Riemannian Analysis of Probability Density Functions with Applications in Vision

Anuj Srivastava
Department of Statistics
Florida State University
anuj@stat.fsu.edu

Ian Jermyn
ARIANA Group, INRIA
Sophia Antipolis, France
ian.jermyn@sophia.inria.fr

Shantanu Joshi
Department of Electrical Engineering
Florida State University
joshi@eng.fsu.edu

Abstract

Applications in computer vision involve statistically analyzing an important class of constrained, non-negative functions, including probability density functions (in texture analysis), dynamic time-warping functions (in activity analysis), and re-parametrization or non-rigid registration functions (in shape analysis of curves). For this one needs to impose a Riemannian structure on the spaces formed by these functions. We propose a "spherical" version of the Fisher-Rao metric that provides closed-form expressions for geodesics and distances, and allows fast computation of sample statistics. To demonstrate this approach, we present an application in planar shape classification.

1. Introduction

Several applications in computer vision, such as texture analysis, activity analysis and shape analysis, use mathematical representations involving a certain class of constrained, non-negative function. To study observed variability within and across classes, one has to develop statistical models on appropriately constrained function spaces. Additionally, one needs to develop metrics, estimators and evaluation tools for the desired inferences. The main difficulty here comes in performing calculus while respecting the nonlinear constraints imposed on these functions. A natural solution is to work on the nonlinear manifolds formed by the allowable functions and to utilize the underlying differential geometries of these manifolds to perform statistics. This general framework has previously been used by several researchers. However, questions remain about the choice of: (i) the representation and (ii) the Riemannian metric. In this paper we focus on a specific set of closely-related constrained functions – probability density functions (pdfs), warping functions, non-rigid registration functions, 1D diffeomorphisms, and re-parametrization functions(some of them are actually identical) and study the different choices of representations. Our goal is to choose a representation and a metric that allow the efficient computation of statistical tools for applications in computer vision.

Before we present different representations and metrics, we specify the functions of interest and their motivating applications. Perhaps the most important example is the use of pdfs in modeling frequencies of pixel values in images. Very commonly images are filtered using pertinent filters and the resulting images are used to estimate probability densities of the filtered pixel values. Applications of this tool include image retrieval [14], and texture synthesis [16, 12]. The second relevant problem is in *activity analysis* where one studies times of occurrences of ordered events in order to classify this activity as a whole. For instance, consider the problem of a person arriving at the airport, checking in at an airline counter, and going to the departure gate to catch a flight. Since these individual events can be performed with random time delays, one has to introduce a time-warping function in order to register and match observations [15, 13]. The time-warping functions are naturally constrained to be non-decreasing and can be viewed as cumulative distribution functions (cdfs). With an additional positivity constraint, the set of functions form the group

of 1D diffeomorphisms, whose 2D counterpart has famously been used in the development of deformable template models for images [8]. The third problem is in analyzing the shapes of closed, planar curves that are available as ordered sets of points. In order to compare them in a manner that is invariant to their parameterizations, one forms a quotient space, called the shape space, that is defined as the space of closed curves modulo all possible re-parameterizations [7]. A related problem is to perform non-rigid registration of points across curves [10, 11].

Although the above three applications are quite different, the sets of constrained non-negative functions used are closely related. As described later, these functions may be represented and studied in several different but equivalent representations (pdfs, cdfs, log-densities, or square-root densities), each of which may be better suited to certain specific tasks. An important step in classifying observations using these functions is to compute distances between functions. This task is accomplished by imposing Riemannian structures on appropriate manifolds formed by these functions. The most natural Riemannian metric in this context is the so-called *Fisher-Rao* metric, which has been used extensively in computer vision [6, 5, 10, 11]. Čencov [2] showed that this is the only metric that is invariant to re-parametrizations of those functions. This metric has also played an important role in information geometry due to the pioneering efforts of Amari [1].

The remaining question is: What choice of representation of functions (and hence the Fisher-Rao metric) is most efficient for our applications? We will demonstrate that the square-root form, defined to be the square-root of a pdf, results in the desired manifold being the unit sphere inside a larger Hilbert space with the usual $\mathbb{L}^2$ metric. In view of the spherical nature of the underlying space, many of the desired quantities (geodesics, exponential maps, inverse of exponential maps) are available in analytic forms. This is in contrast to past usage of the Fisher-Rao metric where metrics and geodesics had to be approximated using numerical methods. In this paper, we will demonstrate the computational advantages of using the square-root form, and its associated Fisher-Rao metric, in vision applications.

The rest of this paper is organized as follows. Section 2 presents three of several applications in computer vision that motivate this work and introduces four different mathematical representations of the functions of interest. Section 3 summarizes the differential geometry of the chosen representation and Section 4 demonstrates the computation of sample means using geometric tools. The paper ends with a demonstration of the proposed representation in a problem of binary shape classification in Section 5 and a brief summary in Section 6.

