



HAL
open science

New Riemannian Priors on the Univariate Normal Model

Salem Said, Lionel Bombrun, Yannick Berthoumieu

► **To cite this version:**

Salem Said, Lionel Bombrun, Yannick Berthoumieu. New Riemannian Priors on the Univariate Normal Model. Entropy, 2014, 16, pp.4015 - 4031. 10.3390/e16074015 . hal-01084330

HAL Id: hal-01084330

<https://hal.science/hal-01084330>

Submitted on 19 Nov 2014

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.

Article

New Riemannian Priors on the Univariate Normal Model

Salem Said *, Lionel Bombrun and Yannick Berthoumieu

Groupe Signal et Image, CNRS Laboratoire IMS, Institut Polytechnique de Bordeaux, Université de Bordeaux, UMR 5218, Talence, 33405, France; E-Mails: lionel.bombrun@u-bordeaux.fr (L.B.); yannick.berthoumieu@u-bordeaux.fr (Y.B.)

* Author to whom correspondence should be addressed; E-Mail: salem.said@u-bordeaux.fr; Tel.: +33-(0)5-4000-6185.

Received: 17 April 2014; in revised form: 23 June 2014 / Accepted: 9 July 2014 /

Published: 17 July 2014

Abstract: The current paper introduces new prior distributions on the univariate normal model, with the aim of applying them to the classification of univariate normal populations. These new prior distributions are entirely based on the Riemannian geometry of the univariate normal model, so that they can be thought of as “Riemannian priors”. Precisely, if $\{p_\theta; \theta \in \Theta\}$ is any parametrization of the univariate normal model, the paper considers prior distributions $G(\bar{\theta}, \gamma)$ with hyperparameters $\bar{\theta} \in \Theta$ and $\gamma > 0$, whose density with respect to Riemannian volume is proportional to $\exp(-d^2(\theta, \bar{\theta})/2\gamma^2)$, where $d^2(\theta, \bar{\theta})$ is the square of Rao’s Riemannian distance. The distributions $G(\bar{\theta}, \gamma)$ are termed Gaussian distributions on the univariate normal model. The motivation for considering a distribution $G(\bar{\theta}, \gamma)$ is that this distribution gives a geometric representation of a class or cluster of univariate normal populations. Indeed, $G(\bar{\theta}, \gamma)$ has a unique mode $\bar{\theta}$ (precisely, $\bar{\theta}$ is the unique Riemannian center of mass of $G(\bar{\theta}, \gamma)$, as shown in the paper), and its dispersion away from $\bar{\theta}$ is given by γ . Therefore, one thinks of members of the class represented by $G(\bar{\theta}, \gamma)$ as being centered around $\bar{\theta}$ and lying within a typical distance determined by γ . The paper defines rigorously the Gaussian distributions $G(\bar{\theta}, \gamma)$ and describes an algorithm for computing maximum likelihood estimates of their hyperparameters. Based on this algorithm and on the Laplace approximation, it describes how the distributions $G(\bar{\theta}, \gamma)$ can be used as prior distributions for Bayesian classification of large univariate normal populations. In a concrete application to texture image classification, it is shown that this leads to an improvement in performance over the use of conjugate priors.

Keywords: Fisher information; Riemannian metric; prior distribution; univariate normal distribution; image classification

1. Introduction

In this paper, a new class of prior distributions is introduced on the univariate normal model. The new prior distributions, which will be called Gaussian distributions, are based on the Riemannian geometry of the univariate normal model. The paper introduces these new distributions, uncovers some of their fundamental properties and applies them to the problem of the classification of univariate normal populations. It shows that, in the context of a real-life application to texture image classification, the use of these new prior distributions leads to improved performance in comparison with the use of more standard conjugate priors.

To motivate the introduction of the new prior distributions, considered in the following, recall some general facts on the Riemannian geometry of parametric models.

