

MODELING SHAPE DYNAMICS DURING CELL MOTILITY IN MICROSCOPY VIDEOS*

Ximu Deng^{*,1}, Rituparna Sarkar^{*,2}, Elisabeth Labruyere², Jean-Christophe Olivo-Marin² & Anuj Srivastava¹

¹ Department of Statistics, Florida State University, Tallahassee, FL, USA

² Bioimage Analysis Unit, Institut Pasteur, Paris, France

ABSTRACT

Statistical analysis of *shape evolution* during cell migration is important for gaining insights into biological processes. This paper develops a time-series model for temporal evolution of cellular shapes during cell motility. It uses *elastic shape analysis* to represent and analyze *shapes* of cell boundaries (as planar closed curves), thus separating cell shape changes from cell kinematics. Specifically, it utilizes Transported Square-Root Velocity Field (TSRVF), to map non-Euclidean shape sequences into a Euclidean time series. It then uses PCA to reduce Euclidean dimensions and imposes a Vector Auto-Regression (VAR) model on the resulting low-dimensional time series. Finally, it presents some results from VAR-based statistical analysis: estimation of model parameters and diagnostics, synthesis of new shape sequences, and predictions of future shapes given past shapes.

Index Terms— cell migration, shape dynamics, elastic shape model, amoeboid motion

1. INTRODUCTION

Cell migration is a crucial process in various biological phenomenon such as embryonic development, wound healing, cancer metastasis and others. Diverse biological and chemical reactions of the intra-cellular as well as extracellular environments invoke rearrangement of cytoskeleton, which in turn leads to cell migration. This migration usually has two components: (1) kinematics, relating to the overall motion of the cell, and (2) morphology, relating to the change in its shape. Quantifying characteristics of cellular deformation – both kinematic and morphological – can provide interesting insights into the cell migration process. In this paper, we are particularly interested in the motility patterns of amoeba, which exhibits a certain type of cell migration. The amoeboid migration is characterized by formation of small membrane protrusion (*a.k.a.* blebs) with sequential contraction and retraction of the cell membrane, thus making it an interesting example for studying the membrane dynamics involved in cell motility. Furthermore, we are mainly interested in isolating and modeling dynamic evolution of *shapes* of amoeba over time, while discarding their kinematics or global motions.

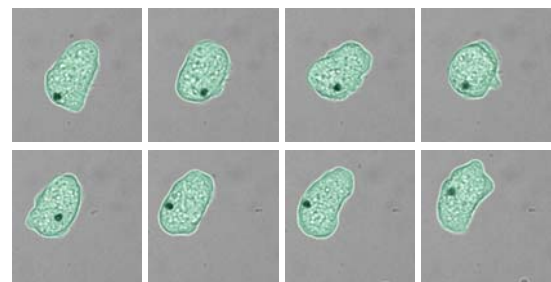


Fig. 1: Segmentations (in green) in bright-field image sequences.

The field of biological shape analysis, using precise mathematical representations and statistical modeling, is an active area of research. While past research has mainly dealt with static shapes [1, 2], our interest lies in dynamics of shape evolution, i.e. How to mathematically and statistically model a shape changing over time? We will consider amoeba as 2D objects represented by their boundaries (simple closed, planar curves) in video frames. We extract these cell boundaries from bright field microscopy images, as these images provide better visibility of cell membrane deformation (see Fig. 1). The cell boundaries are extracted using a deep learning based method [3], tailored specifically to segment cells from bright field microscopy images. An example of segmentation of cells from an image sequence is shown in Fig. 1 (in green). The outer contours extracted from these segmentations are then used in shape modeling.

There exists an extensive literature on quantifying kinematics of single cell motility [4, 5], but few have explored shape evolutions involved in cell motility. A few papers have investigated the dynamics of cell membrane during migration, albeit from a biological perspective [6, 7]. In [8], the authors mainly analyze membrane protrusion characteristics which guide cell motion, while in [9] the authors temporally associate cell boundaries to analyze a wave-like propagation of cell membrane. Tweedy et al. [2] analyze different motility modes adopted by a cell in the migration process using principal components of Fourier shape descriptors. In [1], the authors use various shape features to classify and interpret differences between amoeba populations. However, these works have not accounted for the temporal correlation in shape evolution (a crucial aspect in amoeboid motion), while comput-

*XD AND RS ARE JOINT FIRST AUTHORS OF THIS PAPER.

ing shape features. This calls for employing a more sophisticated model of dynamic shape evolution, a model that is able to characterize complex shapes involved in cell motility.

