HAL open science

Classifying histograms of medical data using information geometry of beta distributions

Alice Le Brigant, Nicolas Guigui, Sana Rebbah, Stéphane Puechmorel

To cite this version:

Alice Le Brigant, Nicolas Guigui, Sana Rebbah, Stéphane Puechmorel. Classifying histograms of medical data using information geometry of beta distributions. 2020. hal-02671122v1

HAL Id: hal-02671122

https://hal.science/hal-02671122v1

Preprint submitted on 2 Jun 2020 (v1), last revised 24 Aug 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classifying histograms of medical data using information geometry of beta distributions

Alice Le Brigant * **Nicolas Guigui** ** **Sana Rebbah** ***,**** **Stphane Puechmorel** ***

* *Université Paris 1 Panthéon Sorbonne, Paris, France*
** *Université Côte d'Azur, Inria Epione, Sophia Antipolis, France*
*** *Ecole Nationale de l'Aviation Civile, Toulouse, France*
**** *Toulouse NeuroImaging Center, Universit de Toulouse, Inserm, UPS, France*

Abstract: In this paper, we use tools of information geometry to compare, average and classify histograms. Beta distributions are fitted to the histograms and the corresponding Fisher information geometry is used for comparison. We show that this geometry is negatively curved, which guarantees uniqueness of the notion of mean, and makes it suitable to classify histograms through the popular K-means algorithm. We illustrate the use of these geometric tools in supervised and unsupervised classification procedures of two medical data-sets, cardiac shape deformations for the detection of pulmonary hypertension and brain cortical thickness for the diagnosis of Alzheimer's disease.

Keywords: Information geometry, histogram analysis, classification, clustering, medical imaging.

1. INTRODUCTION

The differential geometric approach to probability theory and statistics has met increasing interest in the past years, from the theoretical point of view as well as in applications. In this approach, probability distributions are seen as elements of a differentiable manifold, on which a metric structure is defined through the choice of a Riemannian metric. Two very important ones are the Wasserstein metric, central in optimal transport, and the Fisher information metric (also called Fisher-Rao metric), essential in information geometry. Unlike optimal transport, information geometry is foremost concerned with parametric families of probability distributions, and defines a Riemannian structure on the parameter space using the Fisher information matrix, see Fisher (1922). The Fisher information is the Hessian of the Kullback-Leibler divergence, a popular measure of dissimilarity between probability distributions that does not verify the properties of a distance. In parameter estimation, it can be interpreted as the quantity of information contained in the model about unknown parameters. Rao (1992) showed that it could be used to locally define a scalar product on the space of parameters, interpreted as a Riemannian metric.

An important feature of this metric is that it is invariant under any diffeomorphic change of parameterization. Indeed, it is induced by a more general Riemannian metric on the infinite-dimensional space of smooth, non parametric probability densities, and therefore it is invariant to coordinate change; see e.g. Friedrich (1991). It can be seen as a natural choice of metric as it is the only Riemannian metric that is invariant with respect to transformation by a sufficient statistic, or a diffeomorphic transformation of the support in the non-parametric case (Cencov, 2000; Bauer et al., 2016). Arguably the most famous example of Fisher information geometry of a statistical model is that of the univariate Gaussian model, which is hyperbolic. The geometries of other parametric families such as the multivariate Gaussian model (Atkinson and Mitchell, 1981; Skovgaard, 1984), the family of gamma distributions (Arwini and Dodson, 2008; Rebbah et al., 2019), or more generally location-scale models (Said et al., 2019), among others, have also received a lot of attention.

In this work, we focus on beta distributions, a family of probability measures on $[0, 1]$ used to model random variables defined on a compact interval in a wide variety of applications, e.g. in Bayesian inference as conjugate prior for several discrete probability distributions (ONeill and Roberts, 1999), or to model percentages and proportions in genomic studies (Yang and Fang, 2017). Up to our knowledge, the information geometry of beta distributions has not yet received much attention. In Section 2, we give new results and properties for this geometry by deriving the metric matrix, and the geodesic equations. We compute the sectional curvature and show it has negative sign everywhere. This result is of particular interest to ensure the existence of Frchet means. By numerically solving the geodesic equation, we may compute geodesic distances and means between probability distributions.

⋆ Sponsor and financial support acknowledgment goes here. Paper titles should be written in uppercase and lowercase letters, not all uppercase.

In Section 3, we exemplify the use of this geometry in the analysis of histograms extracted from segmented medical images. Traditional statistical methods are challenged in medical data analysis as data-sets often have very few samples ($\sim 10-100$) whereas observations are very high dimensional ($\sim 10^4-10^6$) and missing data are ubiquitous. Moreover anatomical measurements are often obtained through complex processing pipelines that result in various numbers of features (e.g. mesh representations of organs' surface may have different numbers of vertices), resulting in the lack of a common representation space for the data.