2. Motivations & Representations

In this section we present some motivating applications for studying the constrained, non-negative functions that we focus on. Additionally, we present several choices for representing these functions and discuss the structures of the resulting Riemannian manifolds equipped with the Fisher-Rao metric expressed in these representations.

2.1. Motivating Problems

We start by presenting some applications that involve the functions of interest.

1. **Spectral PDFs of Images**: As the first example, we highlight the use of pdfs in spectral analysis of textured, natural, or man-made images. Shown in Figure 1 is an example: the top left panel shows an image I that is then filtered using Gabor filters [3] at different orientations. For each resulting image, one computes a pdf of the gray scale pixel values. The remaining three panels shows examples of such pdfs. In the spectral analysis, each image is represented by a collection of pdfs, generated for a pre-determined bank of Gabor, Laplacian and other derivative filters. Two images are compared by comparing their respective pdfs under the same filters. "Image templates", denoting the central tendency of images in a class, can be defined as "averages" of the corresponding pdfs. Rescaling the image pixels to take values in the range $[0, 1]$, one is interested in tools for computing distances and averages on the set of pdfs on $[0, 1]$.

2. **Time Warping in Activity Analysis**: Consider the problem of activity analysis and classification, where one is interested in studying a predetermined sequence of events that are performed with random time-separations. To compare different instances of the same activity, we need to use time-warping functions that allow registration of such instances. An observation of an activity is denoted by $T \equiv \{t_1, t_2, \dots\}$, where $t_i \in [0, 1]$ are the times of occurrences of these events. For a time-warping function (made precise later) γ, the set $\gamma(T) = \{\gamma(t_i)\}$ becomes another occurrence of that activity with events occurring at $\gamma(t_i)$. Since the sequence of events is maintained, one considers $\gamma(T)$ from the same class as T and would like

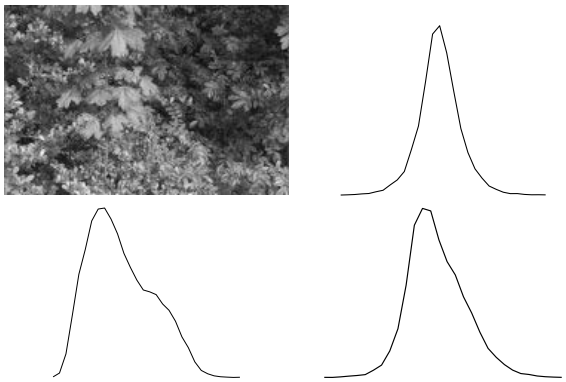

Figure 1. A natural image I (top left) and some pdfs of $I * F(\theta)$, where F is a Gabor filter with orientation θ, for different θs.

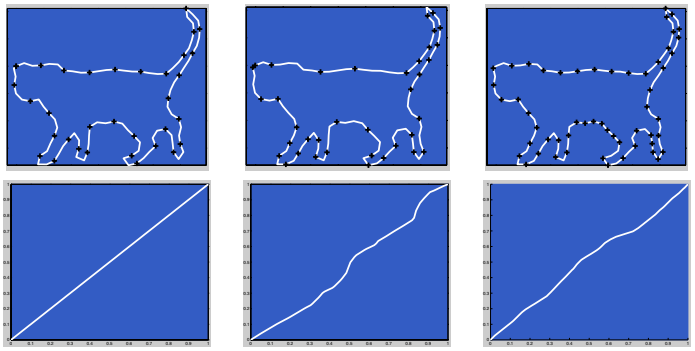

Figure 2. Same curve with three different parameterizations. The left one is sampled at uniform speed, or arc-length parametrization, while the other two have ϕs that are different from the identity.

to cluster and classify it appropriately. To quantify differences between any two activities from the same class, $\gamma_1(T)$ and $\gamma_2(T)$, or to develop empirical statistics from observed activities, one will need to define and compute distances and statistics on the space of allowable γs.

3. **Re-parameterizations in Shape Analysis**: Consider a simple, closed parameterized curve α in the plane. For studying its shape, we resize α to be of unit length. If α is sampled differently, i.e. the sampled points are spaced differently, then the change in its representation is given by $\alpha(\gamma(s))$, where γ is a re-sampling function. Shown in Figure 2 is an example of three different parameterizations of the same shape. The first one is the arc-length parametrization, i.e. $\gamma(s) = s$, while the other are non-uniform speed parameterizations. Given two arbitrarily sampled closed curves α_1 and α_2, we want to compare their shapes, taking different re-samplings into account. This problem is closely related to the non-rigid registration of points across shapes. In practical cases where shapes are observed in presence of noise, one needs to include probability models on space of γs to perform robust inferences.

