Shape Analysis for Complex Systems using Information Geometry tools

Angela De Sanctis, Antonio Gattone

a.desanctis@unich.it, gattone@Economia.uniroma2.it

Università “G. d’Annunzio”, Università “Tor Vergata”

Introduction: Differential Manifolds

The **differential manifolds** are the object of study in the **Differential Geometry**.

They are **topological spaces** locally euclidean then mapped by local coordinates. Differential Geometry proves that all the consequent analysis is intrinsic, that is it does not depend on the choice of the coordinates.

For example the M-**Shere** is locally equivalent to the M-dimensional euclidean space but not globally.

Introduction: Statistical Manifolds

The **statistical manifolds** are the object of study in the **Information Geometry**.

They are **families of probability density** with its local coordinates defined by the model parameters

$$F = \{p(x,\theta)\} \sim \Theta = \{\theta\}$$

For example, the **normal distribution**:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} \tag{1}$$

is univocally identifiable with the parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, where these represent respectively the mean and the standard deviation.

Introduction: Statistical Manifolds

Then we can identify the family of the normal distributions with the half-plane

$$\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$$

More in general a bivariate Gaussian density can be represented as a single point with coordinates $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ on 4-dimensional manifold which is the cartesian product $\Theta_1 \times \Theta_2$ of two half planes.

The most famous statistical manifolds are the Exponential Families, among those there are the Normal and Poisson Families and also the Mixture Family.

Introduction: Fisher-Rao metric

We recall that a **riemannian metric** on a manifold means a metric "compatible" with the system of local coordinates.

From Information Geometry we also know that **Fisher information matrix** induces a riemannian metric on the statistical manifold, called the **Fisher-Rao metric** with metric tensor:

$$g_{ij}(\theta) = \int p(x/\theta)\frac{\partial}{\partial\theta^i}\log p(x/\theta)\frac{\partial}{\partial\theta^j}\log p(x/\theta)dx \qquad (2)$$

$$ds^2 = \sum_{i,j=1}^{M} g_{ij}(\theta)\theta^i\theta^j$$

It is possible to prove that it is the only one consistent with the Maximum Likelihood Principle.

Introduction: Fisher-Rao metric

For the family of the normal distributions, simple calculations lead to:

$$g_{11}(\mu,\sigma) = \frac{1}{\sigma^2}$$

$$g_{22}(\mu,\sigma) = \frac{2}{\sigma^2}$$

$$g_{12}(\mu,\sigma) = g_{21}(\mu,\sigma) = 0$$

More in general, for bivariate gaussian densities $\theta_i = (\mu_i, \sigma_i)$, $(i = 1, 2)$ the Fisher-Rao metric is $ds^2 = \sum_{i=1}^{2} \left[(d\mu_i)^2 + 2(d\sigma_i)^2\right] / \sigma_i^2$

Geodesics distance

It thus follows that, in this case, the distance with respect to the Fisher-Rao metric is

$$d\left\{(\mu_1,\sigma_1)(\mu_2,\sigma_2)\right\} = \sqrt{d_1^2 + d_2^2}$$

where

$$d_i = \sqrt{2} cosh^{-1}\left[\frac{(\mu_{1i}-\mu_{2i})^2 + 2\sigma_{1i}^2 + 2\sigma_{2i}^2}{4\sigma_{1i}\sigma_{2i}}\right]$$

Geodesics

Differential Geometry gives us a concept of **curvature** which is intrinsic that is depends only on the manifold, without considering how it embeds in an euclidean space (Egregium Theorem).

Besides in a riemannian manifold, given two points, there exists only one curve, called **geodesic**, which connects them locally minimizing the distance.

Geodesics

On a statistical manifold the distance is the Fisher information, therefore it means such a curve is what we can obtain by the available information.

It is possible to prove that, with respect to Fisher metric, for the Normal Family the geodesics are those of the **Hyperbolic Plane** that is half-circles with the center on μ-axis and vertical lines.

Patterns

One of the most important properties of Complex Systems is the forming of **patterns**: a band of fish in the sea, a flight of birds in the sky, the dunes in the desert are only few examples.

These patterns are not fixed in time but they evolve with the dynamics of the Complex System and vary according to its level of self-organization.

