Results in Applied Mathematics 18 (2023) 100363

Contents lists available at ScienceDirect

Results in Applied Mathematics

journal homepage: www.elsevier.com/locate/results-in-applied-mathematics

Alpha geodesic distances for clustering of shapes

Angela A. De Sanctis [a,*], Stefano A. Gattone [a], Fotios D. Oikonomou [b]

[a] University "G. d'Annunzio" of Chieti-Pescara, viale Pindaro 42, Pescara, 65127, Italy
[b] University of Patras, Greece

ARTICLE INFO

Article history:
Received 22 December 2022
Received in revised form 15 February 2023
Accepted 21 February 2023
Available online xxxx

MSC:
53Z05
53Z30

Keywords:
Shape analysis
Clustering
Fisher metric
α-connections
Geodesic distance

ABSTRACT

According to Information Geometry, we represent landmarks of a complex shape, as probability densities in a statistical manifold where geometric structures from α-connections are considered. In particular the 0-connection is the Riemannian connection with respect to the Fisher metric. In the setting of shapes clustering, we compare the discriminative power of different shapes distances induced by geodesic distances derived from α-connections. The methodology is analyzed in an application to a data set of aeroplane shapes.

1. Introduction

In the last decades, shapes clustering has played a very important role mostly in medical imaging, computer vision and geometric morphometrics.

Since the shape space is invariant under similarity transformations, that is translations, rotations and scaling, the Euclidean distance on such a space is not really meaningful. Therefore, in Shape Analysis [1], in order to apply standard clustering algorithms to planar shapes, the Euclidean metric has to be replaced by suitable metrics of the shape space. Examples were provided in [2,3], where the Procrustes distance was integrated in standard clustering algorithms such as the k-means. Similarly, [4] applied standard hierarchical or k-means clustering using dissimilarity measures based on the inter-landmark distances. In a model-based clustering framework, [5] and [6] developed a mixture model of offset-normal shape distributions. A more recent literature considers the shape-analysis of curves and functions. Using tools from Differential Geometry and Functional Data Analysis, it provides a rigorous and complete methodological setting in a non-parametric context. We refer the interested reader to [7] for more details.

Since often the profile of an object is not clearly defined, as for example in some medical images, we follow a different approach. We deal with objects whose shapes are based on landmarks [8–10]. These objects can be obtained by medical imaging procedures, curves defined by manually or automatically assigned feature points or by a discrete sampling of the object contours. We assume that each landmark is modeled via a bivariate Gaussian density, where the means are the landmark geometric coordinates and capture uncertainties that arise in landmark placement while the variances inform on the variability across the population of shapes under study. We regard the space of bivariate Gaussian densities as a statistical manifold [11] with the local coordinates given by the model parameters.

* Corresponding author.
E-mail addresses: a.desanctis@unich.it (A.A. De Sanctis), gattone@unich.it (S.A. Gattone), pheconom@upatras.gr (F.D. Oikonomou).

https://doi.org/10.1016/j.rinam.2023.100363

As it is well known, any hierarchical clustering algorithm uses as input the pairwise distances of all possible pairs of points under study. Then, we define distances between landmarks which are induced by the geodesics (geodesic distances) from the natural geometric structures of Information Geometry [12]. We already considered the Fisher metric as a Riemannian metric on the statistical manifold of the Gaussian densities [13,14]. The induced geodesic distance is related with the minimization of information in the Fisher sense. In this paper we generalize the results considering the whole family of α-connections, which includes, for $\alpha = 0$, the Fisher metric as Riemannian connection. By means of a real application, our goal is to evaluate the discriminative power of the different shapes distances obtained by varying the parameter α.

In Section 2, we recall the main definitions and the fundamental results, which concern the geometry of α-connections according to Information Geometry.

Then, in Section 3, clustering of shapes induced by geodesic distances derived from the α-connections is introduced. Two different generalized K-means algorithms, Type 1 and Type 2 are developed in Section 4. While in the Type 1 algorithm the landmark coordinates variances are assumed isotropic across the clusters, in Type 2 the variances are allowed to vary among the clusters. Section 5 presents an application on aeroplane shapes in which the discriminative power of the different α - connections shapes distances is evaluated. Conclusions are given in Section 6.

2. Geometrical structures for a statistical manifold

From Differential Geometry we recall that an affine connection on a manifold N of dimension n is a mapping $\nabla : T(N) \times T(N) \longrightarrow T(N)$, where $T(N)$ is the linear space of the vector fields on N, which satisfies the following conditions:

$$\nabla_{X+Y}Z = \nabla_X Z + \nabla_Y Z$$

$$\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z$$

$$\nabla_{fX}Y = f\nabla_X Y$$

$$\nabla_X fY = f\nabla_X Y + (Xf)Y$$

where Xf denotes the function $p \to X_p f$.