In information geometry [1], it is well known that a parametric model $\{p_\theta; \theta \in \Theta\}$, where $\Theta \subset R^p$, can be equipped with a Riemannian geometry, determined by Fisher's information matrix, say $I(\theta)$. Indeed, assuming $I(\theta)$ is strictly positive definite, for each $\theta \in \Theta$, a Riemannian metric on Θ is defined by:

$$ds^2(\theta) = \sum_{i,j=1}^p I_{ij}(\theta) d\theta^i d\theta^j \quad (1)$$

The fact that the length element Equation (1) is invariant to any change of parametrization was realized by Rao [2], who was the first to propose the application of Riemannian geometry in statistics.

Once the Riemannian metric Equation (1) is introduced, the whole machinery of Riemannian geometry becomes available for application to statistical problems relevant to the parametric model $\{p_\theta; \theta \in \Theta\}$. This includes the notion of Riemannian distance between two distributions, p_θ and $p_{\theta'}$, which is known as Rao's distance, say $d(\theta, \theta')$, the notion of Riemannian volume, which is exactly the same as Jeffreys prior [3], and the notion of Riemannian gradient, which can be used in numerical optimization and coincides with the so-called natural gradient of Amari [4].

It is quite natural to apply Rao's distance to the problem of classifying populations that belong to the parametric model $\{p_\theta; \theta \in \Theta\}$. In the case where this parametric model is the univariate normal model, this approach to classification is implemented in [5]. For more general parametric models, beyond the univariate normal model, similar applications of Rao's distance to problems of image segmentation and statistical tests can be found in [6–8].

The idea of [5] is quite elegant. In general, it requires that some classes $\{\mathcal{S}_L; L = 1, \dots, C\}$, (based on a learning sequence) have been identified with "centers" $\bar{\theta}_L \in \Theta$. Then, in order to assign a test population, given by the parameter θ_t , to a class L^* , it is proposed to choose L^* , which minimizes Rao's distance $d^2(\theta_t, \bar{\theta}_L)$, over $L = 1, \dots, C$. In the specific context of the classification of univariate normal populations [5], this leads to the introduction of hyperbolic Voronoi diagrams.

The present paper is also concerned with the case where the parametric model $\{p_\theta; \theta \in \Theta\}$ is a univariate normal model. It starts from the idea that a class \mathcal{S}_L should be identified not only with a center $\bar{\theta}_L$, as in [5], but also with a kind of “variance”, say γ^2 , which will be called a dispersion parameter. Accordingly, assigning a test population given by the parameter θ_t to a class L should be based on a tradeoff between the square of Rao’s distance $d^2(\theta_t, \bar{\theta}_L)$ and the dispersion parameter γ^2 .

Of course, this idea has a strong Bayesian flavor. It proposes to give more “confidence” to classes that have a smaller dispersion parameter. Thus, in order to implement it, in a concrete way, the paper starts by introducing prior distributions on the univariate normal model, which it calls Gaussian distributions. By definition, a Gaussian distribution $G(\bar{\theta}, \gamma^2)$ has a probability density function, with respect to Riemannian volume, given by:

$$p(\theta|\bar{\theta}, \gamma) \propto \exp\left(\frac{-d^2(\theta, \bar{\theta})}{2\gamma^2}\right) \quad (2)$$

Given this definition of a Gaussian distribution (which is developed in a detailed way, in Section 3), classification of univariate normal populations can be carried out by associating to each class \mathcal{S}_L of univariate normal populations a Gaussian distribution $G(\bar{\theta}_L, \gamma_L^2)$ and by assigning any test population with parameter θ_t to the class L^* , which maximizes the likelihood $p(\theta_t|\bar{\theta}_L, \gamma_L)$, over $L = 1, \dots, C$.

The present paper develops in a rigorous way the general approach to the classification of univariate normal populations, which has just been described. It proceeds as follows.

Section 2, which is basically self-contained, provides the concepts, regarding the Riemannian geometry of the univariate normal model, which will be used throughout the paper.

Section 3 introduces Gaussian distributions on the univariate normal model and uncovers some of their general properties. In particular, Section 3.2 of this section gives a Riemannian gradient descent algorithm for computing maximum likelihood estimates of the parameters $\bar{\theta}$ and γ of a Gaussian distribution.