2.2. Riemannian Representations

In the applications mentioned above, the functions of interest are, mathematically speaking, closely interrelated and can be represented in many ways. The ultimate choice of representation should be dependent on the ease of implementation in the ensuing application. A common issue in all these representations is that the underlying spaces are not vector spaces but are nonlinear (differentiable) manifolds; this creates a need to use tools from differential geometry – Riemannian metrics, geodesics, exponential maps, etc – on these manifolds for defining and computing statistics. As stated earlier, the choice of metric is fixed to be the Fisher-Rao metric because it is the only metric that is invariant to re-parameterization.

Let $\mathcal{D}$ be the set of all 1D diffeomorphisms from the interval $[0, 1]$ to itself. Next we enumerate different choices of representations and the associated forms of the Fisher-Rao metric:

1. **Probability density function** p: Each of the constrained, non-negative function of interest can be written as a pdf. The reason is that the derivative of a function transforms as a pdf under re-parametrization. To simplify discussion, we restrict to the space of pdfs on the interval $[0, 1]$, forming the set:

$$\mathcal{P} = \{p : [0,1] \to \mathbb{R} | \forall s, p(s) \geq 0 \text{ and } \int_0^1 p(s)ds = 1\} .$$

$\mathcal{P}$ is not a vector space. $\mathcal{D}$ provides a group action on $\mathcal{P}$ according to: $\mathcal{D} \times \mathcal{P} \to \mathcal{P}$: $(\gamma, p) = p(\gamma)\dot{\gamma}, \quad p \in \mathcal{P}, \quad \gamma \in \mathcal{D}$.

The Fisher-Rao metric on $\mathcal{P}$ can be stated as follows: for any point $p \in \mathcal{P}$ and the tangent vectors $v_1, v_2 \in T_p(\mathcal{P})$, the inner-product is given by

$$\langle v_1, v_2 \rangle = \int_0^1 v_1(s)v_2(s)\frac{1}{p(s)}ds \ . \qquad (1)$$

Here, $T_p(\mathcal{P})$ is the set of functions tangent to $\mathcal{P}$ at p. Amongst all possible representations, the pdf turns out to be one of the most difficult representations to work with. The main difficulty comes from the need for ensuring $p(s) \geq 0$ for all s. For example, in case one is trying to compute a geodesic

between any two elements of $\mathcal{P}$, it is quite difficult to ensure that p remains non-negative along the whole path. As an aside, we remark that the path $tp_1+(1-t)p_2$, for $0 \leq t \leq 1$ and $p_1, p_2 \in \mathcal{P}$, is not a geodesic between p_1 and p_2 under the Fisher-Rao metric.

2. **Cumulative distribution function** ϕ. Associated with each element of $\mathcal{P}$ is a unique cdf $\phi(s) = \int_0^s p(t)dt$. ϕ is a differentiable mapping from $[0,1]$ to itself. Additionally, if $p > 0$ then ϕ is also an invertible map. Define the set of all cdfs:

$$\Phi = \{\phi : [0,1] \to [0,1] | \forall s, \dot{\phi}(s) > 0, \phi(0) = 0, \phi(1) = 1\} .$$

Φ forms a group with the group operation given by composition, i.e. $\phi_1, \phi_2 \in \Phi$, the group operation is given by $\phi_2(\phi_1(s))$. The identity element of Φ is the function $id(s) = s$. $\mathcal{D}$ is identical to Φ and, thus, acts on Φ using the group operation: $(\gamma, \phi) \to \phi(\gamma)$. The Fisher-Rao metric for this representation is given by: for any $v_1, v_2 \in T_\phi(\Phi)$, we have

$$\langle v_1, v_2 \rangle = \int_0^1 \dot{v}_1(s)\dot{v}_2(s)\frac{1}{\dot{\phi}(s)}ds . \qquad (2)$$

Φ is somewhat easier than $\mathcal{P}$ to analyze in view of its group structure. Also, note that the time-warping functions in activity analysis and the re-parametrization functions (or non-rigid registration functions) in shape analysis can be directly written as elements of Φ or $\mathcal{D}$.

