

A Riemannian Framework for Linear and Quadratic Discriminant Analysis on the Tangent Space of Shapes

Susovan Pal, Roger P. Woods,
UCLA Brain Mapping Center
University of California, Los Angeles
Los Angeles, CA, USA

susovanpal@mednet.ucla.edu, rwoods@ucla.edu

Elizabeth Sowell
Department of Pediatrics
Children’s Hospital Los Angeles
University of Southern California, Los Angeles
Los Angeles, CA, USA
esowell@chla.usc.edu

Suchit Panjiyar
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA, USA

suchit@ucla.edu

Katherine L. Narr, Shantanu H. Joshi
UCLA Brain Mapping Center
University of California, Los Angeles
Los Angeles, CA, USA

narr@ucla.edu, s.joshi@ucla.edu

Abstract

We present a Riemannian framework for linear and quadratic discriminant classification on the tangent plane of the shape space of curves. The shape space is infinite dimensional and is constructed out of square root velocity functions of curves. We introduce the idea of mean and covariance of shape-valued random variables and samples from a tangent space to the pre-shape space (invariant to translation and scaling) and then extend it to the full shape space (rotational invariance). The shape observations from the population are approximated by coefficients of a Fourier basis of the tangent space. The algorithms for linear and quadratic discriminant analysis are then defined using reduced dimensional features obtained by projecting the original shape observations on to the truncated Fourier basis. We show classification results on synthetic data and shapes of cortical sulci, corpus callosum curves, as well as facial midline curve profiles from patients with fetal alcohol syndrome (FAS).

1. Introduction

Often, anatomical features can be compactly represented by curve-based landmarks or boundaries. Depending on the study in question, this practice can have several advantages; i) it is easier for anatomical experts to delineate or identify such curve based features in images or even on 3D parametrized representations of images (surfaces etc.), ii) it is convenient for computational analysis since there is a

large reduction of the feature space that contributes to increased numerical efficiency of representation and analysis, and iii) if the features are easy to identify and trace, they can be quickly used for a first level analysis to determine changes in anatomy due to disease or development.

Briefly, such biological curve-shape based studies are outlined as follows. The white matter morphology of the mid-sagittal cross-section of the corpus callosum in the brain can be represented by a boundary curve whose shape has been implicated in various diseases. The callosal morphology shows abnormalities in fetal alcohol syndrome [4, 18] and in relation to autism spectrum disorders [21]. It has also been implicated in schizophrenia [13] and personality disorders [5]. On the other hand, sulcal curves on the brain are examples of cortical landmarks [9]. With a well established protocol, major sulcal and gyral features (12 ~13 sulci) can be consistently delineated by neuroanatomists [10]. There are other examples of curve shape analysis from biological imaging – cellular shapes from microscopy [2, 1] and echocardial curves from ultrasound images [14].

Most of the above works have explicitly used curve representations of biological objects without focusing on the geometry of the underlying space of such curves, although Wolff et al. [21] use an invariant medial shape representation that can be further used to construct geometric shape spaces. However for most approaches there is a mismatch of the cost function for shape matching and the underlying metric of the space to which these curves belong. Further, most studies have largely focused on population level group differences based on point-wise statistical tests conducted

directly on the matched curves. A recent approach by Bogunovic, Frangi et al. [3] have used the well known LD-DMM framework for the characterization of internal carotid artery curves. They obtained a 70 to 80% cross validation classification rate based on global features (curvatures and torsion ratios etc.) of the curves. While we recognize that curve features should be best tuned according to the application at hand, in this work we adopt a different approach. Our goal here is to construct a classification framework using the full shape of the curves by following a differential geometric approach that exploits the geometry of the shape space. Specifically, in this paper we construct a Fourier basis on the tangent space of the shape space and approximate the shape observations by their tangent vectors parametrized by the basis. In this paper we focus on two simple classifiers – the linear and the quadratic discriminant classifiers, although the framework allows the definition of more advanced classifiers as well.

This paper is organized as follows. Section 2.1 briefly outlines our choice of the curve representation. We use the square root velocity functions (SRVF) [11, 12, 19] to define shapes of curves. We chose the square root velocity formulation for curves as it is an explicit representation and simple to compute. Sections 3 and 4 outline our notion of the mean and covariance on the pre-shape and the shape space respectively. An important ingredient for approximating the tangent space of shapes is the construction of a Fourier basis 4.1, which is used to approximate the tangent space. Finally, section 5 defines the classifier on the tangent space followed by results in section 6 on synthetic data as well as on data with clinical relevance. The novelty of our work has two fronts; a tangent space classification approach defined for this particular geometric representation of curves, and the first attempt to classify fetal alcohol syndrome using the midline curves of 3D facial images. To our knowledge this has not been previously tried or accomplished.

2. Brief background of shape representation

We briefly introduce the square-root velocity function (SRVF), preshape and shape spaces of curves here. For a detailed introduction, please see [11, 19].

2.1. Square Root Velocity representation and Pre-shape spaces of Parametrized curves

Let β be a parametrized curve $\beta : D := [0, 2\pi] \rightarrow \mathbb{R}^n$. We will restrict to those β that are almost everywhere differentiable and their first derivative is in $L^2(D, \mathbb{R}^n)$. For the purpose of studying the shape of β , we will represent it using the square-root velocity function (SRVF) [11, 19] defined as $q : D \rightarrow \mathbb{R}^n$, where $q(t) := \dot{\beta}(t)/\sqrt{\|\dot{\beta}(t)\|}$. This representation includes those curves whose parametrization can become singular in the analysis. Also, for every

$q \in L^2(D, \mathbb{R}^n)$ there exists a curve β (unique up to a translation) such that the given q is the SRV function of that β , given by, $\beta(t) = \int_0^t q(s)\|q(s)\|ds$. To remove scaling variability, we re-scale all curves to be of length 2π . The remaining transformations (rotation, translation, and re-parametrization) will be dealt with differently. The restriction that β is of length 2π translates to the condition that $\int_D \|q(t)\|^2 dt = \int_D \|\dot{\beta}(t)\|^2 dt = 2\pi$. Therefore, the SRV functions associated with these curves are elements of the unit sphere in the Hilbert space $L^2(D, \mathbb{R}^n)$; we will use the notation \mathcal{C}^o to denote this hypersphere.