To bypass these issues, it is thus helpful to consider histograms across the whole region of interest and to represent them in the space of beta distributions. The obtained representation has only two dimensions which is useful for visualization and interpretation of the results. We demonstrates that classification in this space achieves similar accuracy in discriminating patients from healthy controls as in the high-dimensional data space.

2. FISHER GEOMETRY OF BETA DISTRIBUTIONS

2.1 The Fisher information metric

Let Θ be an open set of $\mathbb{R}^d$ and let μ be a measure defined on a measurable space $(E, \mathcal{T})$. A parameterized family of distributions on the parameter space Θ is a set of probability measures absolutely continuous with respect to μ, that is:

$$\mathcal{P}_\Theta = \{p(\cdot;\theta)\mu, \theta \in \Theta\}.$$

When the mapping $\theta \in \Theta \mapsto p(x,\theta)$ is differentiable for μ-almost all $x \in E$ and the model satisfies certain regularity conditions, the Fisher information matrix at θ is defined to be:

$$I(\theta) = \left[E\partial_i l(x,\theta)\partial_j l(x,\theta)\right]_{1\le i,j\le d},$$

where $l(x,\theta) = \log p(x,\theta)$. As an open subset of $\mathbb{R}^d$, Θ is a differentiable manifold with $T_\theta\Theta \simeq \mathbb{R}^d$. The Fisher information defines a Riemannian metric on Θ called the Fisher information metric

$$(u,v) \in T_\theta\Theta \mapsto g_\theta(u,v) = u^t I(\theta) v,$$

where u^t denotes the transpose of the vector u. By a slight abuse of language, we will talk of the geometry of $\mathcal{P}_\Theta$, implicitly referring to the metric structure induced on $\mathcal{P}_\Theta$ by (Θ, g).

Here we consider the parametric family of beta distributions, a family of probability measures on $[0,1]$ with density with respect to the Lebesgue measure parameterized by two positive scalars $x, y > 0$

$$p(t;x,y) = \frac{\Gamma(x+y)}{\Gamma(x)\Gamma(y)} t^{x-1}(1-t)^{y-1}, \quad t \in [0,1].$$

We consider the Riemannian manifold composed of the parameter space $\Theta = \mathbb{R}_+^* \times \mathbb{R}_+^*$ and the Fisher information metric g, and by extension denote by beta manifold the pair $(\mathcal{B}, g)$, where $\mathcal{B}$ is the family of beta distributions

$$\mathcal{B} = \{B(x,y) = p(\cdot;x,y)dt, x > 0, y > 0\}.$$

Here dt denotes the Lebesgue measure on $[0,1]$. The beta family is a natural exponential one with log-partition function

$$\varphi(x,y) = \ln\Gamma(x+y) - \ln\Gamma(x) - \ln\Gamma(y). \tag{1}$$

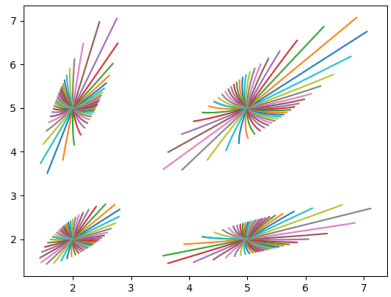

Fig. 1. Geodesic balls in the beta manifold.

and so the Fisher information admits a simple expression:

$$I(x,y) = -\mathrm{Hess}\,\varphi(x,y).$$

Straightforward computations then yield

$$I(x,y) = \begin{bmatrix} \psi'(x) - \psi'(x+y) & -\psi'(x+y) \\ -\psi'(x+y) & \psi'(y) - \psi'(x+y) \end{bmatrix} \tag{2}$$

where ψ denotes the digamma function, i.e.

$$\psi(x) = \frac{d}{dx}\ln\Gamma(x). \tag{3}$$

The matrix form of the Fisher metric g is given by (2), and its infinitesimal length element is

$$ds^2 = \psi'(x)dx^2 + \psi'(y)dy^2 - \psi'(x+y)(dx+dy)^2.$$

2.2 Geodesics and geodesic distance

The distance between two beta distributions is defined as the geodesic distance associated to the Fisher metric in the parameter space

$$d_F(B(x,y), B(x',y')) = \inf_\gamma \int_0^1 \sqrt{g(\dot\gamma(t), \dot\gamma(t))}dt, \tag{4}$$

where the infimum is taken over all paths $\gamma : [0,1] \to \Theta$ such that $\gamma(0) = (x,y)$ and $\gamma(1) = (x',y')$. To compute this distance in practice, one needs to solve the geodesic equation.