We may define Γ_{ij}^k the connection coefficients of ∇ with respect to some coordinate system θ^i to be the n^3 functions determined by the equation: $\nabla_{\partial_i}\partial_j = \Gamma_{ij}^k \partial_k$ where $\partial_i = \frac{\partial}{\partial\theta^i}$. In general, given $X = X^i\partial_i$ and $Y = Y^i\partial_i$, we may write

$$\nabla_X Y = X^i(\partial_i Y^k + Y^j \Gamma_{ij}^k)\partial_k$$

Let ∇ be an affine connection on a Riemannian manifold $(N, g = \langle , \rangle)$. If for all vector fields X, Y, Z it satisfies:

$$Z\langle X, Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle$$

we say that ∇ is a metric connection with respect to g. Using the coordinate expressions of g and ∇, we can rewrite this condition as

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i}$$

where

$$\Gamma_{ij,k} = \Gamma_{ij}^p g_{pk}$$

and the above equation becomes

$$\partial_k g_{ij} = \Gamma_{ki}^p g_{pj} + \Gamma_{kj}^p g_{pi}.$$

We call a connection, which is both metric and symmetric, that means $\Gamma_{ij,k} = \Gamma_{ji,k}$, the Riemannian connection or the Levi-Civita connection with respect to g. For a given g, it exists uniquely, indeed combining both the conditions we obtain:

$$\Gamma_{ij,k} = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij})$$

Besides it is possible to prove that, given any connection ∇ on a Riemannian manifold (N, g), there exists a unique connection ∇^*, called the dual connection of ∇, that satisfies, for all vector fields X, Y, Z, the following identity:

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X^* Z\rangle$$

Then we introduce a totally symmetric cubic $(0, 3)$-tensor (i.e. 3-covariant tensor): $C(X, Y, Z) = \langle \nabla_X Y - \nabla_X^* Y, Z\rangle$.

We define a statistical manifold (N, g, C) a manifold N equipped with a metric tensor g and a totally symmetric cubic tensor C.

Let P a family of probability density functions $p(x;\theta)$ parameterized by $\theta \in \mathbf{R}^k$. It is well known that we can endowed it with a structure of manifold, whose local coordinates are the parameters of the family. As an example, we can consider the family of p-variate Gaussian densities:

$$f(x;\theta=(\mu,\Sigma))=(2\pi)^{\frac{-p}{2}}(\det\Sigma)^{\frac{-1}{2}}\exp\{\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

where $x=(x_1,x_2,\ldots,x_p)^T$, $\mu=(\mu_1,\mu_2,\ldots,\mu_p)^T$ is the mean vector and Σ the covariance matrix. Note that the parameter space has dimension $k=p+\frac{p(p+1)}{2}$. In particular, we are interested in the case $p=2$.

Natural structural conditions in such a manifold are (uniquely) geometrical met by the Riemannian metric induced by Fisher information matrix (Fisher metric) and a family of affine connections (α - connections), which induce on P a structure of statistical manifold.

Precisely, for the Fisher metric, the tensor metric is defined as:

$$g_{ij}(\theta)=E_\theta(\partial_i l_\theta\partial_j l_\theta)=\int\partial_i l(x;\theta)\partial_j l(x;\theta)p(x;\theta)dx$$

where $\partial_i=\frac{\partial}{\partial\theta^i}$, $l_\theta(x)=l(x;\theta)=\ln p(x;\theta)$ and $E_\theta(f)=\int f(x)p(x;\theta)dx$.

Besides, for every real number α, we put:

$$(\Gamma^{(\alpha)}_{ij,k})_\theta=E_\theta[(\partial_i\partial_j l_\theta+\frac{1-\alpha}{2}\partial_i l_\theta\partial_j l_\theta)(\partial_k l_\theta)]$$

then we can define an affine connection $\nabla^{(\alpha)}$ on P by

$$\langle\nabla^{(\alpha)}_{\partial_i}\partial_j,\partial_k\rangle=\Gamma^{(\alpha)}_{ij,k}$$

where $g=\langle,\rangle$ is the Fisher metric. We call $\nabla^{(\alpha)}$ the α-connection. We remark that the α-connection is clearly a symmetric connection. Also the α-connection and the $-\alpha$-connection are dual with respect to the Fisher metric.

It is possible to verify that

$$\nabla^{(\alpha)}=(1-\alpha)\nabla^{(0)}+\alpha\nabla^{(1)}=\frac{1+\alpha}{2}\nabla^{(1)}+\frac{1-\alpha}{2}\nabla^{(-1)}$$

which establishes the importance of some elements of this family of affine connections obtained for particular values of the parameter α.

Due to

$$\begin{aligned}\partial_k g_{ij}&=E_\theta[(\partial_k\partial_i l_\theta)(\partial_j l_\theta)]+E_\theta[(\partial_i l_\theta)(\partial_k\partial_j l_\theta)]+E_\theta[(\partial_i l_\theta)(\partial_j l_\theta)(\partial_k l_\theta)]\\&=\Gamma^{(0)}_{ki,j}+\Gamma^{(0)}_{kj,i}\end{aligned}\tag{1}$$

we can deduce that the 0-connection is the Riemannian connection with respect to the Fisher metric. In general $\nabla^{(\alpha)}$ is not metric if $\alpha\neq 0$.