We will call \mathcal{C}^o the pre-shape space of curves. To impose Riemannian structures on this pre-shape space, we consider their tangent spaces. Since \mathcal{C}^o is a sphere in $L^2([0, 2\pi], \mathbb{R}^n)$, its tangent space at a point q is given by: $T_q(\mathcal{C}^o) = \{v \in L^2([0, 2\pi], \mathbb{R}^n) | \langle v, q \rangle = 0\}$. Here $\langle v, q \rangle$ denotes the inner product in $L^2([0, 2\pi], \mathbb{R}^n)$: $\langle v, q \rangle = \int_0^{2\pi} (v(t) \cdot q(t))_{\mathbb{R}^n} dt$.

The standard metric on $L^2([0, 2\pi], \mathbb{R}^n)$ restricts to one on \mathcal{C}^o . This metric can then be used to determine geodesics and geodesic lengths between elements of these spaces, which is straightforward, \mathcal{C}^o being a hypersphere owing to the global scaling constraint.

2.2. Shape Spaces of Parametrized Curves

By representing a parametrized curve $\beta(t)$ by its SRV function $q(t)$, and imposing the constraint $\int_D \langle q(t), q(t) \rangle dt = 2\pi$, we have taken care of the translation and the scaling variabilities, but the rotation and the re-parametrization variabilities still remain. A rotation is an element of $SO(n)$, the special orthogonal group of $n \times n$ matrices, and a re-parametrization is an element of Γ , the set of all orientation-preserving diffeomorphisms of D .

The rotation and re-parameterization of a curve β are denoted by the actions of $SO(n)$ and Γ on its SRV. While the action of $SO(n)$ is the usual: $SO(n) \times \mathcal{C}^o \rightarrow \mathcal{C}^o$, $(O, q(t)) = Oq(t)$, the action of Γ is derived as follows. For a $\gamma \in \Gamma$, the composition $\beta \circ \gamma$ denotes its re-parameterization; the SRV of the re-parametrized curve is $F(\dot{\beta}(\gamma(t))\dot{\gamma}(t)) = q(\gamma(t))\sqrt{\dot{\gamma}(t)}$, where q is the SRV of β . This gives use the action $\Gamma \times \mathcal{C}^o \rightarrow \mathcal{C}^o$, $(q, \gamma) = (q \circ \gamma)\sqrt{\dot{\gamma}}$. It can be shown that i) the actions of $SO(n)$ and Γ on \mathcal{C}^o commute, and that ii) the action of the product group $\Gamma \times SO(n)$ on \mathcal{C}^o is by isometries with respect to the chosen metric.

Therefore, under the actions of O and Γ , we can define the quotient space of \mathcal{C}^o modulo $\Gamma \times SO(n)$. The orbit of a function $q \in \mathcal{C}^o$ is given by: $[q] = \{O(q \circ \gamma)\sqrt{\dot{\gamma}} | (\gamma, O) \in \Gamma \times SO(n)\}$. In this framework, an orbit is associated with a shape and comparisons between shapes are performed by comparing the orbits of the corresponding curves and, thus, the need for

a metric on the set of orbits. In order for this set to inherit the metric from \mathcal{C}^o , we need the orbits to be closed sets in \mathcal{C}^o . Since these orbits are not closed in \mathcal{C}^o , we replace them by their closures in $L^2(D, \mathbb{R}^n)$. With a slight abuse of notation, we will call these orbits $[q]$. Then, the quotient space \mathcal{S} is defined as the set of all such closed orbits associated with the elements of \mathcal{C}^o , i.e. $\mathcal{S} = \{[q] | q \in \mathcal{C}^o\}$. The differential of this quotient map from \mathcal{C}^o to \mathcal{S} induces a linear isomorphism between $T_{[q]}(\mathcal{S})$ and, the normal space to $[q]$ at any point \tilde{q} in $[q]$. The Riemannian metric on \mathcal{C}^o restricts to an inner product on the normal space which, in turn, induces an inner product on $T_{[q]}(\mathcal{S})$. The fact that $\Gamma \times SO(n)$ act by isometries implies that the resulting inner product on $T_{[q]}(\mathcal{S})$ is independent of the choice of $\tilde{q} \in [q]$. In this manner, \mathcal{S} inherits a Riemannian structure from \mathcal{C}^o . See [11, 19] for computation of geodesics in \mathcal{S} .

3. Statistics on vector subspaces of $L^2([0, 2\pi], \mathbb{R}^3)$

From this section onwards, for practical purposes, we will assume that $n = 3$. Let $V \subset L^2([0, 2\pi], \mathbb{R}^3)$ be a subspace of the Hilbert space $L^2([0, 2\pi], \mathbb{R}^3)$. We will see in the next section that V will be mainly $T_\mu \mathcal{S}$ for this paper, which has been shown to be a subspace of $L^2([0, 2\pi], \mathbb{R}^3)$ [11, 19]. Here, we introduce the concepts of mean and covariance for V -valued data, and their discretized versions for computer implementation. This is done keeping in mind that, in the future, the pre-shape data will be represented by elements in $L^2([0, 2\pi], \mathbb{R}^3)$, and their tangent space will be represented by V , so we can carry the concepts in this section over to shape data. Although the shape curve data is typically continuous, being elements of $L^2([0, 2\pi], \mathbb{R}^3)$, for the purpose of computer implementation, we deal with discretized versions of the shape data, where each datum is considered to be an ordered set of points on elements of $L^2([0, 2\pi], \mathbb{R}^3)$. With this in mind, we define three related concepts of mean and covariance: i) mean and covariance of a random variable $X : \Omega \rightarrow V \subset L^2([0, 2\pi], \mathbb{R}^3)$, ii) mean and covariance of a sample of parameterized curves $\{v_1, \dots, v_N\} \subset V$ generated by the random variable X , and finally, iii) mean and covariance of discretized versions of the curves $\{v_1, \dots, v_N\}$, where each v_j is measured at m equispaced time intervals dividing $[0, 2\pi]$. This is done for the purpose of discretization as we only consider finite number of equispaced points on the curves $\{v_1, \dots, v_N\} \subset L^2([0, 2\pi], \mathbb{R}^3)$.