Proposition 1. The geodesics $t \mapsto (x(t), y(t))$ of the beta parameter space Θ are solutions of

$$\begin{aligned} \ddot x + a(x,y)\dot x^2 + b(x,y)\dot x\dot y + c(x,y)\dot y^2 = 0, \\ \ddot y + a(y,x)\dot y^2 + b(y,x)\dot x\dot y + c(y,x)\dot x^2 = 0, \end{aligned} \tag{5}$$

where

$$a(x,y) = \frac{\psi''(x)\psi'(y) - \psi''(x)\psi'(x+y) - \psi'(y)\psi''(x+y)}{2d(x,y)},$$

$$b(x,y) = -\frac{\psi''(x+y)\psi'(y)}{d(x,y)},$$

$$c(x,y) = \frac{\psi''(y)\psi'(x+y) - \psi'(y)\psi''(x+y)}{2d(x,y)},$$

$$d(x,y) = \psi'(x)\psi'(y) - \psi'(x+y)(\psi'(x) + \psi'(y)).$$

No closed form for the geodesics is known, but they can be computed numerically by solving (5), see Section. 2.4. Nonetheless we can notice that, due to the symmetry of the metric with respect to parameters x and y, the line of equation $x = y$ is a geodesic, where the parameterization is fixed by the unique geodesic equation given by (5).

2.3 Negative curvature and Frchet mean

Extension of basic statistical objects such as the mean, the median or the variance to the setting of Riemannian manifolds is now well known; see e.g. Pennec (2006). A popular choice to define the notion of mean in a Riemannian manifold is the Fréchet mean, also called intrinsic mean, which is defined as the minimizer of the sum of the squared distances to the points that we want to average. In our setting, the intrinsic mean of a set of $\{B_i, i = 1, \ldots, n\}$ of beta distributions is given by

$$\bar{B} = \operatorname*{argmin}_{B \in \mathcal{B}} \sum_{i=1}^{n} d_F(B, B_i)^2.$$

The Fréchet mean is in general not unique and refers to a set. Uniqueness holds however when the curvature of the Riemannian manifold is negative (Karcher, 1977), which is the case here.

Theorem 1. The beta manifold $(\mathcal{B}, d_F)$ has negative sectional curvature. Therefore, any set of beta distributions admits a unique Frchet mean for the Fisher information geometry.

The proof of this theorem, given in the appendix, relies on certain regularity conditions of the polygamma functions, and in particuler on the sub-additivity of the ratio ψ'/ψ'' which was proven recently by Yang (2017).

The uniqueness of the Frchet mean for this geometry is an important argument in favor of its use for supervised and unsupervised classification of histograms. Indeed, the popular K-means algorithm, also called Lloyd's algorithm (Lloyd, 1982), which seeks to minimize intra-cluster variance, relies on the computation of the mean of the clusters. It can be easily extended to non-linear manifolds by replacing the Euclidean distance and mean by the Riemannian distance and Frchet mean.

2.4 Implementation

Geodesics (see Fig. 1) are computed through the exponential and logarithm maps of the Riemannian manifold, which can be implemented by solving respectively the initial and boundary value problems associated to the geodesic equation (5). This allows to approximate the distance, by averaging the norm of the velocity of the discretized geodesic between two points. The Frchet mean can be found in practice using a gradient descent algorithm commonly referred to as the Karcher flow. It simply consists in applying the following update on the current estimate $\hat{B}$ of the mean at each iteration:

$$\hat{B} \leftarrow \exp_{\hat{B}}\left(\frac{\tau}{n}\sum_{i=1}^{n}\log_{\hat{B}}(B_i)\right)$$

for some step size $\tau > 0$. Here exp and log denote the Riemannian exponential and logarithm maps. Our implementation is available in the open-source Python package geomstats `http://geomstats.ai`.

3. APPLICATIONS TO MEDICAL DATA ANALYSIS

In this section with exemplify the theoretical results of the previous section on histograms computed from medical data in two different contexts.

