

# LANDMARK REPRESENTATION OF SHAPES AND FISHER-RAO GEOMETRY

Washington Mio

Department of Mathematics  
Florida State University  
Tallahassee, FL 32306-4510 USA

Xiuwen Liu

Department of Computer Science  
Florida State University  
Tallahassee, FL 32306-4530 USA

## ABSTRACT

We develop computational strategies to calculate geodesics and geodesic distances between plane shapes represented by mixture of Gaussians centered at landmark points with a fixed variance with respect to the information-theoretic Fisher-Rao metric. This representation and metric have been investigated recently by Peter and Rangarajan, but a feasible computational approach was not provided. The algorithms developed are applied to shape clustering and the results are compared to those obtained with other methods.

**Index Terms**— Shapes, shapes of curves, Fisher information, shape geodesic, Fisher-Rao geometry

## 1. INTRODUCTION

The shape of contours of objects is one of the fundamental features that determine the information content of an image. The main goal of this paper is to develop effective computational strategies for the calculation of geodesics in the space of plane shapes represented by a mixture of isotropic Gaussians centered at a collection of ordered landmark points with respect to the Fisher-Rao metric. This information-theoretic approach to the quantification of divergence and similarity of plane shapes has been investigated recently by Peter and Rangarajan [1]; however, the computational feasibility of the model was not established in that paper.

The study of shapes dates back to D'Arcy Thompson [2], but the first more systematic algorithmic treatment of shape representations and metrics is due Bookstein [3] and Kendall [4]. In their work, a shape is represented by a collection of ordered landmark points and a geodesic metric is used to quantify shape divergence. An important feature of such geometric models is that shapes are naturally morphable as geodesics yield natural shape interpolators. In practice, the selection and observation of landmarks are contextual and noisy processes, so it is desirable to incorporate elements of uncertainty into representations and metrics. In [1], this was done by representing the plane shape associated with a collection of landmarks with a mixture of uniform Gaussians of fixed variance

This work was supported in part by NSF grants CCF-0514743 and IIS-0307998, and ARO grant W911NF-04-01-0268.

$\sigma^2$  centered at the landmark points. A question arises: *What is the natural geometry to adopt in such probabilistic representation of shapes?* From a statistical standpoint, monotonicity properties and the invariance of the Fisher-Rao metric under sufficient statistics [5] make this Riemannian structure a natural choice.

Our approach to the calculation of geodesics in the parametric space of uniform mixtures of isotropic Gaussians with fixed variance is numerical. The gradient descent approach of [1] based on the differential equation that governs geodesics is computationally very costly, so we develop techniques to estimate the gradient in a more efficient manner. To demonstrate the usefulness and computational feasibility of the methods introduced in this paper, we apply the algorithm for calculating geodesics and geodesic distances to clustering of plane shapes.

The paper is organized as follows. In Sec. 2, we introduce the probabilistic representation of shapes and the Fisher-Rao metric. Sec. 3 is devoted to numerical aspects of the calculation of geodesics. Examples of geodesics are presented in Sec. 4 and applications to clustering are discussed in Sec. 5.

## 2. SHAPE REPRESENTATION

We study plane shapes represented by  $k$  ordered landmark points  $p_1 = (\theta_1, \theta_2), \dots, p_k = (\theta_{2k-1}, \theta_{2k}) \in \mathbb{R}^2$ , which will be treated as a single vector  $\theta = (\theta_1, \dots, \theta_{2k}) \in \mathbb{R}^{2k}$ . To each  $\theta \in \mathbb{R}^{2k}$ , associate the 2D mixture of  $k$  isotropic Gaussians

$$(1) \quad p(x; \theta) = \frac{1}{2k\pi\sigma^2} \sum_{i=1}^k e^{-\frac{\|x-p_i\|^2}{2\sigma^2}},$$

with a fixed variance  $\sigma^2$ . The Fisher information matrix  $g(\theta)$  is the  $2k \times 2k$  symmetric matrix whose  $(i, j)$ -entry is

$$(2) \quad \begin{aligned} g_{ij}(\theta) &= \iint_{\mathbb{R}^2} p(x; \theta) \frac{\partial}{\partial \theta_i} \log p(x; \theta) \frac{\partial}{\partial \theta_j} \log p(x; \theta) dx \\ &= \iint_{\mathbb{R}^2} \frac{1}{p(x; \theta)} \frac{\partial p}{\partial \theta_i}(x; \theta) \frac{\partial p}{\partial \theta_j}(x; \theta) dx. \end{aligned}$$

This last integrand is computed explicitly in Appendix A.