3.1. Mean and covariance of a $V \subset L^2([0, 2\pi], \mathbb{R}^3)$ -valued random variable:

Let $X : \Omega \rightarrow V \subset L^2([0, 2\pi], \mathbb{R}^3)$ be a random variable with probability measure \mathcal{P} on Ω . Choose an orthonormal Schauder basis (we will interchangeably call it a Fourier basis) W for $\{w_1, w_2, \dots, w_k, \dots\}$ for V . Let

$X(\omega) = \sum_{j=1}^{\infty} X_j(\omega) w_j \forall \omega \in \Omega$, where each $X_j = \langle X, w_j \rangle_V$ is a real-valued random variable, where $\langle \cdot, \cdot \rangle_V$ is the induced inner product on V from $L^2([0, 2\pi], \mathbb{R}^3)$. We define mean or expected value of X , denoted by $E(X)$ as $E(X) := \sum_{j=1}^{\infty} E(X_j) w_j \in V$. Next, we define covariance of X by the linear operator on V given by: $\text{cov}(X)(w_j) := \sum_{k=1}^{\infty} \text{cov}(X_j, X_k) w_k$, and then linearly extending it. Hence, with respect to the orthonormal Schauder basis W , we have the infinite matrix representation for $\text{cov}(X)$ given by $\text{cov}(X)(i, j) := \text{cov}(X_i, X_j) = \text{cov}(\langle X, w_i \rangle, \langle X, w_j \rangle)$.

3.1.1 Motivation for using a Fourier basis

A Fourier basis for V is, by definition, a Schauder basis [16], which is also orthonormal. This orthonormality makes the representation of covariance much easier. The covariance matrix for a random vector $X = (X_1, X_2, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ has the (i, j) -entry $\text{cov}(X_i, X_j) := \text{cov}(\langle X, e_i \rangle, \langle X, e_j \rangle)$. Here e_i denotes the orthonormal basis element $(0, \dots, 1, \dots)$ with 1 in the i -th slot. This matrix can also be identified by the linear map on \mathbb{R}^d and denoted by $\text{cov}(X)$, so that $\text{cov}(X)(e_i) := \sum_j \text{cov}(\langle X, e_i \rangle, \langle X, e_j \rangle) e_j$. Note that the above definition of covariance uses the pairwise covariance between $X_i := \langle X, e_i \rangle$. There are two motivating criteria for choosing a Fourier basis. For an arbitrary vector space V with an orthonormal Fourier basis $\{w_i\}$, the above definition of covariance as an operator on V is realized by replacing e_i by w_i . This is explained in detail in the next subsection (Sec. 3.2). Secondly and importantly, the lower order harmonics of the Fourier basis capture the gradual changes, whereas higher order terms capture details and minutiae. Restricting the basis to finite lower order terms also enables a natural smoothing of functions in the underlying space, which makes the Fourier basis an attractive choice.

3.2. Mean and covariance of a finite discrete sample lying in $V \subset L^2([0, 2\pi], \mathbb{R}^3)$:

Let $\{v_1, \dots, v_n\}$ be a sample generated in that order by the above random variable $X : \Omega \rightarrow V \subset L^2([0, 2\pi], \mathbb{R}^3)$. We define the sample mean by $m := \frac{1}{n} \sum_{i=1}^n v_i$, where the addition takes place in the vector space V . One can see the connection between the sample mean and the (population) mean or expectation of a random variable X as follows: divide the sample space Ω into n number of sub-sample spaces $\{\Omega_1, \dots, \Omega_n\}$ so that $\mathcal{P}(\Omega_j) = \frac{1}{n} \forall 1 \leq j \leq n$, assume that X takes constant values v_j on Ω_j , and then calculate its expected value $E(X) = \int_{\Omega} X d\mathcal{P} = \sum_{j=1}^n \int_{\Omega_j} v_j d\mathcal{P} = \sum_{j=1}^n v_j \int_{\Omega_j} d\mathcal{P} = \sum_{j=1}^n v_j \mathcal{P}(\Omega_j) = \frac{1}{n} \sum_{j=1}^n v_j$, since $\mathcal{P}(\Omega_j) = \frac{1}{n} \forall 1 \leq j \leq n$.

Before defining the sample covariance in V , we revisit the definition of the same in standard \mathbb{R}^d -valued dimensional

data with d features. Recall that, in order to define covariance of a sample of n observations $\{v_1, v_2, \dots, v_n\}$, each with d number of features, we write the (ordered) data matrix in X two ways: $X = [v_1, v_2, \dots, v_n] = [R_1, \dots, R_n]^T$, taking the j -th observation v_j as its j -th column, and i -th feature R_i as its i -th row. Let $\bar{X} := (\bar{R}_1, \dots, \bar{R}_n)^T \in \mathbb{R}^{d \times 1}$, where $\bar{R}_i := \frac{1}{n} R_{ij}$. Note that, \bar{X} is nothing but the sample mean m . Let $\tilde{X} := \bar{X}(1, 1, \dots, 1) \in \mathbb{R}^{d \times n}$. Next, we define the (unbiased) covariance matrix to be the $d \times d$ matrix $\text{cov}(X) := \frac{1}{n-1}(X - \tilde{X})(X - \tilde{X})^T$.

Now, to generalize the above definition to the case of V -valued data, we need to be careful: firstly, V may not be a finite dimensional vector space, and secondly, V does not have the canonical inner product structure of \mathbb{R}^d giving orthogonal canonical directions of the form $(0, \dots, 0, 1, 0, \dots, 0)$, hence the feature directions may be unclear and may depend on the basis chosen. The first difficulty cannot be theoretically overcome as the covariance matrix must have the dimensions corresponding to the number of linearly non-redundant features. We attempt to solve the second difficulty simultaneously by defining features so that the feature directions (but not the feature values) are mutually orthogonal, and so that corresponding feature values are the projection of each shape data into the corresponding feature's direction. An obvious choice is to express each $a \in V$ by $a = \sum_{j=1}^{\infty} a_j w_j$, declare $w_j \in V$ as the j -th feature direction of a , and the corresponding feature value to be the j -th projection $a_j := \langle a, w_j \rangle_V$. This way, the data matrix X of the ordered sample $[v_1, v_2, \dots, v_n]$, as written in the previous paragraph, becomes an ∞ by n matrix with respect to the orthonormal Schauder or Fourier basis $W = \{w_1, w_2, \dots, w_i, \dots\}$ of V , and is written in two ways: $X = [C_1, C_2, \dots, C_j, \dots, C_n] = [R_1, R_2, \dots, R_i, \dots]^T$, where the j -th column C_j is so that its i -th entry, or equivalently the (i, j) -th data matrix entry R_{ij} is $\langle v_j, w_i \rangle_V$. Next, we define $\bar{R}_i := \frac{1}{n} \sum_{j=1}^n R_{ij} \forall 1 \leq i \leq \infty$, $\bar{X} := [\bar{R}_1, \dots, \bar{R}_\infty]^T \in \mathbb{R}^{\infty \times 1}$, and $\tilde{X} := \bar{X}(1, 1, \dots, 1) \in \mathbb{R}^{\infty \times n}$. Then, we can represent $\text{cov}(X) : V \rightarrow V$ in this Fourier basis by the ∞ by ∞ matrix given by $\text{cov}(X) = \frac{1}{n-1}(X - \tilde{X})(X - \tilde{X})^T$.