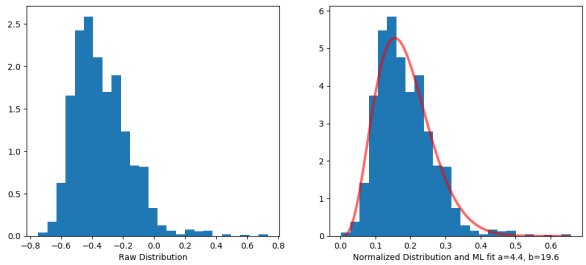

Fig. 2. Example of the normalization and ML representation of a subject's AS distributions in the space of beta distributions

3.1 Data-sets

Cardiac Shape Deformations Firstly, we use a data-base of 204 shapes of right ventricles (RV) of the heart extracted from 3d-echocardiographic sequences (Moceri et al., 2017). 3d-meshes were extracted from the sequences with semi-automatic software (TomTec 4D RV-Function 2.0) and post-processed to compute markers of deformation during the cardiac cycle. We focus on the systole, that is the period during which the RV contracts to eject blood through the pulmonary valve. As a marker of deformation, we use the Area Strain (AS), that represents the local stretching of the RV surface. This feature is of growing interest among clinicians to assess cardiac function. It was shown to be a predictor of survival in pulmonary hypertension (Moceri et al., 2017).

For each subject, the RV at time t is represented by a 3d-mesh of 822 vertices and 1587 triangular cells. Note that all vertices and cells correspond to one-another across subjects, but this may not be the case in most applications. This allows us to compare statistics computed in the original data space and in the space of beta distributions. We compute the relative area change of each cell k between the end of diastole (t_0) and the end of systole (t_1):

$$AS_k = \frac{a_k^{t_1} - a_k^{t_0}}{a_k^{t_0}}, \tag{6}$$

where a_k^t is the area of the triangle k at time t for a given mesh. This results in a distribution of values greater than -1. Due to the tissue's nature, the values should in fact be bounded by constants $p < q$. However, the acquisition noise causes some values to lie outside $[p, q]$ although physically infeasible. We thus normalize the data to lie in the interval $[0, 1]$ by applying the transformation

$$x \mapsto \frac{\min(\max(x, p), q) - p}{q - p}.$$

Finally, we map each patient's histogram to the space of beta distributions by estimating the parameters x, y by maximum likelihood (ML). Fig. 2 represents the data for a given patient with the estimated probability density function and Fig. 3 the obtained representation of the population's histograms in the manifold of beta distributions.

Cortical Thickness maps The measure of brain atrophy is a crucial tool in the analysis of neurodegenerative diseases. In particular, Cortical Thickness (CTh) measured by structural Medical Resonance Imaging (MRI)

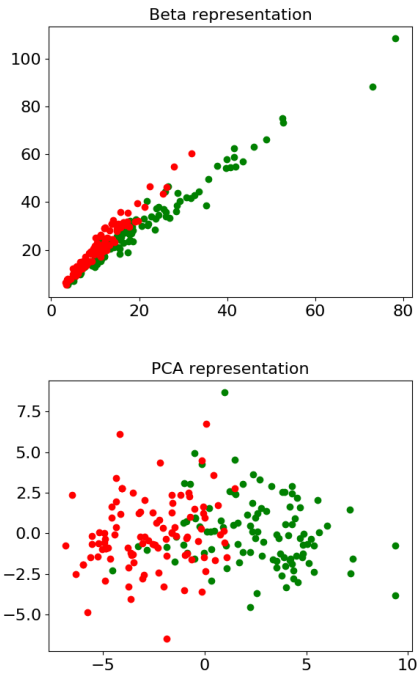

Fig. 3. AS data represented in the parameter space of beta distributions (up) and in the space of two first principal components (down).

has been proposed as a biomarker for the early diagnosis of Alzheimers disease (Pini et al., 2016; Busovaca et al., 2016) and for the prediction of conversion from Mild Cognitive Impairment to Alzheimer's Disease (Wei et al., 2016; Sun et al., 2017). In many of these studies, central tendency measures such as the mean or the median are used to represent the biomarker, either in the whole cortex, in regions of interest or in the voxels, leading to an important loss of information. More recent ones used histogram analysis (Giulietti et al., 2018; Ruiz et al., 2018), and in (Rebbah et al., 2019), the authors use the Fisher geometry of generalized gamma distributions to perform this analysis. However, cortical thickness is a bounded quantity, and therefore beta distributions seem to be more natural candidates to model their distribution. Moreover, the negative curvature of the beta geometry makes it a suitable candidate in clustering procedures based on computation of cluster means, such as the K-mean algorithm used here.

The data used in this application were extracted from MR scans selected from the Alzheimers Disease Neuroimaging Initiative (ADNI) database [1] . Indeed, the initial subjects were not age- and sex-matched and our procedure consisted in randomly selecting subjects. In addition, some of the subjects were excluded because of a low diagnosis reliability (according to ADNI criteria) and others because of unsuccessful CTh measurement due to poor image quality. The resulting population is composed of 143 subjects: 71 healthy controls subjects and 72 Alzheimers disease patients.

CTh was measured from 3D T1-weighted MR images using the Matlab toolbox CorThiZon (Querbes et al., 2009).