For each  $\theta$ , the matrix  $g(\theta)$  defines an inner product  $\langle \cdot, \cdot \rangle_\theta$  on the parameter space  $\mathbb{R}^{2k}$  by

$$(3) \quad \langle v, w \rangle_\theta = v^T g(\theta) w.$$

The Riemannian manifold  $(\mathbb{R}^{2k}, \langle \cdot, \cdot \rangle_\theta)$  is known as the Fisher-Rao information manifold associated with the parametric mixture family. If  $t \mapsto \theta(t)$ ,  $0 \leq t \leq 1$ , is a smooth path in  $\mathbb{R}^{2k}$ , its *energy* is defined as

$$(4) \quad E = \frac{1}{2} \int_0^1 \langle \dot{\theta}(t), \dot{\theta}(t) \rangle_{\theta(t)} dt.$$

Geodesics in the Fisher information manifold are critical points of the energy functional  $E$ . Thus, for numerical calculations of geodesics using gradient descent, we need to develop a stable algorithm to estimate the gradient of  $E$ .

### 3. THE GRADIENT OF $E$

We discretize a path  $\theta(t)$  as a sequence  $\theta(1), \dots, \theta(n)$  of  $n$  points in  $\mathbb{R}^{2k}$ . In this representation, the energy  $E$  can be viewed as a function on  $\mathbb{R}^{2kn}$  and expressed as

$$(5) \quad E = \sum_{i=1}^n \dot{\theta}(i)^T g(\theta(i)) \dot{\theta}(i).$$

Our experiments indicate that the estimation of the gradient of  $E$  based on the partial derivatives of  $E$  in the canonical basis of  $\mathbb{R}^{2kn}$  is numerically very unstable. Thus, at each stage  $\theta(i)$  of the path, we first identify orthogonal directions that are likely to contribute to the energy most significantly and use the partial derivatives along these directions to approximate the gradient. The experiments in Sec. 4 demonstrate that gradient descent using this approach leads to a stable calculation of geodesics.

For  $\theta \in \mathbb{R}^{2k}$ , diagonalize the symmetric matrix  $g(\theta)$  to obtain an orthonormal basis  $\{v_1(\theta), \dots, v_{2k}(\theta)\}$  of  $\mathbb{R}^{2k}$ , with respect to the standard Euclidean metric, formed by eigenvectors of  $g(\theta)$ . We assume that the basis elements are ordered so that the associated eigenvalues  $\lambda_1, \dots, \lambda_{2k} > 0$  form a decreasing sequence. Note that, by (3), this basis is also an orthogonal basis of  $\mathbb{R}^{2k}$  with respect to the inner product  $\langle \cdot, \cdot \rangle_\theta$ .

At a point  $\theta(i)$  of a path, denote the eigenvectors by  $v_{ij}$ ,  $1 \leq j \leq 2k$  and the corresponding eigenvalues by  $\lambda_{ij}$ . The magnitude of an eigenvalue  $\lambda_{ij}$  determines the potential contribution of the associated eigendirection  $v_{ij}$  to the energy of the path. We truncate this basis to an orthonormal set  $\{v_{i1}, \dots, v_{ip}\}$ ,  $p \leq 2k$ , to capture a preset percentage of the energy

$$(6) \quad E_i = \dot{\theta}(i)^T g(\theta(i)) \dot{\theta}(i)$$

of the path at that point. The number  $p$  may vary as we traverse the path, but we assume that it is constant just to simplify notation. The truncation of the basis not only leads to

computational efficiency, but also yields a stable estimation of the gradient, which is done as follows. Since the end points of the path are to remain fixed, the gradient vanishes at these points. For  $1 \leq i \leq n$  and  $1 \leq j \leq p$ , let  $V_{ij} \in \mathbb{R}^{2kn}$  denote the unit vector  $(0, \dots, 0, v_{ij}, 0, \dots, 0) \in \mathbb{R}^{2kn}$ . Then,

$$(7) \quad \partial_{ij} E = \frac{E(\theta + \delta V_{ij}) - E(\theta)}{\delta}.$$