The goal is **modeling them statistically**, using Information Geometry tools, describing step by step their changes. In particular we wish to discover an index which is capable to understand the trend in the self-organization phenomenon and capture eventual crisis signals.

An example of Complex System

The **macula** is the central part of the retina in the eye, where the distinct vision occurs.

A frequent pathology, due to the degrading of that Complex System, is the irreversible reduction of the vision in people more of 65 years old; this pathology is called **macular degeneration due to the age**.

Ophthalmologists distinguish the evolution of the disease in **two phases**: an initial form, called dry, and a terminal form, which can be new-vascular or for atrophy.

The first one is characterized by the presence of some retina's damages, such as drusen, and areas of change of the pigmentation of the epithelium. In this phase people continues to have a discreet level of vision.

An example of Complex System

On the contrary, the second phase produces a serious loss of vision's capacity and it is characterized by the appearance of a central scotoma produced by the development of anomalous new-vessels near the macula. The atrophy is characterized by loss of the retina's normal stratus.

The diagnosis of macular degeneration is made by observing the ocular fundus with ophthalmoscopy and by using recent **imaging techniques**, such as fluor-angiography. Each of such techniques consents the view of the typical damages, their classification, supervision in time and this is very useful to value the efficacy of the therapies.

Shape

Consider a geometric object (e.g. a triangle, a collection of points in the plane, a human head).

The **shape** of the object consists of all information invariant under similarity transformations (i.e. translation, rotation and scaling).

Data from a shape are often realized as a set of points.

Landmarks

Many statistical methods allow us to extract some points, which are representative for the shape and are called **landmarks**. We can mean these landmarks as the centers of the surrounding points.

Landmark coordinates are stored in the $K \times M$ configuration matrix $\mathbf{X}$ with generic element x_{km} (m-coordinate of the k-th landmark), $k = 1 \dots K$, $m = 1, \dots M$

K=number of landmarks

M=dimension

GMM shape representation

In this representation, the shape of the configuration $\mathbf{X}$ is represented by a K-component **Gaussian mixture model** (GMM)

$$h(\mathbf{x};\theta) = \sum_{k=1}^{K} p_k f(\mathbf{x};\mu_k,\Sigma_k)$$

where $\mathbf{x}$ is a generic M-dimensional vector

$$f(\mathbf{x};\mu_k,\Sigma_k) = (2\pi)^{-\frac{M}{2}} |\Sigma_k|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_k)'\Sigma_k^{-1}(\mathbf{x}-\mu_k)\right\}$$

θ is the set of mixture parameters $\theta = \{p_k,\mu_k,\Sigma_k\}_{k=1}^{K}$

GMM Shape representation

Peter and Rangarajan identify K landmarks of the same shape with the mean points of a K-component Gaussian mixture model (GMM).

They consider planar shapes ($M = 2$) and set for $k = 1\ldots, K$:

$p_k = \frac{1}{K}$, $\mu_k = \{x_{k1}, x_{k2}\}$, $\Sigma_k = \boldsymbol{\Sigma} = \boldsymbol{\sigma}^2\mathbf{I}_2$

The landmark positions are the means for:

$$p(\mathbf{x}|\theta) = \frac{1}{2\pi\sigma^2 K}\sum_{i=1}^{K}\exp\{-\frac{\|\mathbf{x}-\mu_k\|^2}{2\sigma^2}\} \quad (3)$$

Peter and Rangarajan approach

In the absence of any a priori knowledge, it is acceptable to put in the model equal weight $\frac{1}{K}$ to every landmark.

The variance σ captures uncertainties that arise in landmark placement and/or the natural variability across a population of shapes. Peter and Rangarajan consider σ as a free parameter, which is isotropic across all components. Therefore they only use the **means** of a (GMM) as the manifold coordinates.

New approach

On the contrary, we consider **variances** as further coordinates for the landmarks of a complex shape, compatibly with the Information Geometry theory. It is clear that a landmark with a big variance informs us that it is not too much representative of its surrounding points. Numerical simulations prove that, in this case, the image shows blots, as a photocopy from a damaged machine.