[1] `http://adni.loni.usc.edu/about/`

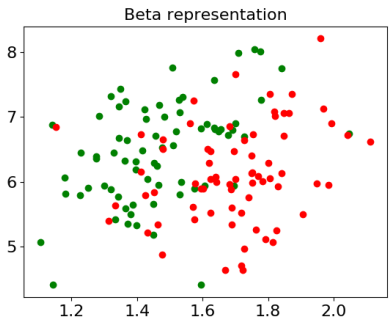

Fig. 4. CTh data represented in the parameter space of beta distributions.

There is no unique definition of CTh. In (MacDonald et al., 2000) it is obtained as the distance between uniquely associated pairs of points on bounding surfaces. Uncoupled methods do not rely on a given set of associated landmarks but create them using techniques coming from shape matching (Tagare, 1997). Finally, a quite different category of algorithms makes use of an elliptic partial differential equation to obtain a field of deformation between the boundaries (Yezzi and Prince, 2003). The CorThiZon toolbox, developed at Inserm, implements a Laplace equation based method that falls within this last category.

For each subject, the data consist in measurements of CTh along the whole cortical ribbon. The overall procedure is the one described in (Querbes et al., 2009), with a normalized voxel size of 1mm along all directions. Due to the variability of the size and shape of the brain among the population of study, we obtain samples of unequal length. This results in the lack of a common representation space for the data, a problem that is solved by the histogram-based approach that we propose here. Indeed, just as for the AS data, we compute for each patient the histogram of CTh measurements, from which we estimate the parameters of a beta distribution by ML. Histograms are normalized beforehand with respect to the maximal CTh value among the population.

3.2 Methodology

In both cases, the subjects are divided in two classes: diseased and controls. In order to assess whether the proposed representation and geometry are relevant, we perform supervised and unsupervised classification in the Riemannian manifold of beta distributions. We compare the results both when using the Riemannian metric and the Euclidean metric on the parameter space $(R_*^+)^2$. We chose to use the K-nearest- neighbor (KNN) and K-means algorithms as they only require computation of distances on non-linear manifolds. Both supervised (SKM) and unsupervised (UKM) versions of K-means are compared. Finally, in the case of the AS data, as we have cell-to-cell correspondence, we can run the algorithms in the original data space R1587 to assess the loss of information when considering only two-parameter distributions instead of the whole histograms. We also compare our method with a standard dimensionality reduction performed by principal component analysis (PCA). This cannot be done for the CTh data since the mapping between values and brain

regions differs across patients. All models are trained and tested in a 5-fold cross-validation fashion, and we assess the classification accuracy in all cases. We used the scikit-learn package (Pedregosa et al., 2011) to perform these experiments.

3.3 Results

Supervised classification Tables 1 and 2 show the mean classification accuracy over 5-fold cross validation for each setting: (1) the space of parameters of the beta distributions fitted to the data-point histograms and, when applicable, (2) the Euclidean space of the original data, and (3) the space spanned by the first two principal components obtained by PCA. The number of neighbors for the KNN algorithm is chosen to be $K = 7$. This choice is compared to other values through 5-fold cross validation, see Fig. 5 and 6.

Table 1. Mean (and standard deviation) of the classification accuracy for the AS data on 5-fold cross-validation

	Original	PCA	Beta	
	Euclidean	Euclidean	Euclidean	Riemannian
KNN	0.81 (0.05)	0.67 (0.32)	0.77 (0.09)	**0.83** (0.05)
SKM	**0.85** (0.03)	0.72 (0.28)	0.66 (0.06)	0.81 (0.06)

Table 2. Mean (and standard deviation) of the classification accuracy for the CTh data on 5-fold cross-validation

	Beta	
	Euclidean	Riemannian
KNN	0.77 (0.05)	**0.79** (0.04)
SKM	0.66 (0.10)	**0.80** (0.05)

These results show that in the space of beta parameters, the Fisher information geometry performs significantly better than Euclidean geometry. Moreover, this performance is comparable to that of classification in the original data space, despite the considerable reduction of dimension (from 1587 to 2). This is illustrated in Fig. 5, where both methods perform alternatively better as the number of neighbors changes.

In comparison, the two-dimensional representation space provided by PCA performs significantly worst on average. This can be explained by the fact that the performance of dimensionality reduction of PCA depends on the choice of training set, and certain train/test splittings lead to poor performance. This is illustrated in Fig. 7: in one of the cross-validation splittings, the test set for each class in the PCA representation is closer to the center of the other class computed from the training test. This has the effect of lowering the average accuracy over the whole cross-validation process. This problem does not occur in the two-dimensional beta representation, as the mapping from one data-point to a pair of parameters (x, y) does not depend on the other points present in the data-set.