3.3. Mean and covariance of a finite set of sample curves in $V \subset L^2([0, 2\pi], \mathbb{R}^3)$ discretized at $(m+1)$ points $t_r, 0 \leq r \leq m$:

So far, the elements of $V \subset L^2([0, 2\pi], \mathbb{R}^3)$ have been treated as continuous curves. For example, in the previous paragraph, $\langle v, w_j \rangle_V$ is the standard Euclidean inner product between the vector valued functions v and w . For real data we would only have discretized values of the sample v , discretized say at $(m+1)$ time points $t_r, 0 \leq r \leq m$ on the interval $[0, 1]$. We are also going to assume here that $t_r = r/m, 0 \leq r \leq m$. With this assumption, the vector $v_j \in V$ becomes the infinite column vector C_j as above, but the k -th entry of C_j or equivalently the data matrix en-

try X_{jk} is no more $\langle v_j, w_k \rangle_V = \int_0^1 \langle v_j(t), w_k(t) \rangle_{\mathbb{R}^3} dt$, but rather its discrete version $\sum_{r=0}^m \langle v_j(t_r), w_k(t_r) \rangle_{\mathbb{R}^3}$. Choose a K big enough, and hence X becomes a $K \times n$ matrix, where one can think of this K as the first K most "prominent" features. This completes the computation of the data matrix X , and subsequently the $K \times K$ covariance matrix $\frac{1}{n-1}(X - \tilde{X})(X - \tilde{X})^T$.

4. Measures of central tendency and deviation on Curve Shapes

In this section, we describe how to compute the means and covariances of data lying on shape manifold that occurs in square root velocity framework [11, 19]. But first we briefly remind the reader of the calculation of the population mean of a sample of shapes in \mathcal{S} . Given a set of observations $\{q_i\}, i = 1, \dots, N$, their Kärcher mean [15] is defined as

$$\mu = \underset{q}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^N \underset{O, \gamma}{\operatorname{argmin}} d(q, \sqrt{\gamma} O(q_i \circ \gamma))^2, i = 1, \dots, N, \quad (1)$$

where $d(\cdot, \cdot)$ is given by the geodesic distance (great circle lengths) on \mathcal{C}^o . This geodesic is given by the exponential map $\exp_{q_1}(f) = q_1 \cos \alpha + f \sin \alpha$, where $\cos \alpha = \langle q_1, q_2 \rangle$ and the initial tangent vector is given by $f = q_2 - \langle q_1, q_2 \rangle$.

One standard way is to construct and better understand the elements of the tangent spaces $V = T_\mu \mathcal{S}$ of the shape spaces, and replicate in the vector space V the mean and covariance calculations as described in the previous two subsections. This calculation will be valid, of course if we can show that V is indeed a subspace of $L^2([0, 2\pi], \mathbb{R}^3)$, and find an orthonormal Schauder or Fourier basis for it.

4.1. Construction of a Fourier basis on the tangent space of the shape space

To describe the elements of $T_\mu \mathcal{S}$, we first note that we have the following orthogonal direct sum: $T_q \mathcal{C}^o = T_\mu \mathcal{S} \oplus T_q \mathcal{O}(q)$, where the preshape $q \in \mathcal{C}^o$ is a point of its orbit $\mathcal{O}(q)$ so that μ is the equivalence class $[q]$ of q , under the action of $SO(3) \times \mathcal{D}$. Also, since \mathcal{C}^o is a hypersphere in $L^2([0, 2\pi], \mathbb{R}^3)$, it is also clear that $T_q \mathcal{C}^o = q^\perp \subset L^2([0, 2\pi], \mathbb{R}^3)$. Hence in order to construct a basis for $T_\mu \mathcal{S}$, we first focus on doing the same for $T_q \mathcal{O}(q)$, the computation of which follows shortly after. Once this is done, it is straightforward to construct a basis for $T_\mu \mathcal{S}$, since the construction of a basis for $T_q \mathcal{C}^o = q^\perp$ is easy, from that of $L^2([0, 2\pi], \mathbb{R}^3)$.

Below we present a general form of an element in $T_q \mathcal{O}(q)$, which in turn helps us to construct a basis for the same. Let c_ϵ be a curve in $\mathcal{O}(q)$ parametrized by ϵ , so that $c_0 = q = Id \circ (q \circ Id) \cdot \sqrt{Id'}$, where the first Id is in $SO(3)$, and the second and third ones are in \mathcal{D} . By

definition of the tangent space, $T_q\mathcal{O}(q)$ consists of the elements $\frac{\partial}{\partial\epsilon}|_{\epsilon=0}c_\epsilon$.