Again from Differential Geometry, let N be an n-dimensional manifold and M an m-dimensional submanifold of N, with coordinate systems θ^i and u^a respectively. If $\partial_i=\frac{\partial}{\partial\theta^i}$ and $\partial_a=\frac{\partial}{\partial u^a}$, by ∇ affine connection on N we can define a directional derivative of a vector field Y on M along a vector of another vector field X on M. In general such a vector $\nabla_{X_p}Y$ is a tangent vector of N but not necessarily a tangent vector of M.

If, for every p in M, $\nabla_{X_p}Y$ is a tangent vector of M then ∇ is a covariant derivative on M and we say that M is autoparallel with respect to ∇. This is equivalent to there existing m^3 functions Γ^c_{ab} on M which satisfy $\nabla_{\partial_a}\partial_b=\Gamma^c_{ab}\partial_c$.

1-dimensional autoparallel submanifolds are called autoparallel curves or geodesics. For a curve $\gamma: t\mapsto\gamma(t)$ the previous condition can be rewritten $\nabla_{\dot\gamma(t)}\dot\gamma(t)=\Gamma(t)\dot\gamma(t)$. Because the 1-dimensional manifolds are necessarily "flat", for a suitable change of variable t, we obtain $\Gamma(t)\equiv 0$ and the equation reduces to $\nabla_{\dot\gamma(t)}\dot\gamma(t)=0$ that means

$$\ddot\gamma^k(t)+\dot\gamma^i(t)\dot\gamma^j(t)(\Gamma^k_{ij})_{\gamma(t)}=0$$

The geodesic with respect to a Riemannian connection is known to locally coincide with the shortest curve joining two points ξ,ξ', measuring length according to:

$$\|\gamma\|=\int_0^1\sqrt{g_{ij}\dot\gamma^i\dot\gamma^j}dt$$

where $\gamma(0)=\xi$ and $\gamma(1)=\xi'$.

In general, from an affine connection on a complete manifold N we can define a distance (geodesic distance) between two points ξ,ξ' of the manifold as the length of the geodesic joining these two points. Then, if $\gamma=\gamma(t)$ is such a curve, we put

$$d(\xi,\xi')=\|\gamma\|$$

Starting from the family of bivariate Gaussian densities with diagonal covariance matrices, which we can write, if $\theta = (\mu, \Sigma)$ with $\mu = (\mu_1, \mu_2)$ and $\Sigma = diag(\sigma_1, \sigma_2)$

$$p(x, y, \mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2\}$$

it is easy to prove that the Fisher matrix is given by:

$$g = diag[\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \frac{2}{\sigma_1^2}, \frac{2}{\sigma_2^2}]$$

For the Riemannian connection with respect to the Fisher metric $\nabla^{(0)}$, it is possible to calculate exactly the geodesics then we can deduce that the closed form of the geodesic distance $d_0 = d_F$ between two bivariate Gaussian densities with diagonal covariance matrices is the following [15]:

$$d_F(\theta, \theta') = \sqrt{2\sum_{i=1}^{2}\left(\ln \frac{|(\frac{\mu_i}{\sqrt{2}}, \sigma_i) - (\frac{\mu'_i}{\sqrt{2}}, -\sigma'_i)| + |(\frac{\mu_i}{\sqrt{2}}, \sigma_i) - (\frac{\mu'_i}{\sqrt{2}}, \sigma'_i)|}{|(\frac{\mu_i}{\sqrt{2}}, \sigma_i) - (\frac{\mu'_i}{\sqrt{2}}, -\sigma'_i)| - |(\frac{\mu_i}{\sqrt{2}}, \sigma_i) - (\frac{\mu'_i}{\sqrt{2}}, \sigma'_i)|}\right)^2} \quad (2)$$

where $\theta = (\mu, \Sigma)$ with $\mu = (\mu_1, \mu_2)$ and $\Sigma = diag(\sigma_1, \sigma_2)$, $\theta' = (\mu', \Sigma')$ with $\mu' = (\mu'_1, \mu'_2)$ and $\Sigma' = diag(\sigma'_1, \sigma'_2)$.

In this paper we are interested in the whole family of geodesic distances d_α induced by the α - connections $\nabla^{(\alpha)}$, for an arbitrary real number α, on the statistical manifold of bivariate Gaussian densities with diagonal covariance matrices.

We previously observed that, in the case $\alpha = 0$, the analytical expression of geodesics is known. The same happens if $\alpha = 1$. Indeed, since Gaussian densities constitute an exponential family which is flat with respect to the 1-connection, [16], geodesics may be expressed using linear equations with respect to affine coordinate systems, that are the canonical parameters. In the general case, for $\alpha \neq 0$ and for $\alpha \neq 1$, analytical expressions of the geodesics induced by these affine connections are not available so we need to use numerical techniques to obtain their approximations and deduce the geodesic distances.