The main challenge in statistical modeling of shape dynamics comes from infinite-dimensionality and nonlinearity of shape spaces. Classical time series models are usually defined for finite-dimensional Euclidean state spaces. We will use the geometry of shape spaces to flatten nonlinear representations of cell shapes into Euclidean variables, and will then utilize PCA to reach finite-dimensionality. Specifically, we will use a combination of exponential, inverse exponential maps and parallel translations on shape spaces, to represent a shape time series by a Euclidean time series. Then, we model this Euclidean time series using a standard vector autoregressive (VAR) model, and perform parameter estimation. The estimated parameters are then used to analyze and compare shape evolutions across observations.

2. MATHEMATICAL REPRESENTATIONS

Here we present some background material on shape analysis of planar closed contours, and use it to derive Euclidean time-series representations of shape sequences.

2.1. Elastic Shape Analysis of Contours

In order to develop a mathematical representations of shape evolutions, we use the elastic shape theory [10, 11, 12]. In this theory, each closed parametrized curve $\beta : \mathbb{S}^1 \rightarrow \mathbb{R}^2$ is represented by its Square-Root Velocity Function (SRVF) $q : \mathbb{S}^1 \rightarrow \mathbb{R}^2$ given by: $q(t) = \frac{\dot{\beta}(t)}{\sqrt{|\dot{\beta}(t)|}}$. The use of SRVF greatly simplifies shape analysis of curves, especially in imposing invariance to rotation, translation, scaling, and re-parameterization of β . These transformations are shape-preserving and should not influence shape analysis results.

The SRVF q is already invariant to the translation of β . If a curve β is rescaled to have length one, then its SRVF q has $\mathbb{L}^2$ norm of one, i.e. $q \in \mathbb{S}_\infty$, the unit Hilbert sphere. If β is rotated by a matrix $O \in SO(2)$, and re-parameterized by a diffeomorphism γ , then the new SRVF is given by $O(q \circ \gamma)\sqrt{\gamma'}$. Since all of these preserve shape of β , the shape of β is represented by the set $[q] = \{O(q \circ \gamma)\sqrt{\gamma'} | O \in SO(2), \gamma \in \Gamma\}$, that contains SRVFs of all possible rotations and reparameterizations of β . The set of all shapes is denoted by $\mathcal{S} = \{[q] | q \in \mathbb{S}_\infty\}$. $\mathcal{S}$ is an infinite-dimensional, nonlinear space and that limits our ability to perform statistical analysis. (For instance, the average of q_1 and q_2 does not result in the average of their shapes.) However, a number of tools have been developed in past to study shapes as elements of $\mathcal{S}$. Fig. 2 show some examples of finding the shortest paths in $\mathcal{S}$, or geodesics, between cell shapes. This geodesic computation can also be used to compute shooting vectors, defined as follows. Given any two shapes $[q_1]$ and $[q_2]$, let $\psi : [0, 1] \rightarrow \mathcal{S}$ be a geodesic path such that $\psi(0) = [q_1]$ and $\psi(1) = [q_2]$.

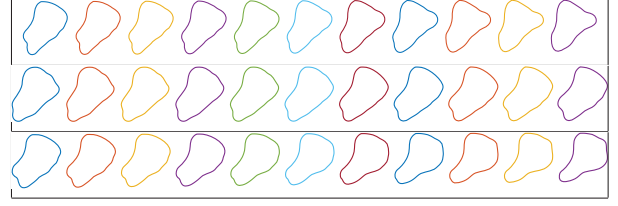


Fig. 2: Examples of shortest paths, or geodesics, between cell shapes treated as elements of $\mathcal{S}$. First and last curves in each row are given and the middle shapes represent geodesic interpolations between them.

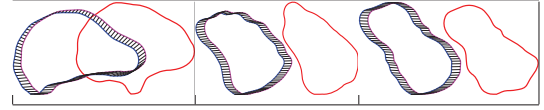


Fig. 3: Examples of computing shape velocities $\dot{\alpha}(\tau)$ in the shape space.

Then, the initial velocity $\dot{\psi}(0) \in T_{[q_1]}(\mathcal{S})$ is called the *shooting vector* from $[q_1]$ to $[q_2]$. These quantities – geodesics, shape distances, and shooting vector – are invariant to rigid motions of curves and their parameterizations, and are instrumental in removing cell kinematics from its morphology.

