. [1601]