On the other side, we also remove the isotropic hypothesis. Indeed, for example, when we have a photograph unfocused on a part of it, we can not state that the information we deduce is uniform.

New approach

The model of Peter and Rangarajan, even if numerically more simple, induced a loss of information in the Fisher sense. But, in some cases, it is reasonable to use that, in particular when the change of the shape is due to **external forces**. Indeed the Authors refer to "deformation of the external space" and unify representation and deformation.

On the contrary, we are interested in the natural evolution of the shape produced by **internal forces** to the system. This is very important, for example in medicine, indeed often the dimmed or stained imagine is the warning of some problems for the involved organ, as we saw in the macular degeneration.

The new model

Set for $k = 1, \ldots, K$:

$p_k = \frac{1}{K}$, $\mu_k = \{x_{k1}, x_{k2}\}$, $\Sigma_k = \sigma_k^2 \mathbf{I}_2$

with $\sigma_k^2 = (\sigma_{k1}^2, \sigma_{k2}^2)$ (variances of the k-th landmark coordinates)

Thus,

$$h(\mathbf{x}|\theta) = \frac{1}{2\pi K} \sum_{k=1}^{K} |\Sigma_k|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{x} - \mu_k)' \Sigma_k^{-1} (\mathbf{x} - \mu_k)\} \qquad (4)$$

GMM Geodesic computation

The Information Matrix of the GMM density

$$h(\mathbf{x};\theta) = \sum_{k=1}^{K} p_k f(\mathbf{x};\mu_k,\Sigma_k)$$

is given by

$$g(\theta) = \sum_{k=1}^{K} p_k g_k(\theta_k)$$

$\theta_k = (\mu_k, \Sigma_k)$ are the parameters of the k-th component of the mixture (we don't consider the term related to the probabilities p_k).

Application: cluster analysis of shapes

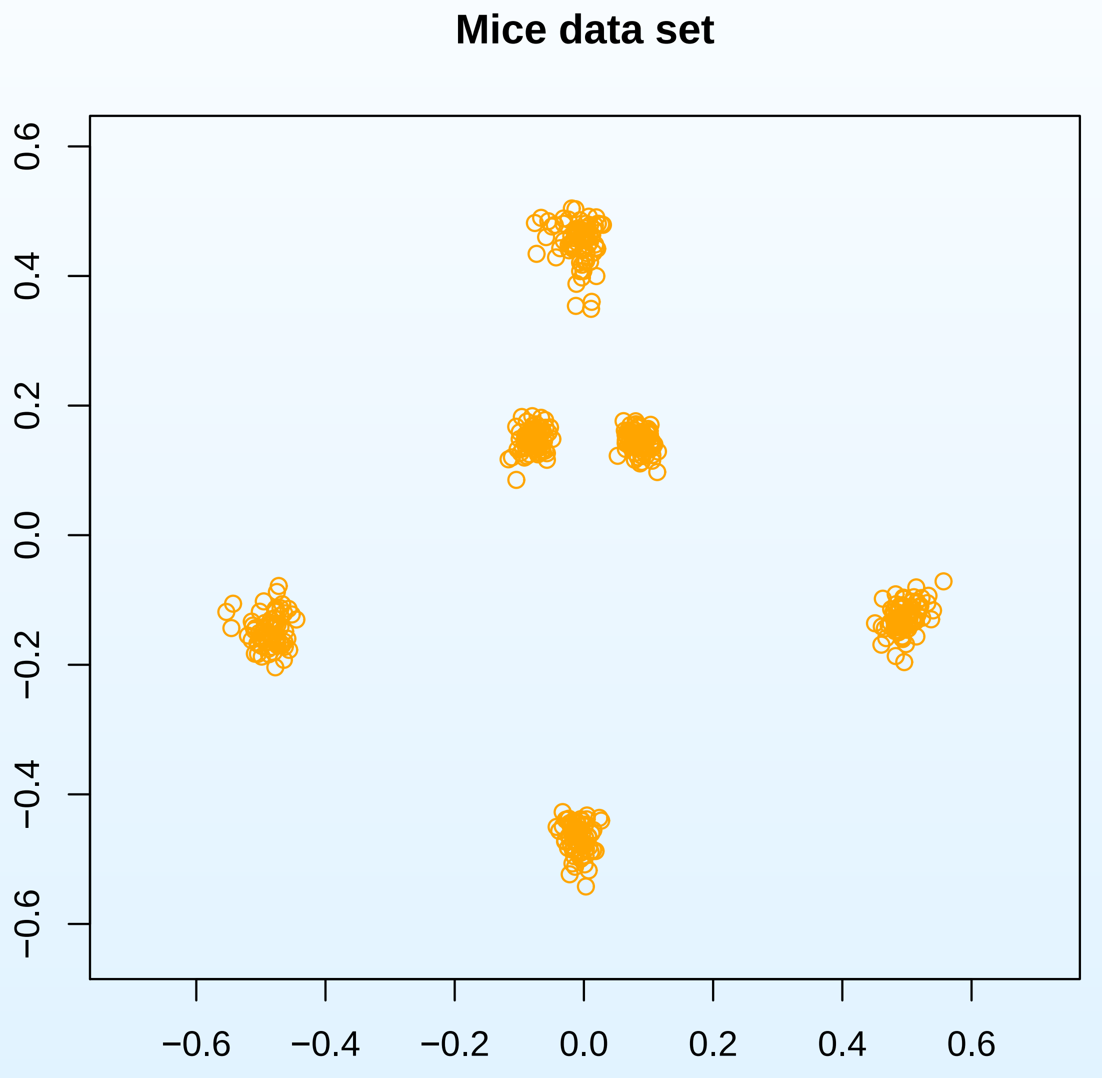
Data set from Dryden and Mardia (1998): **T2 mouse vertebra data**