Section 4 states the general approach to classification of univariate normal populations proposed in this paper. It deals with two problems: (i) given a class \mathcal{S} of univariate normal populations S_i , how to fit a Gaussian distribution $G(\bar{z}, \gamma)$ to this class; and (ii) given a test univariate normal population S_t and a set of classes $\{\mathcal{S}_L, L = 1, \dots, C\}$, how to assign S_t to a suitable class \mathcal{S}_{L^*} .

In the present paper, the chosen approach for resolving these two problems is marginalized likelihood estimation, in the asymptotic framework where each univariate normal population contains a large number of data points. In this asymptotic framework, the Laplace approximation plays a major role [9]. In particular, it reduces the first problem, of fitting a Gaussian distribution to a class of univariate normal populations, to the problem of maximum likelihood estimation, covered in Section 3.2.

The final result of Section 4 is the decision rule Equation (37). This generalizes the one developed in [5] and already explained above, by taking into account the dispersion parameter γ , in addition to the center $\bar{\theta}$, for each class.

In Section 5, the formalism of Section 4 is applied to texture image classification, using the VisTeX image database [10]. This database is used to compare the performance obtained using Gaussian distributions, as in Section 4, to that obtained using conjugate prior distributions. It is shown that Gaussian distributions, proposed in the current paper, lead to a significant improvement in performance.

Before going on, it should be noted that probability density functions of the form (2), on general Riemannian manifolds, were considered by Pennec in [11]. However, they were not specifically used as prior distributions, but rather as a representation of uncertainty in medical image analysis and directional or shape statistics.

2. Riemannian Geometry of the Univariate Normal Model

The current section presents in a self-contained way the results on the Riemannian geometry of the univariate normal model, which are required for the remainder of the paper. Section 2.1 recalls the fact that the univariate normal model can be reparametrized, so that its Riemannian geometry is essentially the same as that of the Poincaré upper half plane. Section 2.2 uses this fact to give analytic formulas for distance, geodesics and integration on the univariate normal model. Finally, Section 2.3 presents, in general form, the Riemannian gradient descent algorithm.

2.1. Derivation of the Fisher Metric

This paper considers the Riemannian geometry of the univariate normal model, as based on the Fisher metric (1). To be precise, the univariate normal model has a two-dimensional parameter space $\Theta = \{\theta = (\mu, \sigma) | \mu \in R, \sigma > 0\}$, and is given by:

$$p_{\theta}(x) = |2\pi\sigma^2|^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3)$$

where each p_{θ} is a probability density function with respect to the Lebesgue measure on R . The Fisher information matrix, obtained from Equation (3), is the following:

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

As in [12], this expression can be made more symmetric by introducing the parametrization $z = (x, y)$, where $x = \mu/\sqrt{2}$ and $y = \sigma$. This yields the Fisher information matrix:

$$I(z) = 2 \times \begin{pmatrix} \frac{1}{y^2} & 0 \\ 0 & \frac{1}{y^2} \end{pmatrix}$$

It is suitable to drop the factor two in this expression and introduce the following Riemannian metric for the univariate normal model,

$$ds^2(z) = \frac{dx^2 + dy^2}{y^2} \quad (4)$$

This is essentially the same as the Fisher metric (up to the factor two) and will be considered throughout the following. The resulting Rao's distance and Riemannian geometry are given in the following paragraph.

2.2. Distance, Geodesics and Volume

The Riemannian metric (4), obtained in the last paragraph, happens to be a very well-known object in differential geometry. Precisely, the parameter space $H = \{z = (x, y) | y > 0\}$ equipped with the metric (4) is known as the *Poincaré upper half plane* and is a basic model of a two-dimensional hyperbolic space [13].

Rao’s distance between two points $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ in H can be expressed as follows (for results in the present paragraph, see [13], or any suitable reference on hyperbolic geometry),

$$d(z_1, z_2) = \operatorname{acosh} \left(1 + \frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{2y_1y_2} \right) \tag{5}$$

where acosh denotes the inverse hyperbolic cosine.