$$\begin{aligned} & \frac{\partial}{\partial\epsilon}|_{\epsilon=0}c_\epsilon \\ &= \frac{\partial}{\partial\epsilon}|_{\epsilon=0}A_\epsilon(q \circ \gamma_\epsilon) \cdot \sqrt{\gamma'_\epsilon} \\ &= \frac{\partial}{\partial\epsilon}|_{\epsilon=0} \partial_\epsilon(q \circ \gamma_0) \cdot \sqrt{\gamma'_0} + A_0 \cdot \frac{\partial}{\partial\epsilon}|_{\epsilon=0}(q \circ \gamma_\epsilon) \cdot \sqrt{\gamma'_0} + \\ & A_0 \cdot (q \circ \gamma_0) \cdot \frac{\partial}{\partial\epsilon}|_{\epsilon=0} \sqrt{\gamma'_\epsilon} \text{ (taking derivative of product of} \\ & \text{functions)} \\ &= Bq + \frac{\partial}{\partial\epsilon}|_{\epsilon=0}(q \circ \epsilon) + q \cdot \frac{\partial}{\partial\epsilon}|_{\epsilon=0} \sqrt{\gamma'_\epsilon}, \text{ where} \\ & B \in \text{Lie}(\text{SO}(3)) \\ &= Bq + (q' \circ \gamma_0) \frac{\partial}{\partial\epsilon}|_{\epsilon=0} \gamma_\epsilon + q \cdot \frac{\partial}{\partial\epsilon}|_{\epsilon=0} \sqrt{\frac{\partial}{\partial t} \gamma_\epsilon(t)} \\ &= Bq + q' \cdot g + q \cdot \frac{1}{2\sqrt{\gamma'_0}} \frac{\partial}{\partial t} \frac{\partial}{\partial\epsilon}|_{\epsilon=0} \gamma_\epsilon(t) \text{ (where } g \in \text{Lie}(\mathcal{D})) \\ &= Bq + q'g + \frac{1}{2}qg' \end{aligned}$$

Algorithm 1 Construction of the data matrix for discretized shapes

- 1: Find the Kärcher mean μ for a set of shapes $\{q_i\}, i = 1, \dots, N$ using Eqn. 1. Find the corresponding tangent vectors from μ to all q_i s be $\{v_i\}$ following Sec. 4.
- 2: Let B be a discrete d -dimensional orthogonalized Schauder basis of $L^2([0, 2\pi], \mathbb{R}^3)$, where $B \equiv \{b_i\}, i = 1, \dots, d$
- 3: Project B on the tangent space $T_\mu(\mathcal{C}^o)$ by $\tilde{B} = \{\tilde{b}_i\}$, where $\tilde{b}_i = b_i - \langle b_i, \mu \rangle \mu, i = 1, \dots, d$.
- 4: Orthogonalize \tilde{B} w.r.t. the inner product given in Sec. 2.1 as $\tilde{B} = \text{orth}(\tilde{B})$.
- 5: Let G_μ be a discrete k -dimensional orthogonal basis of $T_\mu\mathcal{O}(\mu)$ from Sec. 4.1.
- 6: Construct the basis of $T_\mu(\mathcal{S})$ as G , where $G \equiv \{g_i\}$, $g_i = \tilde{b}_i - \sum_{j=1}^k \langle \tilde{b}_i, g_j \rangle g_j, i = 1 \dots d$.
- 7: Finally construct the data matrix as $a_{ij} = \langle v_i, g_j \rangle$, for $i = 1 \dots d, j = 1 \dots N$.

5. Classification problem of shape data: Linear and Quadratic Classifiers of Shapes

Data classification problems fall into the category of a supervised machine learning problems, where we construct classifiers from the given training data, to determine the classification of a (future) test dataset, based on the previously obtained decision boundary of the classifier. One way to perform the above classification is to construct a discriminant classifier using the Bayesian à posterior probability and normality assumption on data according to the central limit theorem. The discriminant classifier is quadratic in general, and becomes linear if the two given classes of data have equal covariance matrices. Keeping this in mind, when the two data sets have almost equal covariances, we perform linear discriminant analysis (LDA) instead of quadratic discriminant analysis(QDA).

Algorithm 2 Algorithm for computing Decision Boundary in QDA/LDA on shape data

- 1: Represent each shape for each training set as a $3 \times (m+1)$ matrix.
- 2: Call the two classes A and B . Let $X := A \cup B$.
- 3: Compute the Kärcher mean $\mu = \mu_X$ of all the matrices (shapes) of X using the Riemannian metric structure of S , the shape space containing all the data. This Kärcher mean is also a $3 \times (m+1)$ matrix.
- 4: Lift each data A_i in A and each data B_i in B to the tangent space $T_{\mu_X}S$ by the inverse of the exponential map from an open subset of $T_\mu S$ to S . Call the lifting map the *logarithm map*. To do this, find the geodesics from μ_X to each A_i in A and each B_i in B in the shape space S , and compute the corresponding velocity vectors in $T_{\mu_X}S \equiv T_\mu S$. These vectors have the same dimension for each data, i.e. each velocity vector is a $3 \times (m+1)$ matrix.
- 5: By slight abuse of notation, call the lift of each $A_i, B_i \in S$ to $T_{\mu_X}S$, as A_i, B_i again. For the rest of the algorithm all A_i, B_i will lie only in $V := T_{\mu_X}S$.
- 6: (Schauder basis construction for $T_q\mathcal{O}(q)$): Following the discussion in section 4.1, compute the Schauder basis $\{g_i\}$ for $T_q\mathcal{O}(q)$.
- 7: (Orthonormal Schauder basis or Fourier basis construction for V): Following the discussion in section 3.2, 3.3, compute the Fourier basis $\{w_k\}$ for $T_\mu S$.
- 8: Use Algorithm 1 to construct the two data matrices X_A, X_B for A, B respectively, having dimensions $K \times N_1, K \times N_2$ respectively.
- 9: (Construction of sample means) Determine the means m_A, m_B of classes A, B as by averaging the columns of X_A, X_B respectively. Define $\tilde{m}_A := m_A(1, 1, \dots, 1), \tilde{m}_B := m_B(1, 1, \dots, 1)$.
- 10: (Construction of sample covariance matrices for discretized shape data in classes A and B): Form the corresponding $K \times K$ covariance matrices $\Sigma_A = \frac{1}{N_1-1}(X_A - \tilde{m}_A)(X_A - \tilde{m}_A)^T, \Sigma_B := \frac{1}{N_2-1}(X_B - \tilde{m}_B)(X_B - \tilde{m}_B)^T$.
- 11: (Estimation of prior probability of classes) $\pi_A := \frac{N_1}{N}, \pi_B := \frac{N_2}{N}$.
- 12: (Quadratic/Linear discriminant functions) Compute: $\delta_A(x) := -0.5\ln(\det\Sigma_A) - 0.5(x - m_A)^T \Sigma_A^{-1}(x - m_A) + \ln(\pi_A) : \mathbb{R}^K \rightarrow \mathbb{R}$ and $\delta_B(x) := -0.5\ln(\det\Sigma_B) - 0.5(x - m_B)^T \Sigma_B^{-1}(x - m_B) + \ln(\pi_B) : \mathbb{R}^K \rightarrow \mathbb{R}$.
- 13: (Quadratic/Linear Decision Boundary) Compute the set: $\{x \in \mathbb{R}^K : \delta_A(x) = \delta_B(x)\}$.