The second thoracic vertebra is a bone lying along the backbone of the mouse and 6 landmarks have been identified by an expert along its outline.

We consider three groups: small, control and large groups of mice, respectively.

We use the **geodesic distance** and cluster the 76 mice.

T2 mouse vertebra data



T2 mouse vertebra data

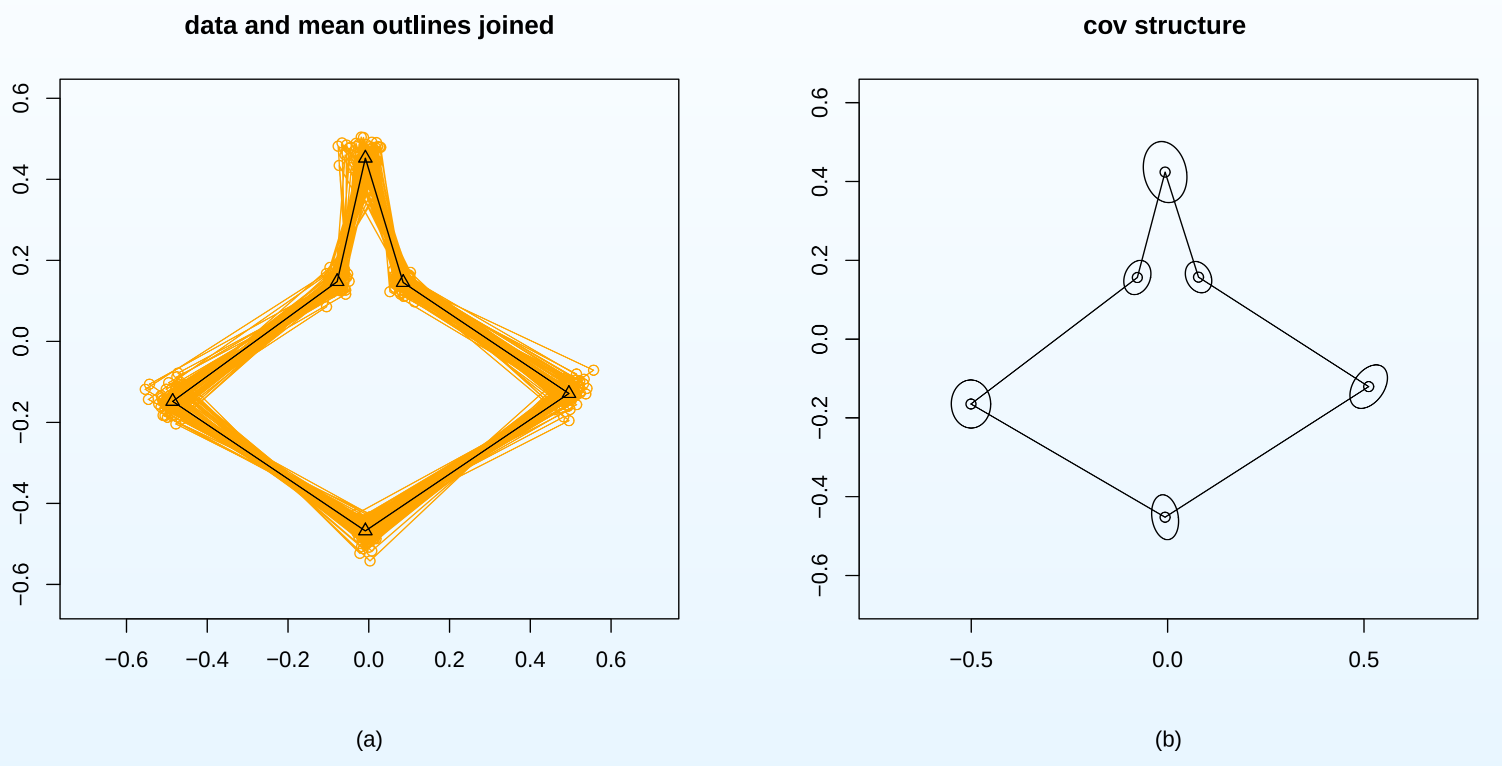
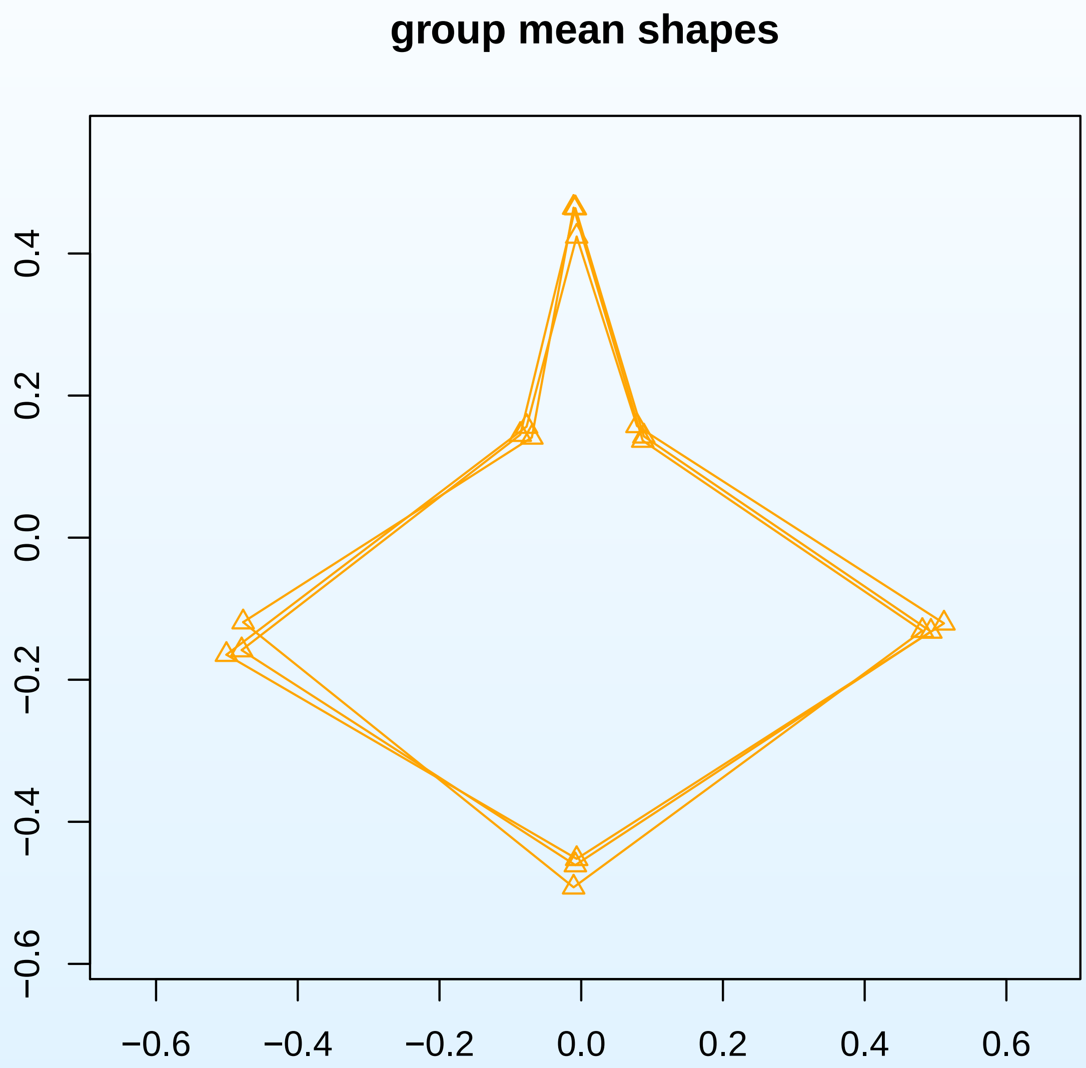