3. Clustering of shapes

We will consider only planar objects, as for example a flat fish or a section of the skull. The "shape" of the object consists of all information invariant under similarity transformations, that is translations, rotations and scaling [1].

Data from a shape are often realized as a set of points. Many methods allow us to extract a finite number of points, which are representative of the shape and are called landmarks.

Suppose we are given a planar shape configuration, S, from a population of shapes. Shape S consists of a fixed number K of labeled landmarks

$$S = \{\mu_1, \mu_2, \ldots, \mu_K\} \quad (3)$$

with generic element $\mu_k = (\mu_{k1}, \mu_{k2})$ for $k = 1, \ldots, K$.

Following [13], the kth landmark, for $k = 1, \ldots, K$, may be represented by a bivariate Gaussian density as follows:

$$f(x; \theta_k) = (\mu_k, \Sigma_k) = (2\pi)^{-1}(\det \Sigma_k)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\} \quad (4)$$

with x being a generic 2-dimensional vector and Σ_k given by

$$\Sigma_k = diag(\sigma_{k1}^2, \sigma_{k2}^2) \quad (5)$$

where σ_{k1}^2 and σ_{k2}^2 are the variances of the k-th landmark.

We remark that, in the previous representation, the means are the geometric coordinates of the landmark and capture uncertainties that arise in landmark placement. The variances are "hidden" coordinates of the landmarks and reflect the natural variability across the population of shapes. Eq. (4) allows to assign to the k-th landmark the coordinates $\theta_k = (\mu_k, \sigma_k)$ on the 4-dimensional manifold which is the product of two upper half planes.

One way to compare shapes of different objects is to first register them on some common coordinate system for removing the similarity transformations [9,17]. Alternatively, Procrustes methods [18] may be used in which objects are scaled, rotated and translated so that their landmarks lie as close as possible to each other with respect to the Euclidean distance.

Let S and S' two planar shapes registered on a common coordinate system using Procrustes method. We parameterize them as follows: $S = (\theta_1, \ldots, \theta_K)$ and $S' = (\theta'_1, \ldots, \theta'_K)$.

The geodesic distances between landmarks allow to define a distance of the two shapes S and S'. Precisely a shape metric for measuring the difference between S and S' can be obtained by taking the sum of the geodesic distances between the corresponding landmarks, according to the following definition:

$$D(S, S') = \sum_{k=1}^{K} d(\theta_k, \theta'_k) \tag{6}$$

Then a classification of shapes, using in turn, as distance d, the geodesic distance d_F induced by Fisher metric and more generally the geodesic distances d_α induced by α - connections, can be done following the methodology that is recalled in the next paragraph.

4. K-means clustering algorithms

In [14] the shape distances are implemented in two different generalized K-means algorithms, Type 1 and Type 2. While in the Type 1 algorithm the landmark coordinates variances are assumed isotropic across the clusters, in Type 2 the variances are allowed to vary among the clusters.

The task is clustering a set of n shapes, $S_1, S_2, \ldots, S_n$ into G different clusters, denoted as $C_1, C_2, \ldots, C_G$. The computational effort of both algorithms is of the order of combinations $C(n, G)$ at the worst case scenario.

4.1. Type 1 algorithm

1 Initial step:
Compute the variability of the kth landmark coordinates $\boldsymbol{\sigma}_k^2 = (\sigma_{k1}^2, \sigma_{k2}^2)$, for $k = 1, \ldots, K$.
Randomly assign the n shapes, $S_1, S_2, \ldots, S_n$ into G clusters, $C_1, C_2, \ldots, C_G$.
For $g = 1, \ldots, G$ calculate the cluster center $c_g = (\theta_1^g, \ldots, \theta_K^g)$ with kth component $\theta_k^g = (\mu_{gk}, \sigma_k^2)$ obtained as $\theta_k^g = \frac{1}{n_g} \sum_{i \in C_g} \theta_k^i$, where n_g is the number of elements in the cluster C_g and θ_k^i is the kth coordinate of S_i given by $\theta_k^i = (\mu_{ik}, \sigma_k^2)$.
2 Classification:
For each shape S_i, compute the distances to the G cluster centers $c_1, c_2, \ldots, c_G$.
The generic distance between the shape S_i and the cluster center c_g is given by:

$$D(S_i, c_g) = \sum_{k=1}^{K} d(\theta_k^i, \theta_k^g).$$

Assign S_i to cluster h that minimizes the distance:

$$D(S_i, c_h) = \min_g D(S_i, c_g).$$

3 Renewal step:
Compute the new cluster centers of the renewed clusters $c_1, \ldots, c_G$.
The kth component of the gth cluster center c_g is defined as $\theta_k^g = \frac{1}{n_g} \sum_{i \in C_g} \theta_k^i$.
4 Repeat 2 and 3 until convergence.