In standard settings, the discriminant analysis is performed when the data lie in some Euclidean space \mathbb{R}^d with its canonical inner product. However, it goes without saying

that the theory is not straightforward for shape-valued data, since i) shape data generally do not lie in vector spaces, and ii) even after using the tangent spaces of the shape spaces to linearize the problem and after truncation of the tangent spaces to make the problem finite dimensional, the inner product of these tangent spaces mostly have non-Euclidean inner products. Here is where the construction of orthonormal Schauder/Fourier basis (Sec. 3.1.1) comes in, which reduces the problem to one in a finite dimensional Euclidean space, following the discussion of the subsections 3.2, 3.3, and 4.1. This allows us to perform standard machine learning techniques on the shape data.

5.1. Algorithmic construction of decision boundary for two classes of shape data

Before performing discriminant analysis, we would like to represent our shapes as features in a matrix, also known as the data matrix. This is achieved as follows. Given a population of N shapes, a mean shape μ and tangent vectors from the mean shape to each of the shapes of the population are computed. Then a tangent space at μ is formed and approximated using an orthonormal Schauder basis. Each tangent vector observation is then projected into this basis space. The coefficients of the projections are used to construct the data matrix. Algorithm 1 shows a procedure to construct a data matrix for a collection of shapes.

We finally outline the procedure for QDA and LDA to compute the decision boundaries for the two classes in Algorithm 2. Our algorithm is a generic one that can be used for both QDA and LDA, since in step 12, the quadratic terms cancel out for the case of equal covariance matrices, producing a linear decision boundary in step 13.

6. Data, Experiments, and Results

In this section we first present results on synthetic data and subsequently show classification results on data from three examples, i) cortical sulcal shapes, ii) a population of corpus callosum shapes of healthy controls and patients with schizophrenia, and iii) facial midline profiles for healthy individuals and patients with severe alcohol exposure from a fetal alcohol syndrome cohort.

6.1. Synthetic Data

Figure 1 shows examples of two classes of synthetic shape data consisting of two bumps (gaussians followed and preceded by powers of half-sinusoids for the two classes respectively). The variation is obtained by changing the pulse width of the gaussians and powers of sinusoids. 100 shapes each of class A and B were generated and their mean shape was computed on the shape space. 10-fold repeated cross validation was performed to test a quadratic classifier. The average cross validation error was obtained as 0.031 (accuracy of 96.9%).

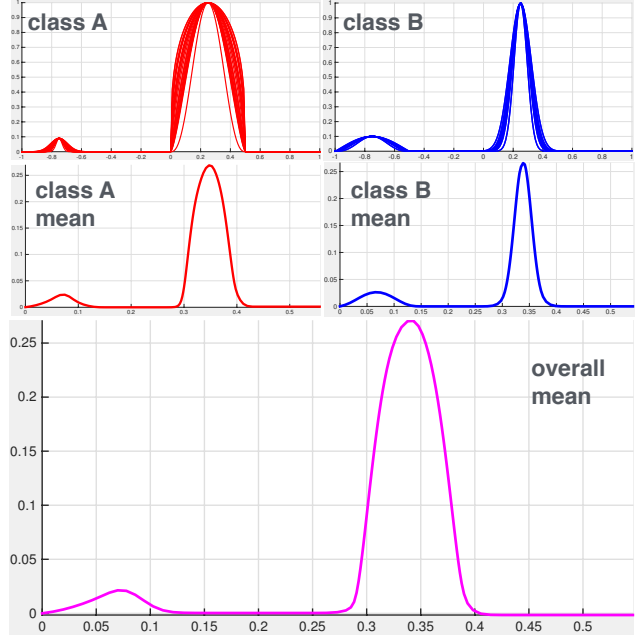


Figure 1. Top row: 100 samples each from shapes of class A and B. Middle row: Mean shapes for the two classes. Bottom row: mean shape of the population.

6.2. Classification of Cortical Sulci

Here we show results of classification of cortical sulci in a population of 175 human subjects from ages 18 to 50 years. Subjects's brains were scanned on a 1.5T MRI Siemens scanner with a 1mm³ isotropic voxel resolution. The MRI images underwent skull stripping and tissue editing, and the gray matter boundary was extracted [7] to obtain triangular mesh representations of the cortex. A set of 26 landmark sulci were traced on the cortical surface by an expert neuroanatomist. Since this paper is focused on discriminant analysis, for the purpose of demonstration, we chose to show results of binary classification for the central and the precentral sulci as shown in Figure 2. Sulcal landmarks were represented as open curves and approximated by the tangent space Fourier representation at the mean shape for both the classes.

Figure 2 shows an example of the precentral and the central sulci traced on an individual subject, the population of the precentral and the central sulci and their respective means, as well as the overall mean of the full dataset. We used a repeated 10-fold cross validation to evaluate the a linear discriminant classifier to obtain an average cross validation error of 12.75% (accuracy of 87.25 %) with a model training error of 5.72%.