Given the shape space $\mathcal{S}$, a video of an evolving cell corresponds to a discrete-time shape sequence in $\mathcal{S}$. Next we develop Euclidean representations for such sequences.

2.2. Transported Velocity Fields

Let $\alpha : I \rightarrow \mathcal{S}$ represent a discrete-time evolution of a shape over the observed times $I = \{\tau_1, \tau_2, \dots\}$. What is a useful mathematical representation for comparing and modeling such processes? As stated earlier, the difficulty comes from the nonlinearity and infinite dimensionality of $\mathcal{S}$. So, we use the changes in the shapes, denoted by $\dot{\alpha}$, instead of shapes α themselves, to represent the dynamics. Here $\dot{\alpha}(\tau_i)$ stands for the shooting vector in going from $\alpha(\tau_i)$ to $\alpha(\tau_{i+1})$. Figure 3 shows three examples of computing shooting vectors. In each panel we show two shapes, and the infinitesimal deformation (shooting vector field) on the first shape that takes the first shape to the second.

Analyzing shooting vectors is still difficult because $\dot{\alpha}(\tau) \in T_{\alpha(\tau)}(\mathcal{S})$ are elements of different tangent spaces at different times, and we need to bring them to the same space for comparisons. This is accomplished using parallel transport.

Definition 1 Transported Square-Root Velocity Field (TSRVF) [13, 14]: For a shape sequence $\alpha : I \rightarrow \mathcal{S}$, define its transported square-root velocity field (TSRVF) according to: $F_\alpha(\tau) = (\dot{\alpha}(\tau))_{\alpha(\tau) \rightarrow \alpha(0)} \in T_{\alpha(0)}(\mathcal{S})$, where $\alpha(\tau) \rightarrow \alpha(0)$ denotes the parallel transport of $\dot{\alpha}(\tau)$ from $\alpha(\tau)$ to $\alpha(0)$ along the path α . Similarly, define the integrated TSRVF (I-TSRVF) to be: $H_\alpha(\tau) = \int_0^\tau F_\alpha(\tau) d\tau$.

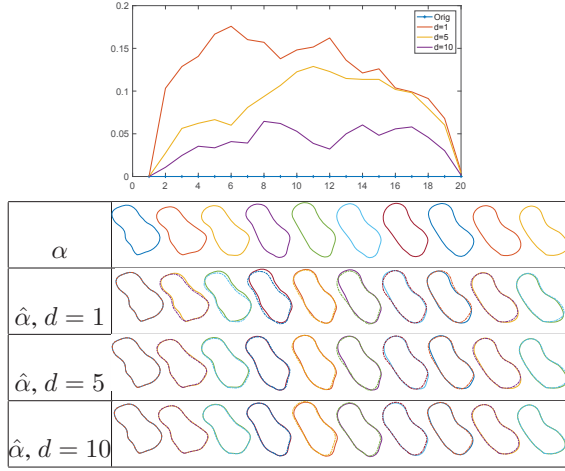


Fig. 4: Reconstruction of a shape sequence α from its Euclidean representations in $\mathbb{R}^d$ for different d . Left panel shows the error associated with each reconstructed shape, while the right column shows the reconstructed sequences.

We will represent a sequence α by either the pair $(\alpha(0), F_\alpha)$ or $(\alpha(0), H_\alpha)$ for the purposes of statistical analysis. These representations are Euclidean and fully invertible. That is, given any one of these pairs, one can reconstruct α using covariant integration and parallel translation.