Starting from z_1 , in any given direction, it is possible to draw a unique geodesic ray $\gamma : R_+ \rightarrow H$. This is a curve having the property that $\gamma(0) = z_1$ and, for any $t \in R_+$, if $\gamma(t) = z_2$ then $d(z_1, z_2) = t$. In other words, the length of γ between z_1 and z_2 is equal to the distance between z_1 and z_2 .

The equation of a geodesic ray starting from $z \in H$ is conveniently written down in complex notation (that is, by treating points of H as complex numbers). To begin, consider the case of $z = i$ (which stands for $x = 0$ and $y = 1$). The geodesic in the direction making an angle ψ with the y -axis is the curve,

$$\gamma_i(t) = \frac{e^{t/2} \cos(\psi/2) i - e^{-t/2} \sin(\psi/2)}{e^{t/2} \sin(\psi/2) i + e^{-t/2} \cos(\psi/2)} \tag{6}$$

In particular $\psi = 0$ gives $\gamma_i(t) = e^t i$ and $\psi = \pi$ gives $\gamma_i(t) = e^{-t} i$. If ψ is not a multiple of π , $\gamma_i(t)$ traces out a portion of a circle, which is parallel to the y -axis, in the limit $t \rightarrow \infty$. For a general starting point z , the geodesic ray in the direction making an angle ψ with the y -axis can be written:

$$\gamma_z(t, \psi) = x + y\gamma_i(t/y, \psi) \tag{7}$$

where $z = (x, y)$ and $\gamma_i(t, \psi)$ is given by Equation (6). A more detailed treatment of Rao’s distance (5) and of geodesics in the Poincaré upper half plane, along with applications in image clustering, can be found in [5].

The Riemannian volume (or area, since H is of dimension 2) element corresponding to the Riemannian metric (4) is $dA(z) = dx dy / y^2$. Accordingly, the integral of a function $f : H \rightarrow R$ with respect to dA is given by:

$$\int_H f(z) dA(z) = \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{f(x, y)}{y^2} dx dy \tag{8}$$

In many cases, the analytic computation of this integral can be greatly simplified by using polar coordinates (r, ϕ) defined with respect to some “origin” $\bar{z} \in H$. Polar coordinates (r, ϕ) map to the point $z(r, \phi)$ given by:

$$z(r, \phi) = \gamma_{\bar{z}} \left(r, \frac{\pi}{2} - \phi \right) \tag{9}$$

where the right-hand side is defined according to Equation (7). The polar coordinates (r, ϕ) do indeed define a global coordinate system of H , in the sense that the application that takes a complex number $re^{i\phi}$ to the point $z(r, \phi)$ in H is a diffeomorphism. The standard notation from differential geometry is:

$$\exp_{\bar{z}} (re^{i\phi}) = z(r, \phi) \tag{10}$$

In these coordinates, the Riemannian metric (4) takes on the form:

$$ds^2(z) = dr^2 + \sinh^2 r d\varphi^2 \tag{11}$$

The integral Equation (8) can be computed in polar coordinates using the formula [13],

$$\int_H f(z)dA(z) = \int_0^{2\pi} \int_0^{+\infty} (f \circ \exp_{\bar{z}}) (re^{i\varphi}) \sinh(r) dr d\varphi \tag{12}$$

where $\exp_{\bar{z}}$ was defined in Equation (10) and \circ denotes composition. This is particularly useful when $f \circ \exp_{\bar{z}}$ does not depend on φ .

2.3. Riemannian Gradient Descent

In this paper, the problem of minimizing, or maximizing, a differentiable function $f : H \rightarrow R$ will play a central role. A popular way of handling the minimization of a differentiable function defined on a Riemannian manifold (such as H) is through Riemannian gradient descent [14].