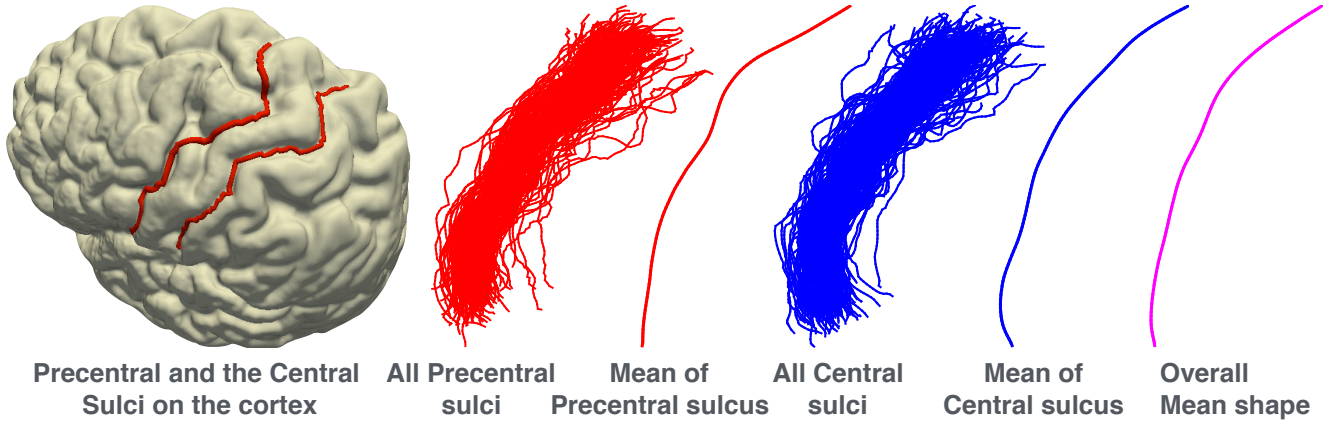


Figure 2. Left to right: examples of the precentral and the central sulci on the cortex, population of all 175 precentral sulci, mean shape of the precentral sulci, population of all 175 central sulci, mean shape of the central sulci, and the mean shape of the overall population.

6.3. Classification of Schizophrenia Diagnostic Status using Corpus Callosus Shapes

Emerging evidence [17] suggests the important role of white matter abnormalities in schizophrenia. Consequently one expects disease effects on callosal morphology in patients. Here we show results of classification of schizophrenia diagnosis in a population of 40 subjects (20 healthy controls and 20 patients with schizophrenia) from ages 18 to 46 years. MRI images for all subjects were acquired on a 1.5T Siemens scanner with 1mm^3 isotropic voxel resolution. An expert neuroanatomist traced the callosal boundary on the mid sagittal slice for each subject. The boundaries were represented in the shape space and the tangent space Fourier approximation was obtained for each callosal shape. Figure 3 shows the callosal shapes for controls and patients. After a repeated 10-fold cross validation to evaluate the linear classifier, we obtained an average cross validation error of 14.67% (accuracy of 85.33 %) with a model training error of 5%.

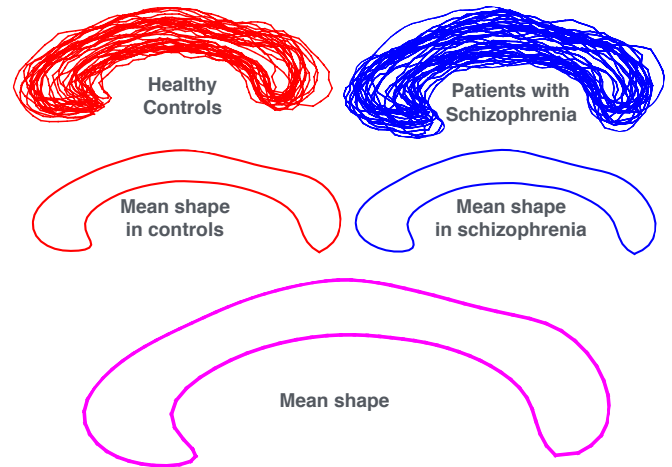


Figure 3. Top row: 20 corpus callosus shapes each for both healthy controls and patients with schizophrenia. Middle row: Mean shapes for the healthy controls and schizophrenia. Bottom row: mean shape of the population.

6.4. Classification of Severe Alcohol Exposure using Facial Midline Curves

Fetal alcohol spectrum disorders arise from moderate to heavy prenatal alcohol exposure to the fetus during development of the central nervous system (CNS). Fetal alcohol syndrome (FAS), which is a more severe case of FASD is usually characterized identifiable facial dysmorphology [20], growth deficits, and behavioral problems. Children and adolescents diagnosed with FAS exhibit a variety of minor to severe facial abnormalities including flat nasal bridge, smooth philtrum (space between nasal base and upper lip), short palpebral fissure (space between eyelids), thin upper lip, and a flat mid-face [8]. Here we attempt to classify patients with FAS from healthy controls using the midline curve from 3D facial surfaces. Our data consisted of 45 subjects (21 with FAS and 24 healthy controls) from ages

5 to 16 years. 3D facial range scans were acquired using a Minolta Vivid 910 laser scanner (1/3 frame CCD 340K pixels, triangulation light block method, scan range- 0.6 - 2.5 m, 0.3 sec. peak acquisition time). The facial midline curve was automatically extracted as an intersection of a coronal center plane passing through the nose and the facial surface. Shapes of all curves were represented by their SRVFs and a tangent space was constructed at the mean shape as shown in section 4. Figure 4 shows the facial midline shapes for healthy controls and patients with FAS, and the mean shape of the population. A linear discriminant classifier was trained on this data and a repeated 10-fold cross validation was performed. This data yielded an average cross validation error of 0.2378 (accuracy of 76.22%) with a training error of 8%.

Finally, all the results are summarized in a tabular form

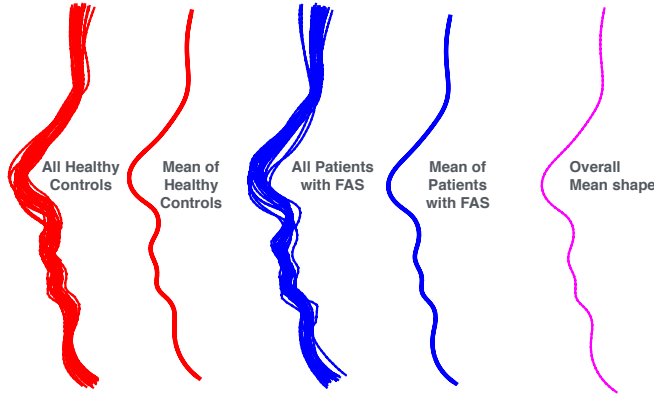


Figure 4. Left to right: midline facial profile curves for 24 healthy controls, mean shape of healthy controls, 21 facial profile curves for patients with FAS, mean shape of patients, and the mean shape of the overall population.

Table 1. Classifier Performance for different types of data.