4.2. Type 2 algorithm

1 Initial step:
Randomly assign the n shapes, $S_1, S_2, \ldots, S_n$ into G clusters, $C_1, C_2, \ldots, C_G$.
In each cluster compute the variability of the kth landmark coordinates $\boldsymbol{\sigma}_{gk}^2 = (\sigma_{gk_1}^2, \sigma_{gk_2}^2)$, for $k = 1, \ldots, K$ and $g = 1, \ldots, G$.
Calculate the cluster center $c_g = (\theta_1^g, \ldots, \theta_K^g)$ with kth component $\theta_k^g = (\mu_{gk}, \sigma_{gk}^2)$ obtained as $\theta_k^g = \frac{1}{n_g} \sum_{i \in C_g} \theta_k^i$ for $g = 1, \ldots, G$, where n_g is the number of elements in the cluster C_g and $\theta_k^i = (\mu_{ik}, \sigma_{gk}^2)$ for $i \in C_g$.
2 Classification:
For each shape S_i, compute the distances to the G cluster centers $c_1, c_2, \ldots, c_G$.
The generic distance between the shape S_i and the cluster center c_g is given by:

$$D(S_i, c_g) = \sum_{k=1}^{K} d(\theta_k^i, \theta_k^g).$$

Assign S_i to cluster h that minimizes the distance:

$$D(S_i, c_h) = \min_g D(S_i, c_g).$$

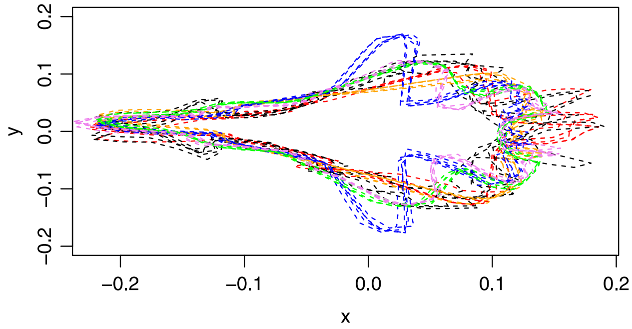

Fig. 1. Aeroplane shape classes.

3 Renewal step:
Update the variability of the kth landmark coordinates in each cluster by computing $\boldsymbol{\sigma}^2_{gk} = (\sigma^2_{gk_1}, \sigma^2_{gk_2})$, for $k = 1, \ldots, K$ and for $g = 1, \ldots, G$.
Calculate the new cluster centers of the renewed clusters $c_1, \ldots, c_G$.
The kth component of the gth cluster center c_g is defined as $\theta^g_k = \frac{1}{n_g} \sum_{i \in C_g} \theta^i_k$.

4 Repeat 2 and 3 until convergence.

5. Application

We consider a set of thirty aeroplanes (thirty shapes), of six different classes (six clusters), with each one of them comprised of sixty dots (sixty landmarks). These data were taken from "Anthony Bagnall, Jason Lines, William Vickers and Eamonn Keogh, The UEA & UCR Time Series Classification Repository, www.timeseriesclassification.com". Each cluster contains five aeroplanes. All the aeroplanes were first registered using Procrustes analysis [1]. The sample of registered shapes is depicted in Fig. 1.

According to Type 1 or Type 2 algorithm, we randomly assign the 30 shapes into 6 clusters and gradually reassign them to cluster that minimizes the distance D computed as above using the α - geodesic.

In order to find the α - geodesic for the α - connection $\gamma(t)$ from the given point $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ to the (also given) point $\theta' = (\mu'_1, \mu'_2, \sigma'_1, \sigma'_2)$ we suppose that,

$$\gamma^i(t) = \gamma^i(0) + t \frac{d\gamma^i}{dt}\bigg|_{t=0} + \frac{t^2}{2} \frac{d^2\gamma^i}{dt^2}\bigg|_{t=0}$$

for i=1,2,3,4 i.e. we consider the Taylor expansion of $\gamma^i(t)$ for up to second order. From the equations of the geodesics we have

$$\frac{d^2\gamma^i}{dt^2} = -\Gamma^{\alpha i}_{jk} \frac{d\gamma^j}{dt} \frac{d\gamma^k}{dt},$$

so the above expansion takes the form

$$\gamma^i(t) = \gamma^i(0) + t \frac{d\gamma^i}{dt}\bigg|_{t=0} - \frac{t^2}{2} \left(\Gamma^{\alpha i}_{jk} \frac{d\gamma^j}{dt} \frac{d\gamma^k}{dt} \right)\bigg|_{t=0}$$

with $\gamma(0) = \theta$. So, in order to have an approximate expression for the geodesic $\gamma(t)$, we want to know the first order derivatives $\frac{d\gamma^j}{dt}\Big|_{t=0}$ at t=0. These can be computed if we demand

$$\gamma(1) = \theta' = (\mu'_1, \mu'_2, \sigma'_1, \sigma'_2),$$

so we have a non-linear system of four equations with four unknowns which we solve numerically. Finally, we compute numerically the integral,

$$d(\theta, \theta') = \int_0^1 \sqrt{g_{ij} \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt}} dt,$$

from which we compute the distance D above. Both computations were done with mathematica. As a comparison, a K-means algorithm was implemented with the Procrustes distance [1], a well-known shape-distance commonly used for landmark-based planar shapes.