3. **Log density function**: Several past papers have used the logarithm of p to represent and analyze probability densities. Of course, this representation requires the function p to be strictly positive. The corresponding representation space is:

$$\mathcal{L} = \{\nu : [0,1] \to \mathbb{R} | \int_0^1 \exp(\nu(s))ds = 1\} .$$

The group $\mathcal{D}$ acts on $\mathcal{L}$ according to: $(\gamma, \nu) \to \nu(\gamma) + \log(\dot{\gamma})$. The Fisher-Rao metric in this representation is given by: for $v_1, v_2 \in T_\nu(\mathcal{L})$:

$$\langle v_1, v_2 \rangle = \int_0^1 v_1(s)v_2(s)\exp(\nu(s))ds . \qquad (3)$$

Although this representation has shown some success in texture analysis [9], there are a few major limitations here. Firstly, the pdf p should be strictly positive to have a logarithmic representation. Secondly, the Riemannian structure on the space is complicated and one has to use numerical techniques to compute geodesics on this space. For example, Mio et al. [9] use a shooting method to find geodesics between any two log-density functions. As demonstrated through an example later, this approach often results in paths that may not reach the target function and thus leads to large errors.

4. **Square-root density function**: The final choice of representation is the square root function: $\psi = \sqrt{p}$. Due to the nature of the square root, this function is not a unique representation of p; uniqueness can be imposed by assuming ψ to be non-negative. Note that there is no requirement for $p > 0$ for this representation to work. Here one considers the space:

$$\Psi = \{\psi : [0,1] \to \mathbb{R} | \psi \geq 0, \int_0^1 \psi^2(s)ds = 1\} .$$

The group action of $\mathcal{D}$ on Ψ is given by the mapping: $(\gamma, \psi) = \psi(\gamma)\sqrt{\dot{\gamma}}$. For any two tangent vectors $v_1, v_2 \in T_\psi(\Psi)$, the Fisher-Rao metric is given by:

$$\langle v_1, v_2 \rangle = \int_0^1 v_1(s)v_2(s)ds . \qquad (4)$$

The space of sqaure-root densities can be viewed as the non-negative orthant of the unit sphere in a Hilbert space. In this larger space, not only are the inner products of tangent vectors defined, but also the inner products of the elements of the Hilbert space itself (and hence of elements of Ψ). The distance in the larger space between two elements is simply the norm of their difference. Note that this is not the same as the distance in Ψ, which is the unit sphere.

Eqns. 1-4 are different ways of writing the same metric; they have simply been expressed in different coordinate systems. Additionally, they are invariant to re-parametrization.

Theorem 1 *The Fisher-Rao metric is invariant to the action of* $\mathcal{D}$.

Proof: We prove this using the square-root representation but the proof is similar for all other representations. Let $v_1, v_2 \in T_\psi(\Psi)$ for some $\psi \in \Psi$ and $\gamma \in \mathcal{D}$ be a re-parametrization function. The re-parametrization action takes ψ to $\psi(\gamma)\sqrt{\dot{\gamma}}$ and v_i to $\tilde{v}_i \equiv v_i(\gamma)\sqrt{\dot{\gamma}}$. The inner product after re-parametrization is given by:

$$\begin{aligned} \int_0^1 \tilde{v}_1(s)\tilde{v}_2(s)ds &= \int_0^1 v_1(\gamma(s))\sqrt{\dot{\gamma}(s)}v_2(\gamma(s))\sqrt{\dot{\gamma}(s)}ds \\ &= \int_0^1 v_1(t)v_2(t)dt, \quad t = \gamma(s) . \end{aligned}$$

which is the same as before the action by γ and, hence, invariant. $\square$

Which of these different representations should one choose for texture, activity, and shape analysis? In this paper, we propose the use of the square-root form for Riemannian analysis of constrained functions. The biggest advantage of this choice is that the resulting space Ψ is simply a unit sphere inside a larger Hilbert space with the $\mathbb{L}^2$ metric. The differential geometry of a sphere is well understood. There are closed form expressions for geodesics, exponential maps, inverse exponential maps and, consequently, sample statistics on a sphere. The condition that $\psi \geq 0$ is not too constraining; this amounts to restricting to a positive orthant of the unit sphere and does not impose any additional computational burden.

3. Differential Geometry of Ψ

In this section we specify the formulae for computing geodesics and other geometric quantities needed in vision applications. As stated above, the main advantage of selecting Ψ for analysis is that it is a convex subset of thea unit sphere in $\mathbb{L}^2$ and many of the geometric expressions are already known.

- **Geodesic Distance**: Given any two functions ψ_1 and ψ_2 in Ψ, the length of the geodesic connecting them in Ψ is given by:

$$d(\psi_1, \psi_2) = \cos^{-1}\langle \psi_1, \psi_2 \rangle \tag{5}$$

  where the inner product is as defined in Eqn. 4, but now applied to the elements of Ψ rather than the tangent vectors.

- **Geodesic**: The geodesic between two points is easily derived by noting that the radial projection to unit norm of the straight line joining the two points is the geodesic between the two points. The straight line is given by $\tilde{\psi}(t) = (1-t)\psi_0 + t\psi_1$. The projection to Ψ is then

$$\psi(t) = \frac{(1-t)\psi_1 + t\psi_2}{t^2 + (1-t)^2 + 2t(1-t)\langle \psi_1, \psi_2 \rangle} \ ,$$

  where it should be noted that t is not the arc-length parametrization.