The TSRVF F_α and I-TSRVF H_α are time series in a vector space $T_{\alpha(0)}(\mathcal{S})$ and can be used for statistical modeling and analysis. Since TSRVF is a parallel transport of the velocity vector, it represents the velocity of the original shape sequence, and since I-TSRVF is an integral of TSRVF it represents the original shapes. The only remaining issue is that these vector fields are still infinite dimensional. Let $\Pi : T_{\alpha(0)}(\mathcal{S}) \rightarrow \mathbb{R}^d$ denote a linear projection to a finite-dimensional subspace $\mathbb{R}^d$ of the tangent space $T_{\alpha(0)}(\mathcal{S})$. This projection can come from PCA of observed TSRVFs, for example. Let $x_\alpha(\tau) = \Pi(F_\alpha(\tau))$ (correspondingly $y_\alpha(\tau) = \Pi(H_\alpha(\tau))$) denote the resulting d -dimensional, Euclidean process associated with the original shape sequence α ; we shall call it TSRVF-PCA representation. This forward mapping is summarized by the steps: $\alpha \in \mathcal{S}^I \rightarrow \hat{\alpha} \in (TS)^I \rightarrow F_\alpha \in T_{\alpha(0)}(\mathcal{S}) \xrightarrow{\Pi} x_\alpha \in (\mathbb{R}^d)^I$. Using appropriate constraints, one can invert Π for use in mapping the Euclidean time series x_α back to the shape space $\mathcal{S}$; call this reconstruct sequence $\hat{\alpha}$. This way one can perform analysis in $(\mathbb{R}^d)^I$ and visualize results in the shape space. Fig. 4 shows reconstruction error $(d_s(\alpha(\tau), \hat{\alpha}_d(\tau)))$ versus τ , for a shape sequence α and different values of d . It also shows some examples $\hat{\alpha}$ for different values of d . The reconstruction error is minimal for $d = 10$, and very high for $d = 1$. The value of $d = 5$ seems to provide a balance between low dimension and reconstruction error. (Similarly, one can repeat this process for y_α and reconstruct α .)

3. VAR MODELS IN TSRVF-PCA SPACE

Now we have Euclidean d -dimensional representations $x_\alpha, y_\alpha : I \rightarrow \mathbb{R}^d$ associated with a shape sequence $\alpha : I \rightarrow \mathcal{S}$. We impose a statistical model on this representation and use it to analyze given amoeba shape sequences. For this, we can use one of several existing models involving auto-regression and moving averages. We will make a notational simplification that time indices are natural numbers $I = \{1, 2, 3, \dots, T_0\}$. A potential model for the resulting discrete-time series is a VAR(p) model:

$$x_\alpha(\tau) = c + \sum_{j=1}^p A_j x_\alpha(\tau - j) + \epsilon_\alpha(\tau), \tau \in I. \quad (1)$$

where: p is the model lag, $c \in \mathbb{R}^d$ is a constant, $A_j \in \mathbb{R}^{d \times d}$ are coefficient matrices, and $\epsilon_\alpha(\tau) \in \mathbb{R}^d$ is the observation noise, modeled as *i.i.d* multivariate normal with mean zero and covariance Σ . Given an observed sequence x_α one can estimate model parameters (c , $\{A_j\}$ and Σ) using the generalized least square error criterion. To facilitate estimation, we rewrite the VAR(p) model as $x_\alpha^T(\tau) = z^T(\tau)\beta + \epsilon_\alpha^T(\tau)$, where the superscript T denotes transpose, $z(\tau) \doteq (1, x_\alpha^T(\tau - 1), \dots, x_\alpha^T(\tau - p))$ and $\beta^T \doteq [c, A_1, \dots, A_p]$. Using a vector-matrix notation, we obtain $\text{vec}(\mathbf{X}) = (\mathbf{I}_k \otimes \mathbf{Z})\text{vec}(\beta) + \text{vec}(\epsilon)$, where $\otimes$ denotes the tensor product. Note that the covariance matrix of $\text{vec}(\epsilon)$ is $\Sigma_\alpha \otimes \mathbf{I}_{T_0-p}$. The least-square estimate of β is obtained by minimizing $S(\beta) = \text{tr}[(\mathbf{X} - \mathbf{Z}\beta)\Sigma_\alpha^{-1}(\mathbf{X} - \mathbf{Z}\beta)^T]$, resulting in: $\hat{\beta} = (\mathbf{Z}^T \mathbf{Z})^{-1}(\mathbf{Z}^T \mathbf{X}) =$

$$\left[\sum_{\tau=p+1}^T z(\tau)z^T(\tau) \right]^{-1} \sum_{\tau=p+1}^T z(\tau)x_\alpha^T(\tau).$$

Note that this estimator does not depend on Σ . Model selection (estimation of p) is performed using an information criteria, such as AIC, BIC, or HQ. AIC usually tends to choose large numbers of lags since it asymptotically overestimates the order with positive probability, whereas BIC and HQ estimate the order consistently under fairly general conditions. Figure 5 shows a plot of BIC values for VAR(p) model versus p for 12 different training sequences. These results show that $p = 3, 4$ usually provides the best model fit to the data.