The results are depicted in Figs. 2, 3, 4 for Type 1 algorithm and Figs. 5, 6, 7 for Type 2. (in Type 2 algorithm we have set the value of variance to be 1E-9 when necessary, to avoid divisions by zero). Type 1 algorithm clearly outperforms

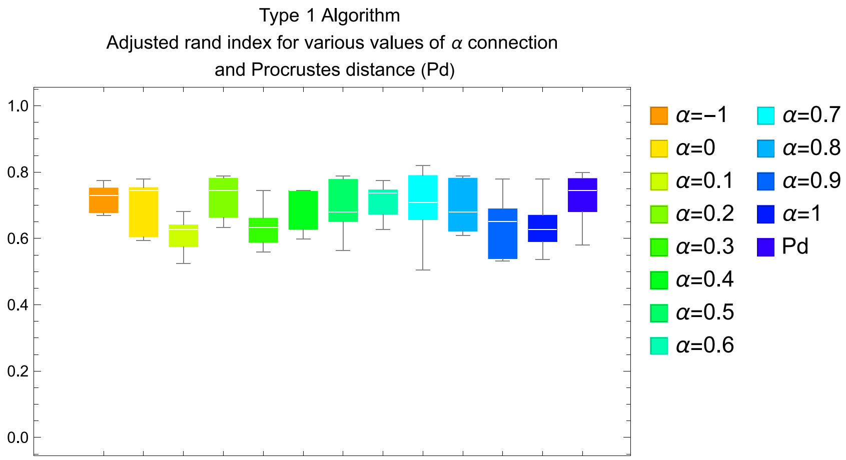

Fig. 2. Type 1 algorithm. Adjusted rand index for various types of distance.

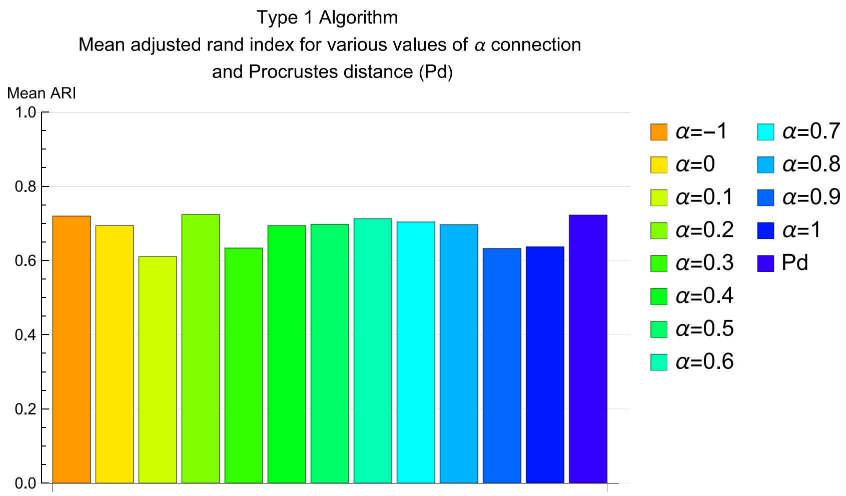

Fig. 3. Type 1 algorithm. Mean adjusted rand index for various types of distance.

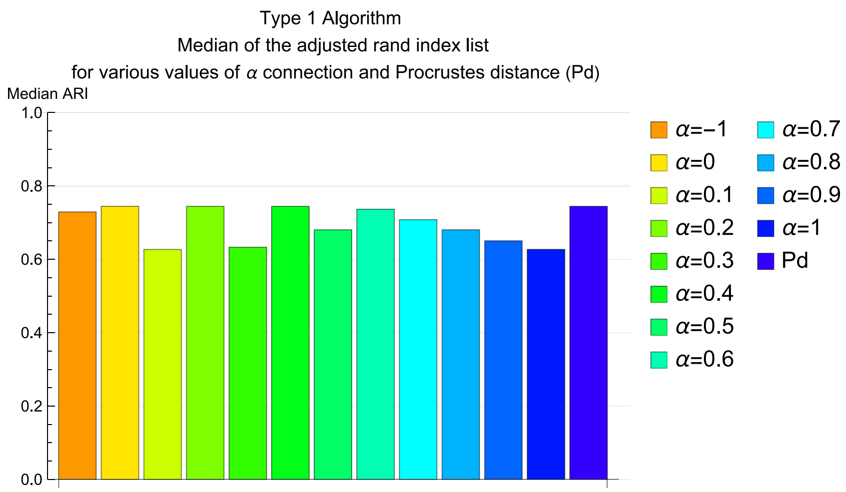

Fig. 4. Type 1 algorithm. Median of the adjusted rand index list for various types of distance.