The gradient is estimated as

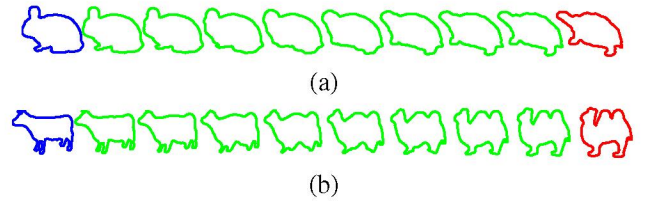
$$(8) \quad \nabla E(\theta) \approx \sum_{1 \leq i \leq n, 1 \leq j \leq p} (\partial_{ij} E) V_{ij}.$$

*Remark.* Note that in the estimation of partial derivatives, the energy of each perturbed path has only two terms that differ from those of the original path. This observation has significant implications in computations.

### 4. GEODESICS

For shapes represented by  $\theta_a, \theta_b \in \mathbb{R}^{2k}$ , a geodesic between  $\theta_a$  and  $\theta_b$  with respect to the Fisher-Rao metric is calculated using gradient descent to minimize the functional  $E$  over paths in  $\mathbb{R}^{2k}$  connecting  $\theta_a$  to  $\theta_b$ . The gradient search is initialized with the linear path  $\theta: [0, 1] \rightarrow \mathbb{R}^{2k}$  parameterized with constant speed.

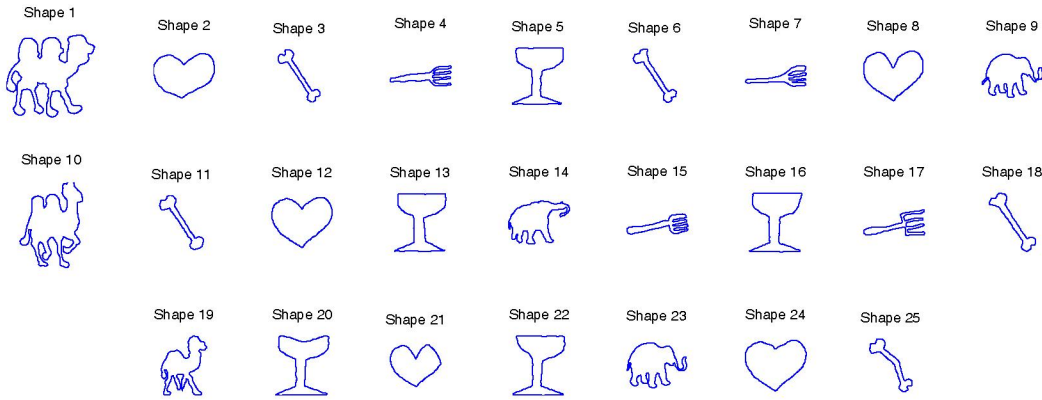
Fig. 1 shows two examples of Fisher-Rao shape geodesics calculated with 99 landmark points. Plots of the energy  $E$  versus the number of iterations during the gradient search are shown in Fig. 2. The choice and correspondence of landmarks were done with an alignment algorithm that utilizes dynamic programming, as discussed in [6]. This is a variant that uses velocity fields of the alignment procedure studied in [7]; similar shape alignment techniques have been studied by other authors. Additional examples of geodesics are shown in Fig. 3.



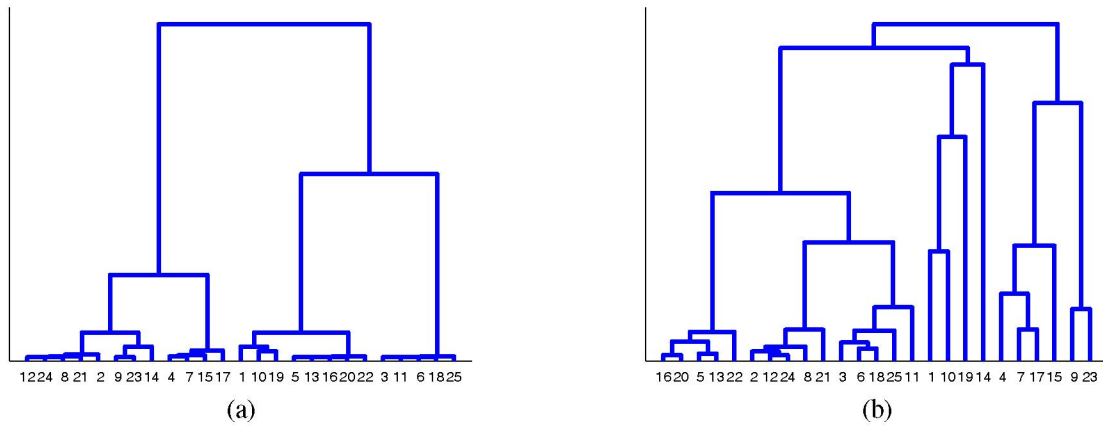
**Fig. 1.** Examples of geodesics in the Fisher-Rao metric computed with 99 landmark points.

