Geodesic Finite Mixture Models

Edgar Simo-Serra
esimo@iri.upc.edu
Carme Torras
torras@iri.upc.edu
Francesc Moreno-Noguer
fmoreno@iri.upc.edu

Institut de Robòtica i Informàtica Industrial (CSIC-UPC)
08028, Barcelona, Spain

The use of Riemannian manifolds and their statistics has recently gained popularity in a wide range of applications involving non-linear data modeling. For instance, they have been used to model shape changes in the brain [1] and human motion [3]. In this work we tackle the problem of approximating the Probability Density Function (PDF) of a potentially large dataset that lies on a *known* Riemannian manifold. We address this by creating a completely data-driven algorithm consistent with the manifold, i.e., an algorithm that yields a PDF defined exclusively on the manifold.

In the proposed finite mixture model, we simultaneously consider multiple tangent spaces, distributed along the whole manifold as seen in Fig. 1. We draw inspiration on the unsupervised Expectation Maximization (EM) algorithm from [2], which given data lying in an Euclidean space, automatically computes the number of model components that Minimize a Message Length (MML) cost. By representing each component as a distribution on the tangent space at its corresponding mean on the manifold, we are then able to generalize the algorithm to Riemannian manifolds and at the same time mitigate the accuracy loss produced when using a single tangent space.

Given an input dataset, [2] starts by randomly initializing a large number of components. During the Maximization (M) step, the MML criterion is used to annihilate those components not well supported by the data. In addition, upon EM convergence, the least probable mixture component is also forcibly annihilated and the algorithm continues until a minimum number of components is reached.

In order to extend [2] to Riemannian manifolds, we define each mixture component as a normal distribution on its own tangent space $T_{\mu_k}\mathcal{M}$, with a mean μ_k and a concentration matrix $\Gamma_k = \Sigma_k^{-1}$:

$$p(x|\theta_k) \approx \mathcal{N}_{\mu_k}\left(0,\, \Sigma_k^{-1}\right)$$

where $\theta_k = (\mu_k, \Sigma_k)$. The mean μ_k is defined on the manifold $\mathcal{M}$, while the concentration matrix Γ_k is defined on the tangent space $T_{\mu_k}\mathcal{M}$ with the mean at the origin. Specifically, our algorithm proceeds as follows:

Let us assume we have K components after iteration $t-1$. Then, in the E-step we compute the *responsibility* that each component k takes for every sample x_i:

$$w_k^{(i)} = \frac{\alpha_k(t-1)p(x_i|\theta_k(t-1))}{\sum_{k=1}^{K} \alpha_k(t-1)p(x_i|\theta_k(t-1))} ,$$

for $k = 1,\ldots,K$ and $i = 1,\ldots,N$, and where $\alpha_k(t-1)$ are the relative weights of each component k.

In the M-step we update the weight α_k, the mean μ_k and covariance Σ_k for each of the components according to:

$$\alpha_k(t) = \frac{1}{N}\sum_{i}^{N} w_k^{(i)} = \frac{w_k}{N}, \qquad \mu_k(t) = \arg\min_{p} \sum_{i=1}^{N} d\left(\frac{N}{w_k} w_k^{(i)} x^{(i)},\, p\right)^2$$

$$\Sigma_k(t) = \frac{1}{w_k}\sum_{i=1}^{N}\left(\log_{\mu_k(t)}(x^{(i)})\right)\left(\log_{\mu_k(t)}(x^{(i)})\right)^{\top} w_k^{(i)}$$

where $d(\cdot,\cdot)$ is the geodesic distance between two points and $log_{\mu}(\cdot)$ is an operator that maps a point from the manifold $\mathcal{M}$ to the tangent space $T_p\mathcal{M}$ at point μ.

After each M-step, we eliminate the components whose accumulated responsibility w_k is below a threshold. A score for the remaining components based on MML is then computed. This EM process is repeated until convergence of the score or until reaching a minimum number of components K_{min}. If this number is not reached, the component with the least responsibility is eliminated and the EM process is repeated. Finally, the configuration with minimum score is retained, yielding a resulting distribution with the form

$$p(x|\theta) = \sum_{k=1}^{K} \alpha_k p(x|\theta_k) .$$

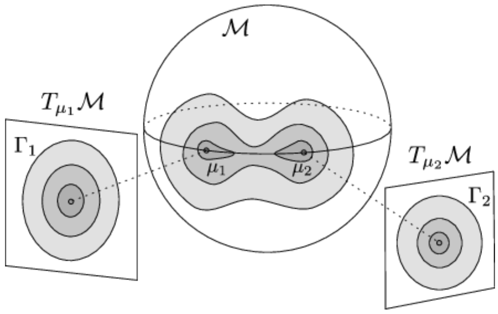
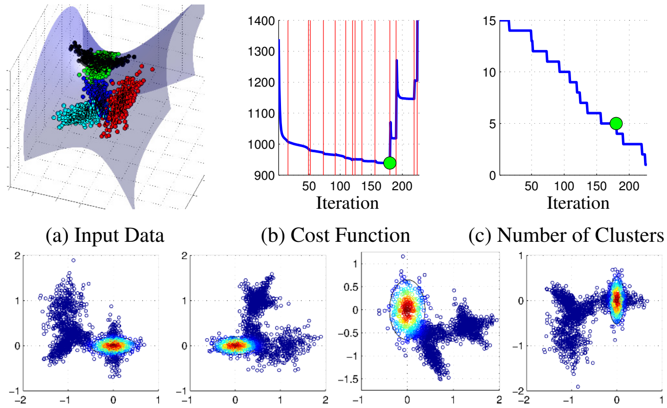


Figure 1: Illustration of the proposed mixture model approach. Each mixture component has its own tangent space, ensuring the consistency of the model while minimizing accuracy loss.



(a) Input Data (b) Cost Function (c) Number of Clusters

(d) Cluster #1 (e) Cluster #2 (f) Cluster #3 (g) Cluster #4

Figure 2: **Quadratic Surface Example.** (a) Section of the manifold with the input data. (b) Evolution of the cost function where vertical lines represent iterations in which a cluster is annihilated. The optimal mixture is marked with a green dot. (c) Evolution of the number of clusters. Some of the clusters from the solution are shown in (d) to (g).

We validate our method by providing extensive results on both synthetic and real examples. In particular, we show results on synthetic examples of a sphere and a quadric surface (see Fig. 2), and on a large and complex dataset of human poses, where the proposed model is used as a regression tool for hypothesizing the geometry of occluded parts of the body. We show that our approach outperforms the traditionally used Euclidean Gaussian Mixture Model, von Mises distributions and approaches using a single tangent space.

[1] B. C. Davis, E. Bullitt, P. T. Fletcher, and S. Joshi. Population Shape Regression from Random Design Data. In *International Conference on Computer Vision*, 2007.

[2] M. Figueiredo and A. Jain. Unsupervised Learning of Finite Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):381–396, 2002.

[3] S. Sommer, F. Lauze, S. Hauberg, and M. Nielsen. Manifold Valued Statistics, Exact Principal Geodesic Analysis and the Effect of Linear Approximations. In *European Conference on Computer Vision*, 2010.