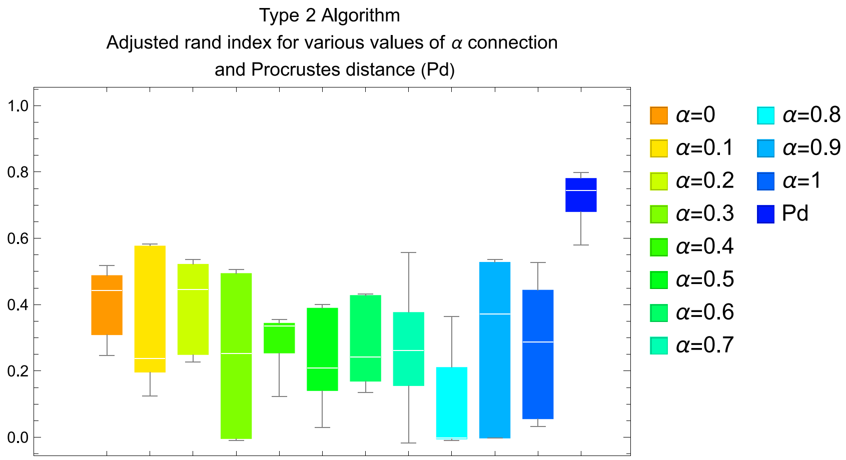

Fig. 5. Type 2 algorithm. Adjusted rand index for various types of distance.

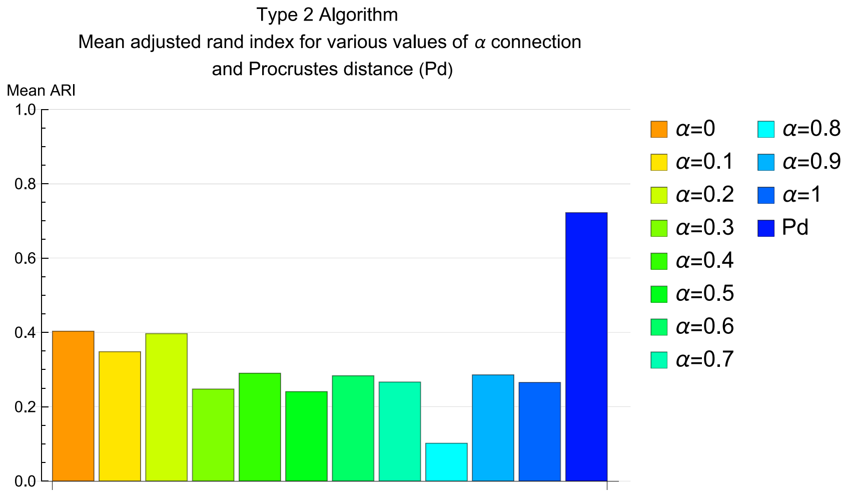

Fig. 6. Type 2 algorithm. Mean adjusted rand index for various types of distance.

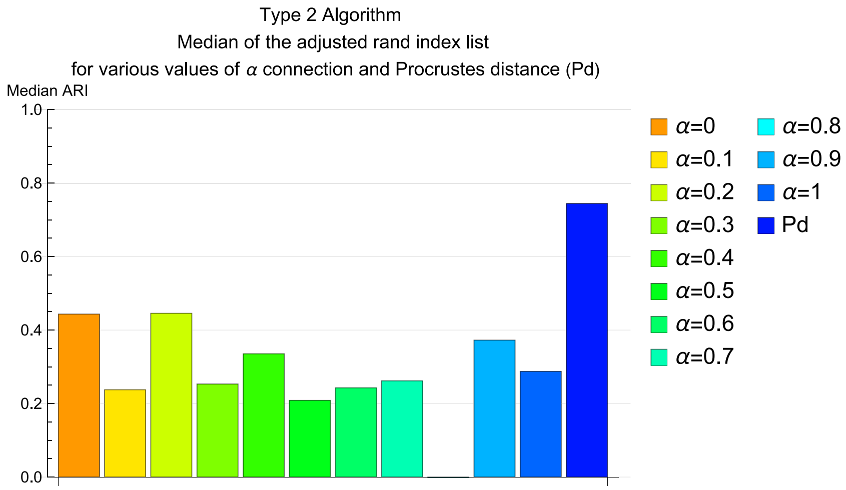

Fig. 7. Type 2 algorithm. Median of the adjusted rand index list for various types of distance.

Type 2. Indeed, in the data, the landmark variability does not change across the different clusters. Thus, estimating the variability in each cluster deteriorates the clustering results because of the small sample size. Precisely, in general, Type 1 algorithm uses only one estimate of the variance, assumed equal across the clusters, instead, Type 2 algorithm uses a different estimate of the variance for each cluster. Thus, if the variability is different across the clusters, then the Type 2 algorithm should outperform Type 1 and if the variability is homogeneous across the clusters, then, Type 1 should outperform Type 2.