### 5. SHAPE CLUSTERING

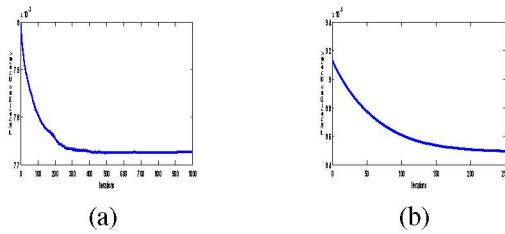
To illustrate the usefulness of the methods in shape analysis, we apply the algorithm for computing Fisher-Rao shape geodesics to shape clustering. Fig. 4 shows 25 shapes from



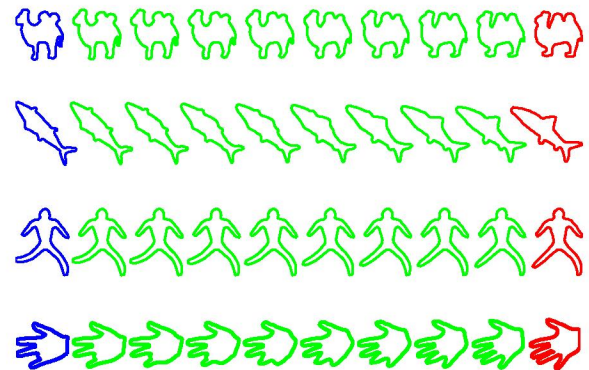
**Fig. 4.** 25 shapes from the LEMS database.



**Fig. 5.** Dendrograms associated with (a) the Fisher-Rao metric and (b) the geodesic metric of [8].



**Fig. 2.** Energy versus the number of iterations during the gradient search for the geodesics shown in Figs. 1(a) and 1(b).



**Fig. 3.** Shape geodesics in the Fisher-Rao metric.

the LEMS database that, intuitively, can be naturally grouped into 6 clusters. To cluster the shapes, we first compute all pairwise geodesic distances and then use a hierarchical clustering procedure. We start with 25 clusters, each consisting of a single shape, and merge them sequentially using minimal cluster distance as merging criterion until 6 clusters remain. For comparison purposes, we carry out the same experiment with the geodesic metric of [8]. The dendrogram associated with these geodesic metrics are shown in Figs. 5(a) and 5(b),

respectively. The six clusters obtained with the Fisher-Rao metric are shown in Fig. 6; all six clusters are intuitively correct and are compact and well separated as seen from the dendrogram. Hierarchical clustering with the metric developed in [8] does not lead to good results. For example, shape 14 forms



a one-element cluster and it does not get grouped with shapes 9 and 23. Moreover, the clusters obtained are not well separated, as indicated in the dendrogram. To circumvent some of these problems, a variant of the *K-Means Algorithm* was investigated in [9] for the geodesic metric of [8]. However, the computational costs associated with the K-Means Algorithm of [9] is high, so that a metric such as Fisher-Rao that produces good results with a hierarchical procedure is very desirable.

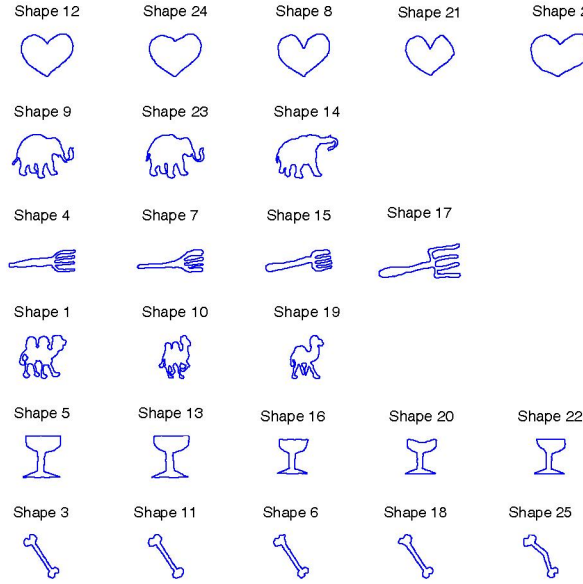


Fig. 6. Clusters obtained with the Fisher-Rao metric.