  Conversion to the distance parametrization leads to the following expression for the geodesic:

$$\psi(t) = \frac{1}{\sin(s_{12})}\left[\sin(s_{12} - t)\psi_1 + \sin(t)\psi_2\right] \ , \tag{6}$$

  where $\cos(s_{12}) = \langle \psi_1, \psi_2 \rangle$ is the cosine of the geodesic distance between the two points.

- **Exponential Map**: The geodesic can be written in terms of a direction v in $T_{\psi_1}(\Psi)$:

$$G_t(v) = \cos(t)\psi_1 + \sin(t)\frac{v}{|v|} \ , \tag{7}$$

  where $v \in T_{\psi_1}(\Psi)$ and $\langle v, \psi_1 \rangle = 0$. As a result, the exponential map, $\varepsilon : T_{\psi_1}(\Psi) \rightarrow \Psi$, has a very simple expression:

$$\exp_{\psi_1}(v) = \cos(|v|)\psi_1 + \sin(|v|)\frac{v}{|v|} \ . \tag{8}$$

  The exponential map is a bijection if we restrict $|v|$ so that $|v| \in [0, \pi)$.

- **Inverse Exponential Map**: For any $\psi_1, \psi_2 \in \Psi$, we define $v \in T_{\psi_1}(\Psi)$ to be the inverse exponential of ψ_2 if $\exp_{\psi_1}(v) = \psi_2$; we will use the notation $\exp^{-1}_{\psi_1}(\psi_2) = v$. This is computed using the following steps:

$$\begin{aligned} u &= \psi_2 - \langle \psi_2, \psi_1 \rangle \psi_1 \\ v &= u \cos^{-1}(\langle \psi_1, \psi_2 \rangle) / \sqrt{\langle u, u \rangle} \ . \end{aligned} \tag{9}$$

Since we have simple analytical expressions for computing these quantities, the resulting statistical analysis of elements of Ψ is much simpler than in other representations. For instance, the geometry of $\mathcal{L}$ (using log-density coordinates) is too complicated to derive analytical expression for geodesics. Instead, one uses a numerical approach. Mio et al. [9] use a *shooting approach* for constructing geodesics on $\mathcal{L}$. The main idea is, given two log-densities ν_1 and ν_2 in $\mathcal{L}$, to find a tangent direction $v \in T_{\nu_1}(\mathcal{L})$ such that a geodesic along v (constructed numerically) reaches ν_2 in unit time. This optimal direction v is found by minimizing a *miss error*, defined as the Euclidean distance between the function reached (for the current v) and ν_2. There are several disadvantages associated with this numerical approach. Firstly, one may not always be able to solve this minimization problem globally using numerical techniques. Secondly, the resulting geodesic may get close to ν_2 but not quite reach it.

Figure 3 highlights the problem in using the shooting method in forming a geodesic under the log-density representation. The left picture shows a path that has been computed using the shooting method in the space $\mathcal{L}$ of the log-density functions, while the right path is computed using analytical expressions in the space Ψ of the square-root functions. (All the functions are displayed in terms of their pdfs for comparisons.) In each panel, the top and the bottom curves are the given densities p_1 and p_2, while the intermediate curves denote equally spaced points on the geodesics. In the left

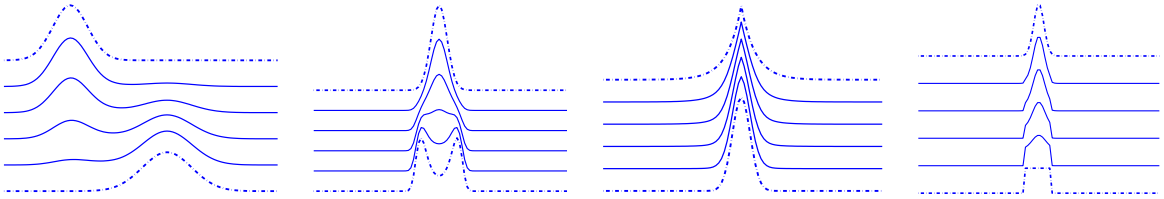

Figure 4. Geodesic paths in Ψ between some interesting square-root densities. All functions are displayed using their corresponding values in $\mathcal{P}$ for convenience.