Interestingly, the clustering results vary according to the value of α used in computing the distance. In this application, for example, neither the Fisher metric ($\alpha = 0$) nor the 1-connection delivers the best results, being the best α value equal to 0.2 which compares favourably with the results obtained with the Procrustes distance. The α value reflects the (kind of) "metric" most suitable to identify the cluster structure. Indeed, since $\nabla^{(\alpha)} = (1-\alpha)\nabla^{(0)} + \alpha\nabla^{(1)}$, varying α is as having a mixture between two geometries: one is like the Procrustes one since the geodesics are straight lines when $\alpha = 1$, and the other, when $\alpha = 0$, is that induced from the Fisher–Rao. The best alpha value will depend on the data and, in particular, to what "metric" is more relevant for identifying the cluster structure.

6. Conclusions

After having represented shapes based on landmarks as points in a statistical manifold, in this paper, we deal with shape distances derived from the geodesic distances induced by α - connections. Then, the discriminative power, with respect to the α parameter, of these shapes distances is evaluated in the setting of K-means clustering. An application, for a data set of aeroplane shapes, shows that the value of *alpha* affects the performance of the algorithm in terms of recovery of the true classification structure. This result, on the one hand, it enlarges the application of the α-connections in the context of cluster analysis; on the other, it arises the problem of how to select the *good* value of α in real applications where the ground truth is unknown. A possible way of handling this problem is to select the value of *alpha* by using internal validity indices such as the pseudo F index [19] or the GAP statistic [20]. The preliminary results have also shown the validity of the presented numerical approximation of the geodesics based on the quadratic Taylor expansion.

This work is a first foray into the introduction of a novel family of shape distances and it suffers from a few drawbacks which represent possible paths for future research.

Firstly, we have yet to perform a complete numerical comparison to the main state of the art distances used in shape analysis (for an update and complete view see [7]).

Secondly, the proposed Alpha geodesic distances are not invariant with respect to shape preserving transformations such as rotation, translation, and scaling. Most recent approaches incorporate at least one type of equivalence relation such as the representation based on the Square-Root-Velocity function (re-parameterization invariant) [21] used for geodesic computation between planar shapes. We remark, however, there is no arbitrariness in the results as long as one works with shapes coordinates, *i.e.* registered shapes aligned as closely as possible through linear mapping.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have used previously published data.

References

[1] Dryden IL, Mardia KV. Statistical shape analysis. London: John Wiley & Sons; 1998.
[2] Stoyan D, Stoyan H. A further application of D.G. Kendall's procrustes analysis. Biom J 1990;32:293–301.
[3] Amaral G, Dore L, Lessa R, Stosic B. K-means algorithm in statistical shape analysis. Commun Stat-Simul C 2010;39:1016–26.
[4] Lele S, Richtsmeier J. An invariant approach to statistical analysis of shapes. New York: Chapman & Hall/CRC; 2001.
[5] Huang C, Styner M, Zhu H. Clustering high-dimensional landmark-based two-dimensional shape data. J Am Stat Assoc 2016;110:946–61.
[6] Kume A, Welling M. Maximum likelihood estimation for the offset-normal shape distributions using EM. J Comput Graph Stat 2010;19:702–23.
[7] Srivastava A, Klassen EP. Functional and shape data analysis. New York: Springer; 2016.
[8] Bookstein FL. Morphometric tools for landmark data: Geometry and biology. Cambridge University Press; 1991.
[9] Kendall DG. Shape manifolds, Procrustean metrics and complex projective spaces. Bull Lond Math Soc 1984;16:81–121.
[10] Cootes T, Taylor C, Cooper D, Graham J. Active shape models-their training and application. Comput Vis Image Underst 1995;61:38–59.
[11] Murray MK, Rice JW. Differential geometry and statistics. Chapman & Hall; 1984.
[12] Amari S, Nagaoka H. Methods of information geometry. Translations of mathematical monographs, vol. 191, Providence: AMS & Oxford University Press; 2000.
[13] Gattone S, Sanctis AD, Russo T, Pulcini D. A shape distance based on the Fisher–Rao metric and its application for shapes clustering. Phisica A 2017;487:93–102.
[14] Gattone S, Sanctis AD, Puechmorel S, Nicol F. On the geodesic distance in shapes K-means clustering. Entropy 2018;20(9):647.
[15] Costa S, Santos S, Strapasson J. Fisher information distance: A geometrical reading. Discrete Appl Math 2015;197:59–69.
[16] Ay N, Jost J. Information geometry. A series of modern surveys in mathematics, Cham: Springer; 2017.
[17] Bookstein FL. Size and shape spaces for landmark data in two dimensions. Statist Sci 1986;1:181–242.

[18] Goodall CR. Procrustes methods in the statistical analysis of shape. J R Stat Soc 1991;53:285–339.
[19] Calinski T, Harabasz J. A dendrite method for cluster analysis. Commun Stat - Theory Methods 1974;3:1–27.
[20] Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a dataset via the gap statistic. J R Stat Soc B 2001;63(2):411–23.
[21] Srivastava A, Klassen E, Joshi SH, Jermyn IH. Shape analysis of elastic curves in Euclidean spaces. IEEE Trans Pattern Anal Mach Intell 2011;33(7):1411–28.