Here, the definition of Riemannian gradient is reviewed, and a generic description of Riemannian gradient descent is provided. The Riemannian gradient of f is here defined as a mapping $\nabla f : H \rightarrow C$ with the following property:

$$\frac{1}{y^2} \times \text{Re} \{ \nabla f(z) h^* \} = \text{Re} \{ df(z) h^* \} \tag{13}$$

for any complex number h , where Re denotes the real part, $*$ denotes conjugation and df is the “derivative”, $df = (\partial f/\partial x) + (\partial f/\partial y) i$. For example, if $f(z) = y$, it follows from Equation (13) that $\nabla f(z) = y^2$.

Riemannian gradient descent consists in following the direction of $-\nabla f$ at each step, with the length of the step (in other words, the step size) being determined by the user. The generic algorithm is, up to some variations, the following:

```

INPUT       $\hat{z} \in H$                 % Initial guess
WHILE       $\|\nabla f(\hat{z})\| > \varepsilon$     %  $\varepsilon \approx 0$  machine precision
            $\hat{z} \leftarrow \exp_{\hat{z}}(-\lambda \nabla f(\hat{z}))$  %  $\lambda > 0$  step size, depends on  $\hat{z}$ 
END WHILE
OUTPUT      $\hat{z}$                        % near critical point of  $f$ 
    
```

Here, in the condition for the while loop, $\|\nabla f(z_k)\|$ is the Riemannian norm of the gradient $\nabla f(z_k)$. In other words,

$$\|\nabla f(z_k)\|^2 = \frac{1}{y_k^2} \times \text{Re} \{ \nabla f(z_k) \nabla f(z_k)^* \}$$

Just like a classical gradient descent algorithm, the above Riemannian gradient descent consists in following the direction of the negative gradient $-\nabla f(\hat{z})$, in order to define a new estimate. This is repeated as long as the gradient is sensibly nonzero, in the sense of the loop condition.

The generic algorithm described above has no guarantee of convergence. Convergence and behavior near limit points depends on the function f , on the initialization of the algorithm and on the step sizes λ . For these aspects, the reader may consult [14](Chapter 4).

3. Riemannian Prior on the Univariate Normal Model

The current section introduces new prior distributions on the univariate normal model. These may be referred to as “Riemannian priors”, since they are entirely based on the Riemannian geometry of this model, and will also be called “Gaussian distributions”, when viewed as probability distributions on the Poincaré half plane.

Here, Section 3.1 defines in a rigorous way Gaussian distributions on H (based on the intuitive Formula (2)). A Gaussian distribution $G(\bar{z}, \gamma)$ has two parameters, $\bar{z} \in H$, called the center of mass, and $\gamma > 0$, called the dispersion parameter. Section 3.2 uses the Riemannian gradient descent algorithm Section 2.3 to provide an algorithm for computing maximum likelihood estimates of \bar{z} and γ . Finally, Section 3.3 proves that \bar{z} is the Riemannian center of mass or Karcher mean of the distribution $G(\bar{z}, \gamma)$, (Historically, it is more correct to speak of the “Fréchet mean”, since this concept was proposed by Fréchet in 1948 [15]), and that γ is uniquely related to mean square Rao’s distance from \bar{z} .

The reader may wish to note that the results of Section 3.3 are not used in the following, so this paragraph may be skipped on a first reading.

3.1. Gaussian Distributions on H

A Gaussian distribution $G(\bar{z}, \gamma)$ on H is a probability distribution with the following probability density function:

$$p(z|\bar{z}, \gamma) = \frac{1}{Z(\gamma)} \exp\left(\frac{-d^2(z, \bar{z})}{2\gamma^2}\right) \tag{14}$$

Here, $\bar{z} \in H$ is called the center of mass and $\gamma > 0$ the dispersion parameter of the distribution $G(\bar{z}, \gamma)$. The squared distance $d^2(z, \bar{z})$ refers to Rao’s distance (5). The probability density function (14) is understood with respect to the Riemannian volume element $dA(z)$. In other words, the normalization constant $Z(\gamma)$ is given by:

$$Z(\gamma) = \int_H f(z)dA(z) \quad f(z) = \exp\left(\frac{-d^2(z, \bar{z})}{2\gamma^2}\right)$$