3.1. Synthesizing Shape Sequences

One of the ways of validating a dynamical model on shapes is to synthesize new sequences from this model and visualize the results. Let $x_\alpha : I \rightarrow \mathbb{R}^d$ represent Euclidean representation of an observed shape sequence α . Let $\hat{\beta}, \hat{\Sigma}$ denote parameters estimated from x_α using generalized least squares, as mentioned above. We can use the model in Eqn. 1, with estimated parameters, to generate a new sequence $\tilde{x}_\alpha : I \in \mathbb{R}^d$, and then map this synthesized sequence back to the original shape space $\mathcal{S}$.

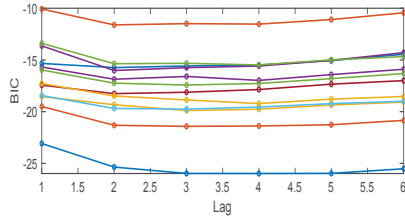


Fig. 5: Model selection: BIC values associated with different model lags (p) for different training sequences.

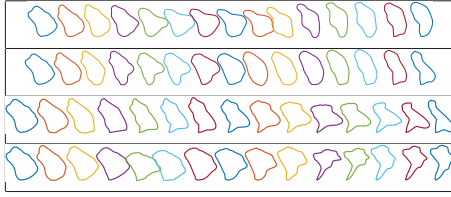


Fig. 6: Synthesis of shape sequences using estimated VAR(1) model: Top row shows the original sequence, and the bottom three row show three random synthesized shape sequences, for $d = 5$.

We demonstrate this idea using an example in Fig. 6. The top row shows the original sequence α whose Euclidean representation is used to fit the VAR(1) model. The bottom three rows show examples of synthesized, random shape sequences, first in $\mathbb{R}^d$ ($d = 5$) and the mapped to $\mathcal{S}$, using the estimated model. Note that $\alpha(0)$ is kept same in all simulations. The realistic nature of synthesized images underline the validity of the proposed dynamical model.

3.2. Predict Future Shapes

Another way to test the shape dynamical model is to predict future shapes using past sequence data. Consider prediction of $x_\alpha(T_0 + h)$, for $h > 0$, using data observed up to time T_0 . The best linear predictor of $x_\alpha(T_0 + h)$ is $\hat{x}_\alpha(T_0 + h|T_0) = \sum_{j=1}^p \hat{A}_j \hat{x}_\alpha(T_0 + h - j|T_0)$, where $\{\hat{A}_j\}$ are the estimated parameter matrices. Note that the h -step prediction also implies predicting all the intermediate shapes from $T_0 + 1$ to $T_0 + h$. If the true sequence is known, then one can quantify the prediction error using: $\frac{1}{h} \sum_{i=1}^h \|x_\alpha(T_0 + i) - \hat{x}_\alpha(T_0 + i|T_0)\|^2$.

We present some experimental results from this prediction using the estimated VAR model. In this experiment, we use several long shape-sequences, each corresponding to a different movie of amoeba motility. In each case we use $d = 5$ to form TSRVF-PCA representation of the shape sequence. The first sequence has time points $\tau = 1, 2, \dots, 179$, so the resulting Euclidean time series $x_\alpha \in \mathbb{R}^{5 \times 178}$. We split the shape sequence into training set ($\tau = 1, 2, \dots, 120$) and testing set ($\tau = 121, \dots, 178$), estimate parameters for VAR(p) model using the training set, and then predict elements of the test sequence recursively. Figure 7 shows some results

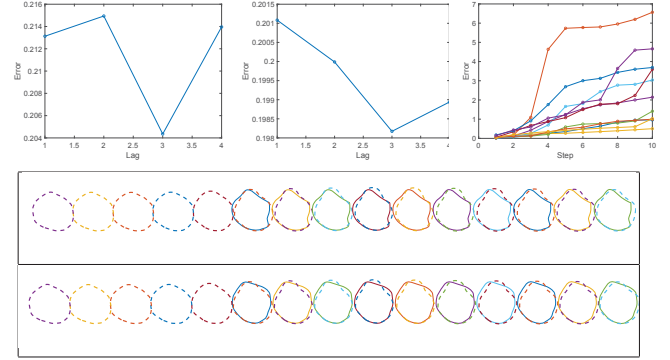


Fig. 7: Top: Prediction errors under the VAR model for different lags p (left is for one sequence and middle is average over 10 sequences) and different step sizes h (right). Bottom: The two rows show predicted sequences associated with VAR model $p = 1, 3$. The solid lines denote predicted shapes and dashed lines denote true data.

from this experiment. The top-left plot shows prediction errors for different values of p in the model on the test part of this sequence. The result indicate that $p = 3$ provides the best results. The bottom two rows show examples of predicted shapes overlaid on the true shape for different values of p . For each sequence, the first five shapes denote the past data (shapes at time $T_0 - 4, T_0 - 3, \dots, T_0$) in dotted lines and the remaining curves denote the predicted shapes (solid lines). By visual inspection, one can see that the prediction error for VAR(3) is smaller than that for VAR(1).