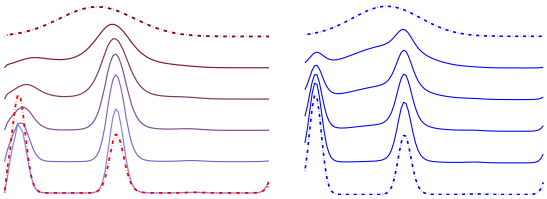

Figure 3. Limitations of shooting method. Computation of a geodesic using the shooting method may not reach the target pdf (left picture) while no such problem exists for the analytical form (right picture) available for the square-root density representation. Geodesics are displayed using the corresponding elements in $\mathcal{P}$ for convenience.

picture, the geodesic starting from $\nu_1 = \log(p_1)$ never quite reaches the point $\nu_2 = \log(p_2)$. In comparison, the geodesic computed using the analytical expression available for the square-root coordinates (right picture) has no such problem. Figure 4 shows some additional examples of geodesic paths between elements of Ψ (displayed using the corresponding elements in $\mathcal{P}$).

4. Sample Statistics on Ψ

An important ingredient in the statistical analysis of constrained functions is the computation of sample statistics such as means and covariances. For example, given a few observations of an activity, each resulting from a different time-warping function, one is interested in computing a *template* of that activity that involves taking an average of all the observed time-warping functions. Similarly, given a collection of pdfs from a set of images (under the same filter), we may use an "average" pdf to characterize the typical statistics of this set. To define and compute means, we use the notion of Karcher mean [4] as follows: For a number of observations $\psi_1, \psi_2, \ldots, \psi_n$, define their Karcher mean as: $\bar{\psi} = \text{argmin}_{\psi \in \Psi} \sum_{i=1}^{n} d(\psi, \psi_i)^2$, where d is taken to be the geodesic distance on Ψ. The search for $\bar{\psi}$ is performed using a gradient approach where an

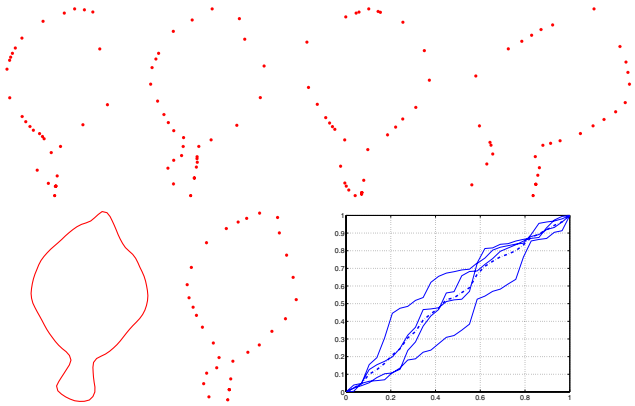

Figure 6. Tow row: Four observations of a shape (shown in bottom left) at randomly sampled points, i.e. with arbitrary speed functions ϕs shown in bottom right. Bottom row: the middle panel shows this shape sampled at the mean ϕ; the mean ϕ is shown on the right in broken line.

estimate is iteratively updated according to:

$$\mu \rightarrow \exp_\mu(\epsilon v), \quad v = \frac{1}{n} \sum_{i=1}^{n} \exp_\mu^{-1}(\psi_i)$$

where exp and $\exp^{-1}$ are given in Eqns. 8 and 9, respectively. The scalar $\epsilon > 0$ is a step size for iteration and is generally taken to be smaller than $\frac{1}{2}$.

Shown in Figure 5 are some examples of ψs and their Karcher means. As earlier, all the displays are in the form of pdfs $p = \psi^2$. In this particular example, the original pdfs are made of Gaussians with different means and variances, although that parametric form has not been utilized in computing the Karcher means.

Another example for computing averages, in context of shape analysis, is shown in Figure 6. The top row shows four randomly sampled versions of the shape shows in the bottom left. The bottom middle panel shows this shape sampled using a γ that is average of the four individual γ_is. These re-parametrization functions are shown in the bottom right panel.

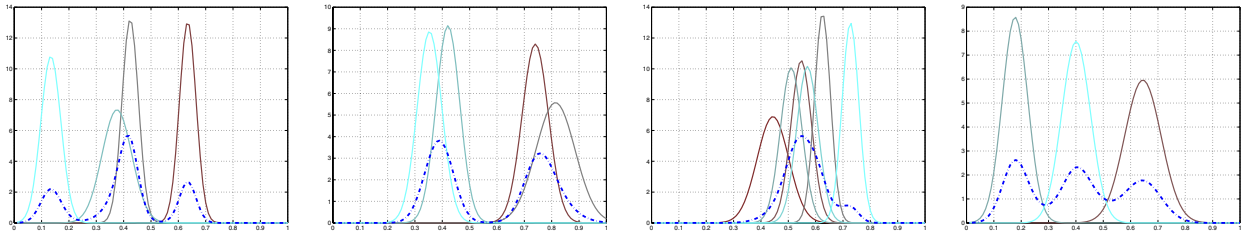

Figure 5. Examples of Karcher means of elements of Ψ. Each panel shows some Gaussian densities $p_i = \psi_i^2$ with different means and variances in solid lines. Superimposed on them in broken line is their Karcher mean $\bar{p} = \bar{\psi}^2$.