Figure 2:

T2 mouse vertebra data



Cluster analysis on T2 mouse vertebra data

We consider the pairwise distances of all shapes in order to evaluate the discriminative power of the proposed Fisher-Rao metric.

Distance matrix was applied in a hierarchical clustering algorithm.

Three distances were used:

- the proposed Fisher-Rao distance (FR^*)
- Fisher-Rao distance with fixed variance (FR)
- the Procrustes distance (shape distance) (PD)

Results of the cluster analysis

Distance	PD	FR^*	$FR - Q_1$	$FR - Q_2$	$FR - Q_3$
aRand-index	0.223	0.964	0.867	0.478	0.478
classification error	0.381	0.013	0.052	0.118	0.118

Q1, Q2 and Q3 denote different values for the variance parameter (first, second and third quartile).

aRand is the adjusted Rand index of Hubert and Arabie (1985) for comparing partitions.

Possible applications

We deduce that the new model allow us a better reconstruction of the shape with respect to that of Peter and Rangarajan or the classical models of shape analysis.

As possible application we observe that, if two shapes are represented by mixture models, whose parameters map points on the statistical manifold, it is possible to use Fisher-Rao metric to construct a geodesic between them which will inform us on the **intermediate shapes** (landmarks and their variances). That intrinsic path will drive the reconstruction of the real intermediate shapes in the external space.

Possible applications

We note that $\sigma^i(t)$ analysis gives us information regarding the dispersion of the real points of the shape around their means $\mu^i(t)$, when t is varying.

If $\sigma^i(t)$ increases in time, we lose the detailed resemblance to the original shape and, when it is a complex pattern, we can deduce a **loss of the self-organization** as connecting phenomenon of the system.

Besides, the study of the instantaneous speed of $\theta^i(t)$ allows us, in the short time, a forecast of the evolution of the pattern and of the eventual tendency to break up of the system.

Bibliography 1

1. S. Amari, Information Geometry of the EM and em Algorithms for Neural Networks, Neural Networks 8, 9, (1995) 1379-1408.

2. S. Amari and H. Nagaoka, Methods of Information Geometry, 191, Oxford University Press (2000).

3. C. S. Bertuglia and F. Vaio, Nonlinearity, Chaos, and Complexity: The Dynamics of Natural and Social Systems, Oxford University Press (2005).

4. I. L. Dryden and K. V.Mardia, Statistical Shape Analysis, John Wiley & Sons, London (1998).

Bibliography 2

5. W. J. Krzanowski, Rao's Distance between Normal Populations That Have Common Principal Components, Biometrics, 52, 4 (1996).

6. M.K. Murray and J. W. Rice, Differential Geometry and Statistics, Chapman and Hall (1984).

7. G. Nicolis, Introduction to Nonlinear Science, Cambridge University (1995).

8. A. Peter and A. Rangarajan, Information Geometry for landmark shape analysis: unifying shape representation and deformation, IEEE Transactions on pattern analysis and machine intelligence 31, 2 (2009).