## 6. SUMMARY AND COMMENTS

We developed a novel geometric algorithm to calculate shape geodesics and geodesic distances using a probabilistic representation of shapes and the information theoretic Fisher-Rao metric. Several examples of shape geodesics were provided and a hierarchical shape clustering experiment was used to illustrate the usefulness of the methodology introduced. Larger clustering experiments have been carried out, but results are not reported due to limitation of space.

Since the emphasis of the paper is on computational feasibility, we did not present the details of the problem of choosing landmarks and establishing correspondences in an automated manner. This was only briefly discussed in Sec. 4 and will be further explained elsewhere.

## 7. REFERENCES

[1] A. Peter and A. Rangarajan, “Shape analysis using the Fisher-Rao Riemannian metric: unifying shape representation and deformation,” Technical report, University of Florida, 2005.

[2] D. W. Thompson, *On Growth and Form: The Complete Revised Edition*, Dover, 1992.

[3] F. L. Bookstein, “Size and shape spaces for landmark data in two dimensions,” *Statistical Science*, vol. 1, pp. 181–242, 1986.

[4] D. G. Kendall, “Shape manifolds, Procrustean metrics and complex projective spaces,” *Bulletin of London Mathematical Society*, vol. 16, pp. 81–121, 1984.

[5] S. Amari and H. Nagaoka, *Methods of Information Geometry*, AMS and Oxford University Press, New York, 2000.

[6] W. Mio, X. Liu, and M. K. Hurdal, “On alignment and shape of curves,” Technical report, Florida State University, 2006.

[7] I. Cohen, N. Ayache, and P. Sulger, “Tracking points on deformable objects using curvature information,” in *Lecture Notes in Computer Science*, 1992, vol. 588.

[8] E. Klassen, A. Srivastava, W. Mio, and S. Joshi, “Analysis of planar shapes using geodesic paths on shape manifolds,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 372–383, 2004.

[9] A. Srivastava, S. Joshi, W. Mio, and X. Liu, “Statistical shape analysis: Clustering, learning and testing,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 590–602, 2005.

## A. CALCULATION OF PARTIAL DERIVATIVES

Write the coordinates of a point  $x \in \mathbb{R}^2$  as  $x = (x_1, x_2)$ . For  $1 \leq i \leq k$ ,

$$\frac{\partial p}{\partial \theta_{2i-1}}(x; \theta) = \frac{x_1 - \theta_{2i-1}}{2k\pi\sigma^4} e^{-\frac{\|x - p_i\|^2}{2\sigma^2}}$$

and

$$\frac{\partial p}{\partial \theta_{2i}}(x; \theta) = \frac{x_2 - \theta_{2i}}{2k\pi\sigma^4} e^{-\frac{\|x - p_i\|^2}{2\sigma^2}}.$$

Thus,

$$\begin{aligned} \frac{\partial p}{\partial \theta_{2i-1}}(x; \theta) \frac{\partial p}{\partial \theta_{2j-1}}(x; \theta) &= \frac{(x_1 - \theta_{2i-1})(x_1 - \theta_{2j-1})}{4k^2\pi^2\sigma^8} e^{-\frac{\|x - p_i\|^2 + \|x - p_j\|^2}{2\sigma^2}}; \\ \frac{\partial p}{\partial \theta_{2i}}(x; \theta) \frac{\partial p}{\partial \theta_{2j}}(x; \theta) &= \frac{(x_2 - \theta_{2i})(x_2 - \theta_{2j})}{4k^2\pi^2\sigma^8} e^{-\frac{\|x - p_i\|^2 + \|x - p_j\|^2}{2\sigma^2}}; \\ \frac{\partial p}{\partial \theta_{2i-1}}(x; \theta) \frac{\partial p}{\partial \theta_{2j}}(x; \theta) &= \frac{(x_1 - \theta_{2i-1})(x_2 - \theta_{2j})}{4k^2\pi^2\sigma^8} e^{-\frac{\|x - p_i\|^2 + \|x - p_j\|^2}{2\sigma^2}}. \end{aligned}$$