Unsupervised classification The clustering performance of the K-means algorithm is shown for the same settings in Tables 3 and 4. These results confirm that the Fisher information geometry in the space of beta parameters is more suited to compare histograms than Euclidean geometry. Unlike in the supervised setting, the two-dimensional representation given by PCA performs well, since in this case the whole data-set was used to perform the dimensionality reduction.

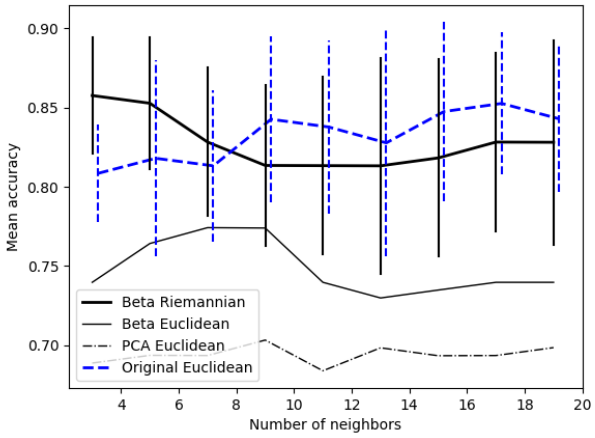

Fig. 5. Mean accuracy of KNN classification on the AS data over 5-fold cross validation. Vertical segments indicate standard deviation for the original Euclidean and the beta Riemannian representations.

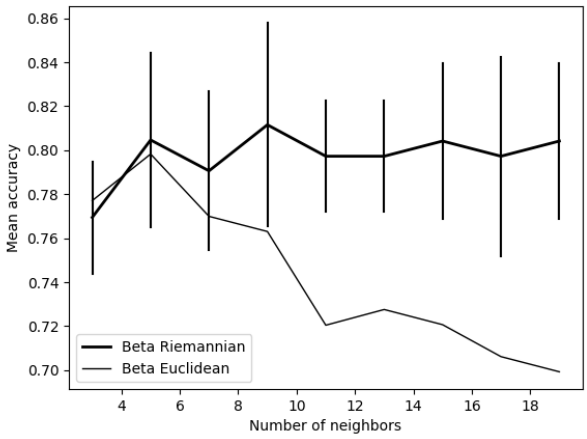

Fig. 6. Mean accuracy of KNN classification on the CTh data over 5-fold cross validation. Vertical segments indicate standard deviation for the beta Riemannian representation.

Table 3. Clustering accuracy for the AS data

	Beta		Original	PCA
	Euclidean	Riemannian	Euclidean	Euclidean
UKM	0.60	0.80	**0.84**	**0.84**

Table 4. Clustering accuracy for the CTh data

	Beta	
	Euclidean	Riemannian
UKM	0.61	**0.82**

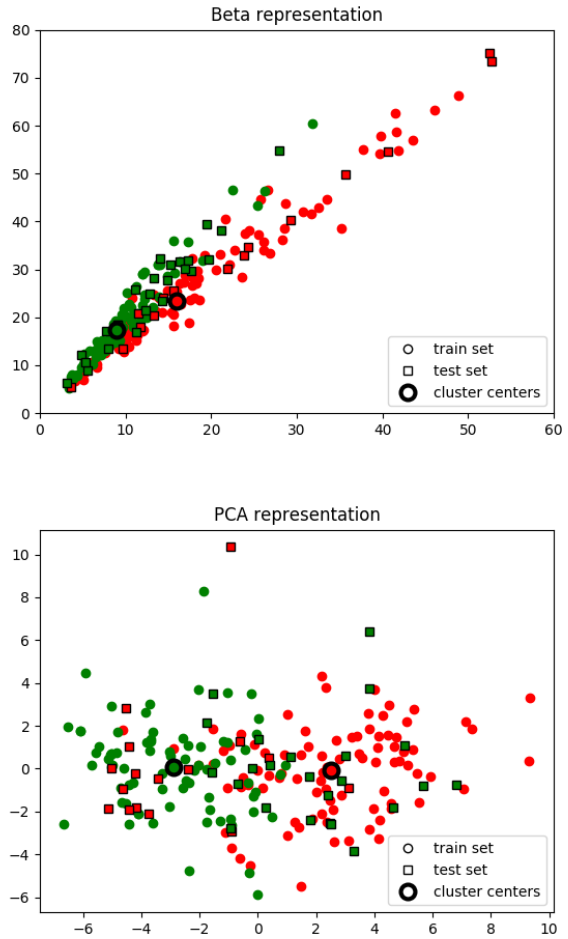

Fig. 7. Supervised K-means on the AS strain data in beta and PCA representations.