Example	Classifier	N (size)	Training Accuracy	Testing Accuracy
toy	qda	200	100%	96.9%
sulci	lda	175	94.28%	87.25%
corpus callosum	lda	40	95%	85.33%
facial curves	lda	45	92%	76.22%

in Table 6.4.

7. Discussion

We presented a geometric approach for defining classifiers on the shape space of curves. In this work, we focused on quadratic and linear discriminant analysis, although the approach is general and allows for defining more advanced classification techniques on the shape space. While we obtained high accuracy on synthetic data as well as corpus callosum shapes, we obtained a moderate accuracy on FAS facial dysmorphology data. This could be explained by the fact that the FAS data consisted of just the midline facial curves instead of the whole face. We expect the accuracy to improve after combining the shape classifier with anthropomorphic measures such as the palpebral fissure spans or the philtrum lengths.

Future work will focus on implementing more advanced machine learning approaches for shapes using the geometric framework proposed in this paper.

8. Acknowledgment

This research was supported by National Institute on Alcohol Abuse and Alcoholism award K25 AA024192.

References

- [1] M. Ambühl, C. Brepsant, J.-J. Meister, A. Verkhovsky, and I. Sbalzarini. High-resolution cell outline segmentation and tracking from phase-contrast microscopy images. *Journal of microscopy*, 245(2):161–170, 2012. 1
- [2] X. An, Z. Liu, Y. Shi, N. Li, Y. Wang, and S. Joshi. Modeling dynamic cellular morphology in images. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 340–347, 2012. 1
- [3] H. Bogunović, J. M. Pozo, R. Cárdenes, M. C. Villa-Uriol, R. Blanc, M. Pötin, and A. F. Frangi. Automated landmarking and geometric characterization of the carotid siphon. *Medical Image Analysis*, 16(4):889–903, 2012. 2
- [4] F. L. Bookstein, P. D. Sampson, P. D. Connor, and A. P. Streissguth. Midline corpus callosum is a neuroanatomical focus of fetal alcohol damage. *Anatomical Record–Advances In Integrative Anatomy And Evolutionary Biology*, 269(3):162–174, 2002. 1
- [5] J. Downhill, M. Buchsbaum, T. Wei, J. Spiegel Cohen, E. Hazlett, M. Haznedar, J. Silverman, and L. Siever. Shape and size of the corpus callosum in schizophrenia and schizotypal personality disorder. *Schizophrenia Research*, 42(3):193–208, 2000. 1
- [6] J. Glaunès, A. Qiu, M. I. Miller, and L. Younes. Large deformation diffeomorphic metric curve mapping. *International Journal of Computer Vision*, 80(3):317–336, 2008.
- [7] C. Holmes, D. MacDonald, J. Sled, A. Toga, and A. Evans. Cortical peeling: CSF/grey/white matter boundaries visualized by nesting isosurfaces. *Visualization in Biomedical Computing*, 1131:99–104, 1996. 6
- [8] K. Jones and D. Smith. Recognition of the fetal alcohol syndrome in early infancy. *The Lancet*, 302(7836):999–1001, 1973. 7
- [9] S. Joshi, R. Cabeen, B. Sun, A. Joshi, B. Gutman, A. Zamyanyan, S. Chakrapani, I. Dinov, R. Woods, and A. Toga. Cortical sulcal atlas construction using a diffeomorphic mapping approach. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 357–366, 2010. 1
- [10] S. H. Joshi, R. P. Cabeen, A. A. Joshi, B. Sun, I. Dinov, K. L. Narr, A. W. Toga, and R. P. Woods. Diffeomorphic sulcal shape analysis on the cortex. *IEEE transactions on medical imaging*, 31(6):1195–1212, 2012. 1
- [11] S. H. Joshi, E. Klassen, A. Srivastava, and I. Jermyn. A novel representation for Riemannian analysis of elastic curves in R^n . *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2007. 2, 3, 4
- [12] S. H. Joshi, E. Klassen, A. Srivastava, and I. Jermyn. Removing shape-preserving transformations in square-root elastic (SRE) framework for shape analysis of curves. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 387–398. Springer, 2007. 2
- [13] S. H. Joshi, K. L. Narr, O. R. Philips, K. H. Nuechterlein, R. F. Asarnow, A. W. Toga, and R. P. Woods. Statistical shape analysis of the corpus callosum in schizophrenia. *Neuroimage*, 64:547–559, 2013. 1

- [14] S. H. Joshi and A. Srivastava. Intrinsic bayesian active contours for extraction of object boundaries in images. *International journal of computer vision*, 81(3):331–355, 2009. [1](#)
- [15] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977. [4](#)
- [16] A. Kufner, O. John, and S. Fucik. *Function spaces*, volume 3. Springer Science & Business Media, 1977. [3](#)
- [17] K. Roy, J. C. Murtie, B. F. El-Khodori, N. Edgar, S. P. Sardi, B. M. Hooks, M. Benoit-Marand, C. Chen, H. Moore, P. O’Donnell, et al. Loss of erbb signaling in oligodendrocytes alters myelin and dopaminergic function, a potential mechanism for neuropsychiatric disorders. *Proceedings of the National Academy of Sciences*, 104(19):8131–8136, 2007. [7](#)
- [18] E. Sowell, S. Mattson, P. Thompson, T. Jernigan, E. Riley, and A. Toga. Mapping callosal morphology and cognitive correlates - Effects of heavy prenatal alcohol exposure. *Neurology*, 2001. [1](#)
- [19] A. Srivastava, E. Klassen, S. H. Joshi, and I. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1451–1428, 2011. [2](#), [3](#), [4](#)
- [20] M. Suttie, T. Foroud, L. Wetherill, J. L. Jacobson, C. D. Molteno, E. M. Meintjes, H. E. Hoyme, N. Khaole, L. K. Robinson, E. P. Riley, S. W. Jacobson, and P. Hammond. Facial dysmorphism across the fetal alcohol spectrum. *Pediatrics*, 131(3):e779–e788, 2013. [7](#)
- [21] J. J. Wolff, G. Gerig, J. D. Lewis, T. Soda, M. A. Styner, C. Vachet, K. N. Botteron, J. T. Ellison, S. R. Dager, A. M. Estes, et al. Altered corpus callosum morphology associated with autism over the first 2 years of life. *Brain*, 138(7):2046–2058, 2015. [1](#)