We also compute average prediction error versus p for ten different test shape sequences. This average is shown in the top-middle plot of Fig. 7. Similar to the earlier results, the error is smallest for $p = 3$. We also study the change in prediction error versus the step size h , while keeping $p = 3$ fixed. The resulting curves for individual sequences are in the top right panel. As expected, the prediction error increases with h but the increase is different for different sequences. One can use this prediction error for further statistical analysis on shape prediction.

Lastly, we note that statistical analysis using TSRVF-PCA representation outperform those obtained using the I-TSRVF-PCA representation.

4. SUMMARY

This paper introduces a statistical model for evolution of amoeba shapes during cell motility, isolating shape changes from cell kinematics. It represents a shape sequence by its transported SRVF, which when combined with PCA results in a low-dimensional Euclidean time-series called *TSRVF-PCA*. Then, it imposes a VAR model on TSRVF-PCA to model temporal evolution of shapes. Finally, it discusses estimation of model parameters using least squares and BIC. It also validates this model by synthesizing new shape sequences and using the model to successfully predict future shapes.

5. REFERENCES

- [1] A. Dufour et al., “Signal processing challenges in quantitative 3-d cell morphology: More than meets the eye,” *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 30–40, 2014.
- [2] L. Tweedy et al., “Distinct cell shapes determine accurate chemotaxis,” *Scientific reports*, vol. 3, pp. 2606, 2013.
- [3] R. Sarkar, S. Mukherjee, E. Labruyre, and J.-C. Olivo-Marin, “Learning to segment clustered amoeboid cells from brightfield microscopy via multi-task learning with adaptive weight selection,” *arXiv preprint arXiv: 2005.09372*, 2020.
- [4] D. Campos et al., “Persistent random motion: Uncovering cell migration dynamics,” *Journal of theoretical biology*, vol. 267, no. 4, pp. 526–534, 2010.
- [5] H. Miyoshi et al., “Characteristics of trajectory in the migration of amoeba proteus,” *Protoplasma*, vol. 222, no. 3–4, pp. 175–181, 2003.
- [6] A. Boquet-Pujadas et al., “Bioflow: a non-invasive, image-based method to measure speed, pressure and forces inside living cells,” *Scientific reports*, vol. 7, no. 1, pp. 9178, 2017.
- [7] W. R. Holmes and L. Edelstein-Keshet, “A comparison of computational models for eukaryotic cell shape and motility,” *PLoS computational biology*, vol. 8, no. 12, pp. e1002793, 2012.
- [8] D. Tsygankov et al., “Cellgeo: a computational platform for the analysis of shape changes in cells with complex geometries,” *Journal of Cell Biology*, vol. 204, no. 3, pp. 443–460, 2014.
- [9] M. Driscoll et al., “Cell shape dynamics: from waves to migration,” *PLoS computational biology*, vol. 8, no. 3, 2012.
- [10] S. H. Joshi, E. Klassen, A. Srivastava, and I. H. Jermyn, “A novel representation for riemannian analysis of elastic curves in $\mathbb{R}^n$,” in *Proceedings of IEEE CVPR*, 2007, pp. 1–7.
- [11] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn, “Shape analysis of elastic curves in euclidean spaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1415–1428, 2011.
- [12] A. Srivastava and E. Klassen, *Functional and Shape Data Analysis*, Springer Series in Statistics, 2016.
- [13] Z. Zhang, J. Su, E. Klassen, H. Le, and A. Srivastava, “Rate-invariant analysis of covariance trajectories,” *Journal of Mathematical Imaging and Vision*, vol. 60, no. 8, pp. 1306–1323, 2018.
- [14] Z. Zhang, E. Klassen, and A. Srivastava, “Phase-amplitude separation and modeling of spherical trajectories,” *Journal of Computational and Graphical Statistics*, vol. 27, no. 1, pp. 85–97, 2018.