4. CONCLUSION

In this work, we have studied the geometry of the beta manifold and shown in particular that it was negatively curved. We have illustrated its use as a non linear representation space for histograms of medical data, which presents the advantage of being low-dimensional and can easily be visualized. This common representation is also particularly helpful when the number of measurements vary from one subject to another. The examples studied here tend to show that the Fisher information metric is a more adapted choice of metric on this space than the Euclidean one to compare and classify histograms. Finally, the dimensionality reduction performed in the process does not result in a significant loss in performance. Moreover, unlike a standard method like PCA, the quality of the dimensionality reduction does not rely on the choice and size of the data-set: each data-point is mapped to an element of the beta manifold independently of the others. This makes the beta representation particularly interesting for small data-sets. Comparison with other geometries, such as that of the gamma and generalized gamma manifolds, and exploration of other supervised and unsupervised classification algorithms will be the object of future work.

ACKNOWLEDGEMENTS

The authors would like to warmly thank Pamela Moceri (CHU Nice) and Nicolas Duchateau (Université Lyon 1) for acquiring and curating the heart data. N. Guigui has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement G-Statistics No 786854). This research has been conducted within the FP2M federation (CNRS FR 2036)

REFERENCES

Arwini, K. and Dodson, C.T. (2008). *Information geometry: near randomness and near independence*. Springer Science & Business Media.

Atkinson, C. and Mitchell, A.F. (1981). Rao's distance measure. *Sankhyā: The Indian Journal of Statistics, Series A*, 345–365.

Bauer, M., Bruveris, M., and Michor, P.W. (2016). Uniqueness of the Fisher–Rao metric on the space of smooth densities. *Bulletin of the London Mathematical Society*, 48(3), 499–506.

Busovaca, E., Zimmerman, M.E., Meier, I.B., Griffith, E.Y., Grieve, S.M., Korgaonkar, M.S., Williams, L.M., and Brickman, A.M. (2016). Is the Alzheimers disease cortical thickness signature a biological marker for memory? *Brain imaging and behavior*, 10(2), 517–523.

Cencov, N.N. (2000). *Statistical decision rules and optimal inference*. 53. American Mathematical Soc.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604), 309–368.

Friedrich, T. (1991). Die Fisher-information und symplektische strukturen. *Mathematische Nachrichten*, 153(1), 273–296.

Giulietti, G., Torso, M., Serra, L., Spanò, B., Marra, C., Caltagirone, C., Cercignani, M., Bozzali, M., and (ADNI), A.D.N.I. (2018). Whole brain white matter histogram analysis of diffusion tensor imaging data detects microstructural damage in mild cognitive impairment and Alzheimer's disease patients. *Journal of Magnetic Resonance Imaging*, 48(3), 767–779.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5), 509–541.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129–137.

MacDonald, D., Kabani, N., Avis, D., and Evans, A.C. (2000). Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage*, 12(3), 340 – 356.

Moceri, P., Duchateau, N., Baudouy, D., Schouver, E.D., Leroy, S., Squara, F., Ferrari, E., and Sermesant, M. (2017). Three-dimensional right-ventricular regional deformation and survival in pulmonary hypertension. *European Heart Journal - Cardiovascular Imaging*, 19(4), 450–458. doi:10.1093/ehjci/jex163.

ONeill, P.D. and Roberts, G.O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal*

of the Royal Statistical Society: Series A (Statistics in Society), 162(1), 121–129.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1), 127.

Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., and Frisoni, G.B. (2016). Brain atrophy in Alzheimers disease and aging. *Ageing research reviews*, 30, 25–48.

Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Dmonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P., and Initiative, T.A.D.N. (2009). Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain*, 132(8), 2036–2047. doi:10.1093/brain/awp105.

Rao, C.R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in statistics*, 235–247. Springer.

Rebbah, S., Nicol, F., and Puechmorel, S. (2019). The geometry of the generalized gamma manifold and an application to medical imaging. *Mathematics*, 7(8), 674.

Ruiz, E., Ramírez, J., Górriz, J.M., Casillas, J., Initiative, A.D.N., et al. (2018). Alzheimers disease computer-aided diagnosis: Histogram-based analysis of regional MRI volumes for feature selection and classification. *Journal of Alzheimer's Disease*, 65(3), 819–842.

Said, S., Bombrun, L., and Berthoumieu, Y. (2019). Warped Riemannian metrics for location-scale models. In *Geometric Structures of Information*, 251–296. Springer.

Skovgaard, L.T. (1984). A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 211–223.