5. An Experiment on Shape Classification

To demonstrate the strength of this framework, we consider a simple problem in binary shape classification. We observe a shape from one of two classes, sampled randomly, and in the presence of additive noise, and the goal is to decide which class it came from.

The setup is as follows. Assume that a curve is to be sampled at n points, and let τ_0 denote the uniform sampling on $[0,1]$, i.e. $\tau_0(i) = i/n$ in arc-length parametrization. Then, the set $\mathcal{T} = \{\gamma(\tau_0) | \gamma \in \mathcal{D}\}$ is the set of all possible n-point samplings of a unit-length curve. Any two elements of $\mathcal{D}$ are defined to be equivalent if they result in the same sampling, i.e. $\gamma_1 \sim \gamma_2$ if $\gamma_1(\tau_0) = \gamma_2(\tau_0)$. This equivalence relation partitions $\mathcal{D}$ into the set of equivalence classes $\mathcal{D}/\sim$. Elements of $\mathcal{D}/\sim$ are sets of type:

$$[\gamma] = \{\tilde{\gamma} \in \mathcal{D} | \tilde{\gamma}(\tau_0) = \gamma(\tau_0)\} \quad \subset \mathcal{D} \ .$$

An observation of a curve α_t, sampled at $\tau \in \mathcal{T}$, can be modeled as:

$$\alpha^d(i) = \alpha_t(\tau(i)) + w(i) \ , \quad t = 1, 2, \qquad (10)$$

where α_t is the true underlying curve and w is white Gaussian noise. Shown in Figure 7 is a pictorial example of this setup. The top row shows the two templates, α_1 and α_2, associated with the two classes and the middle row shows four examples of α^d (α_1 is the underlying curve) with increasing variance of w from left to right.

For a given observation, the posterior probability that it belongs to the j^{th} class is: $P(j|\alpha^d) = P(j) \int_{\tau \in \mathcal{T}} P(\alpha^d|\tau, j) P(\tau|j) d\tau$. Using the action of $\mathcal{D}$ on $\mathcal{T}$, we can rewrite this as an integral on $\mathcal{D}/\sim$:

$$P(j) \int_{\gamma \in \mathcal{D}/\sim} P(\alpha^d|[\gamma], j) P([\gamma]|j) d[\gamma] \ , \qquad (11)$$

Under certain conditions, the integral in Eqn. 11 can be approximated by evaluating the integrand at the maximum likelihood estimate of $[\gamma]$. Let $\hat{[\gamma]}_j$ be the optimal sampling of the template α_j according to:

$$[\hat{\gamma}_j] = \underset{\mathcal{D}/\sim}{\operatorname{argmax}} P(\alpha^d|[\gamma], j) \ .$$

This optimization is performed using a dynamic programming algorithm, which automatically selects a representative element from $[\hat{\gamma}_j]$: the piecewise linear curve obtained by linearly connecting the points $\{(\tau_0(i), \ \hat{\gamma}_j(\tau_0(i))), i = 1, 2, \ldots, n\}$.

So the posterior probability can be approximated by:

$$P(j|\alpha^d) \approx P(j) P(\alpha^d|[\hat{\gamma}_j], j) P([\hat{\gamma}_j]|j) \ .$$

For our experiment, the terms on the right were chosen as follows:

- For Gaussian noise w, the likelihood function becomes $P(\alpha^d|[\gamma], j) \propto \exp(-E_j([\gamma]))$, where

$$E_j([\gamma]) = \sum_{i=1}^{n} \left(|\alpha^d(i) - \alpha_j(\gamma(\tau_0(i)))|^2 \right) \ . \qquad (12)$$

- The prior on γ is defined as follows. Using the techniques presented in Section 4, we can compute the average re-parametrization function $\bar{\gamma}_j$ associated with prior observations in each class and form a prior $P([\gamma]|j) \propto e^{-d(\gamma, \bar{\gamma}_j)^2}$ on $\mathcal{D}$. One can also use the term $e^{-d(\gamma, id)^2}$, which uses the squared distance from the identity in Φ as the prior energy. Here γ stands for the piecewise-linear representative of the class $[\gamma]$. These distances are computed using the square-root representations of γ and id (or $\bar{\gamma}$) in Ψ and then using the geodesic distance between them in Ψ (Eqn. 5).