Sun, Z., van de Giessen, M., Lelieveldt, B.P.F., and Staring, M. (2017). Detection of conversion from Mild Cognitive Impairment to Alzheimer's disease using longitudinal brain MRI. *Frontiers in Neuroinformatics*, 11. doi:10.3389/fninf.2017.00016. URL `https://doi.org/10.3389/fninf.2017.00016`.

Tagare, H.D. (1997). Deformable 2-D template matching using orthogonal curves. *IEEE Transactions on Medical Imaging*, 16(1), 108–117.

Totaro, B. (2004). The curvature of a Hessian metric. *International Journal of Mathematics*, 15(04), 369–391.

Wei, R., Li, C., Fogelson, N., and Li, L. (2016). Prediction of conversion from Mild Cognitive Impairment to Alzheimer's disease using MRI and structural network features. *Frontiers in Aging Neuroscience*, 8. doi: 10.3389/fnagi.2016.00076. URL `https://doi.org/10.3389/fnagi.2016.00076`.

Yang, S. and Fang, Z. (2017). Beta approximation of ratio distribution and its application to next generation sequencing read counts. *Journal of applied statistics*, 44(1), 57–70.

Yang, Z.H. (2017). Some properties of the divided difference of psi and polygamma functions. *Journal of Mathematical Analysis and Applications*, 455(1), 761 – 777. doi:https://doi.org/10.1016/j.jmaa.2017.05.081.

Yezzi, A. and Prince, J. (2003). An Eulerian PDE approach for computing tissue thickness. *IEEE Transactions on Medical Imaging*, 22(10), 1332–1339. doi: 10.1109/TMI.2003.817775.

Appendix A. PROOF OF PROPOSITION 1

The geodesic equations are given by

$$\ddot{x} + \Gamma^x_{xx}\dot{x}^2 + 2\Gamma^x_{xy}\dot{x}\dot{y} + \Gamma^x_{yy}\dot{y}^2 = 0$$
$$\ddot{y} + \Gamma^y_{xx}\dot{x}^2 + 2\Gamma^y_{xy}\dot{x}\dot{y} + \Gamma^y_{yy}\dot{y}^2 = 0$$

where the Γ^k_{ij}'s denote the Christoffel symbols of the second kind. These can be obtained from the Christoffel symbols of the first kind Γ_{ijk} and the coefficients g^{ij} of the inverse of the metric matrix

$$\Gamma^k_{ij} = \Gamma_{ijl}g^{kl}.$$

Here we have used the Einstein summation convention. Since the Fisher metric is a Hessian metric (see Totaro (2004)), the Christoffel symbols of the first kind can be obtained as

$$\Gamma_{ijk} = \frac{1}{2}\varphi_{ijk},$$

where φ is the log-partition function (1) and lower indices denote partial derivatives. Straightforward computation yields the desired equations.

Appendix B. PROOF OF THEOREM 1

The sectional curvature of a Hessian metric is given by

$$K = \frac{1}{4(\det g)^2}R_{xyxy}$$

where

$$\begin{aligned} R_{xyxy} = &-\varphi_{yy}(\varphi_{xxx}\varphi_{xyy} - \varphi^2_{xxy}) \\ &+ \varphi_{xy}(\varphi_{xxx}\varphi_{yyy} - \varphi_{xxy}\varphi_{xyy}) \\ &+ \varphi_{xx}(\varphi_{xxy}\varphi_{yyy} - \varphi^2_{xyy}). \end{aligned}$$

Here lower indices denote partial derivatives with respect to the corresponding variables, and φ is the log-partition function (1). Their computation yields

$$\varphi_{xxx} = \psi''(x) - \psi''(x+y),$$
$$\varphi_{yyy} = \psi''(y) - \psi''(x+y),$$
$$\varphi_{xxy} = \varphi_{xyy} = -\psi''(x+y),$$

and the determinant of the metric is given by

$$\det g = d(x,y) = \psi'(x)\psi'(y) - \psi'(x+y)(\psi'(x) + \psi'(y)).$$

This gives

$$\begin{aligned} K(x,y) = &(\psi''(x+y)(\psi'(x)\psi''(y) + \psi''(x)\psi'(y)) \\ &-\psi'(x+y)\psi''(x)\psi''(y))/(4d(x,y)^2). \end{aligned}$$

Factorizing the numerator by $\psi''(x)\psi''(y)\psi''(x+y)$ yields

$$K(x,y) = \frac{\psi''(x)\psi''(y)\psi''(x+y)}{4\,d(x,y)^2}(F(x) + F(y) - F(x+y)),$$

where $F = \psi'/\psi''$. Since ψ'' is negative, the first factor is negative. Moreover, it has been shown in (Yang, 2017, Corollary 4) that the function F is sub-additive. This means that the second factor is positive, yielding the desired result.