The bottom row of Figure 7 shows the results of this binary classification performance in two cases: one when only the likelihood term in Eqn. 12 is used (broken line) and when the full posterior is used (solid line). For the left plot, the underlying sampling $\tau = \gamma(\tau_0)$ was closer to τ_0 than for the plot on the right. The classification performance was computed using 1000 Monte

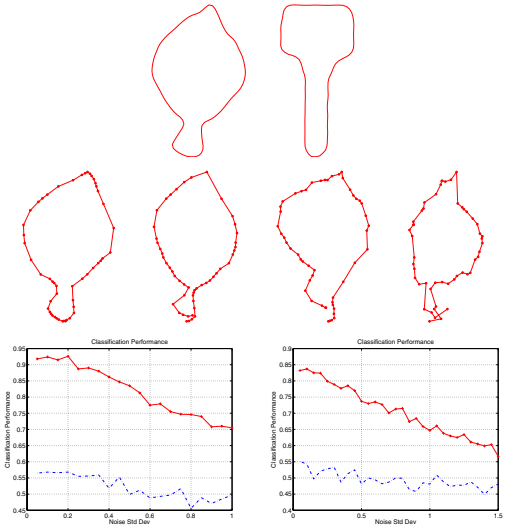

Figure 7. Top row: Templates for the two object classes. Middle row: Observations of class 1 template at random γ_0s and increasing noise from left to right. Bottom row: Plot of classification performance versus the observation noise. Left panel shows the case when γ_0 is much closer to *id* and the right panel shows the opposite case. The broken line denotes the maximum-likelihood solution and the solid line denotes the maximum a-posterior solution.

Carlo trials at each noise level. This experiment clearly shows the utility of the prior term $P([\gamma]|j)$ in Eqn. 12 in the classification process. Furthermore, our framework computes this distance very efficiently as a distance on a unit sphere.

6. Summary

We have proposed a "spherical" version of the Fisher-Rao metric for imposing a Riemannian structure on a collection of related spaces: the space of pdfs, time-warping functions, re-parametrization functions, etc. The proposed metric is both computationally and analytically simpler, and allows efficient computation of statistics on a larger class of functions than previously used metrics. We have demonstrated this idea using an application in planar shape classification.

Acknowledgements

This work was partially supported by EU project MUSCLE (FP6-507752), by the MUSCLE "Shape modelling" E-team, and by INRIA/Florida State University Associated Team "SHAPES". Additionally, Anuj Srivastava and Shantanu Joshi were supported in part by grants ARO W911NF-04-1-0268, AFOSR FA9550-06-1-0324, and ARO W911NF-04-1-0113. Anuj Srivastava was also supported by a Visiting Professorship from INRIA in summer 2006.

References

[1] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*. Oxford University Press, 2000. 2

[2] N. N. Čencov. *Statistical Decision Rules and Optimal Inferences*, volume 53 of *Translations of Mathematical Monographs*. AMS, Providence, USA, 1982. 2

[3] B. Jahne, H. Haubecker, and P. Geibler. *Handbook of Computer Vision and Applications, Volume 2*. Academic Press, 1999. 2

[4] H. Karcher. Riemann center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977. 6

[5] S. J. Maybank. Detection of image structures using the Fisher information and the Rao metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1579–1589, 2004. 2

[6] S. J. Maybank. The Fisher-Rao metric for projective transformations of the line. *Int. J. Comput. Vision*, 63(3):191–206, 2005. 2

[7] P. W. Michor and D. Mumford. Riemannian geometries on spaces of plane curves. *Journal of the European Mathematical Society*, 8:1–48, 2006. 2

[8] M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *Intern. Journal of Computer Vision*, 41(1/2):61–84, 2002. 2

[9] W. Mio, D. Badlyans, and X. Liu. In *EMMCVPR 2005*, volume 3757 of *Lecture Notes in Computer Science*, pages 18–33. Springer, 2005. 4, 5

[10] A. Peter and A. Rangarajan. A new closed-form information metric for shape analysis. In *Proc. of MICCAI, Copenhagen*, 2006. 2

[11] A. Peter and A. Rangarajan. Shape matching using the fisher-rao riemannian metric: Unifying shape representation and deformation. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2006. 2

[12] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int'l Journal of Computer Vision*, 40(1):49–71, 2000. 1

[13] C. Rao, M. Shah, and T. F. Syeda-Mahmood. Invariance in motion analysis of videos. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, pages 518–527, 2003. 1

[14] E. Spellman, B. C. Vemuri, and M. Rao. Using the KL-center for efficient and accurate retrieval of distributions arising from texture images. In *Proc. of IEEE CVPR*, volume 1, pages 111–116, 2005. 1

[15] A. Veeraraghavan and A. K. Roy-Chowdhury. The function space of an activity. In *Proc. of IEEE CVPR*, volume 1, pages 959–968, 2006. 1

[16] S. C. Zhu, Y. N. Wu, and D. Mumford. "Minimax entropy principles and its application to texture modeling". *Neural Computation*, 9(8):1627–1660, November 1997. 1