Using polar coordinates, as in Equation (12), it is possible to calculate this integral explicitly. To do so, let (r, φ) , whose origin is \bar{z} . Then, $d^2(z, \bar{z}) = r^2$ when $z = z(r, \varphi)$, as in Equation (9). It follows that:

$$(f \circ \exp_{\bar{z}})(r, \varphi) = \exp\left(\frac{-r^2}{2\gamma^2}\right) \tag{15}$$

According to Equation (12), the integral $Z(\gamma)$ reduces to:

$$Z(\gamma) = \int_0^{2\pi} \int_0^{+\infty} \exp\left(\frac{-r^2}{2\gamma^2}\right) \sinh(r) dr d\varphi$$

which is readily calculated,

$$Z(\gamma) = 2\pi \times \sqrt{\frac{\pi}{2}} \gamma \times e^{\frac{\gamma^2}{2}} \times \operatorname{erf}\left(\frac{\gamma}{\sqrt{2}}\right) \tag{16}$$

where erf denotes the error function. Formula (16) completes the definition of the Gaussian distribution $G(\bar{z}, \gamma)$. This definition is the same as suggested in [11], with the difference that, in the present work, it has been possible to compute exactly the normalization constant $Z(\gamma)$.

It is noteworthy that the normalization constant $Z(\gamma)$ depends only on γ and not on \bar{z} . This shows that the shape of the probability density function (14) does not depend on \bar{z} , which only plays the role of a location parameter. At a deeper mathematical level, this reflects the fact that H is a homogeneous Riemannian space [13].

The probability density function (14) bears a clear resemblance to the usual Gaussian (or normal) probability density function. Indeed, both are proportional to the exponential minus the “square distance”, but in one case, the distance is interpreted as Euclidean distance and, in the other (that of Equation (14)) as Rao’s distance.

3.2. Maximum Likelihood Estimation of \bar{z} and γ

Consider the problem of computing maximum likelihood estimates of the parameters \bar{z} and γ of the Gaussian distribution $G(\bar{z}, \gamma)$, based on independent samples $\{z_i\}_{i=1}^N$ from this distribution. Given the expression (14) of the density $p(z|\bar{z}, \gamma)$, the log-likelihood function $\ell(\bar{z}, \gamma)$ can be written,

$$\ell(\bar{z}, \gamma) = -N \log\{Z(\gamma)\} - \frac{1}{2\gamma^2} \sum_{i=1}^N d^2(z_i, \bar{z}) \quad (17)$$

Since \bar{z} only appears in the second term, the maximum likelihood estimate of \bar{z} , say \hat{z} , can be computed first. It is given by the minimization problem:

$$\hat{z} = \operatorname{argmin}_{z \in H} \frac{1}{2} \sum_{i=1}^N d^2(z_i, z) \quad (18)$$

In other words, the maximum likelihood estimate \hat{z} minimizes the sum of squared Rao distances to the samples z_i . This exhibits \hat{z} as the Riemannian center of mass, also called the Karcher or the Fréchet mean [16], of the samples z_i .

The notion of Riemannian center of mass is currently a widely popular one in signal and image processing, with applications ranging from blind source separation and radar signal processing [17,18] to shape and motion analysis [19,20]. The definition of Gaussian distributions, proposed in the present paper, shows how the notion of Riemannian center of mass is related to maximum likelihood estimation, thereby giving it a statistical foundation.

An original result, due to Cartan and cited in Equation [16], states that \hat{z} , as defined in Equation (18), exists and is unique, since H , with the Riemannian distance (4), has constant negative curvature. Here, \hat{z} is computed using Riemannian gradient descent, as described in Section 2.3. The cost function f to be minimized is given by (the factor N^{-1} is conventional),

$$f(z) = \frac{1}{2N} \sum_{i=1}^N d^2(z_i, z) \quad (19)$$

Its Riemannian gradient $\nabla f(z)$ is easily found by noting the following fact. Let $f_i(z) = (1/2)d^2(z, z_i)$. Then, the Riemannian gradient of this function is (see [21] (page 407)),

$$\nabla f_i(z) = \log_z(z_i) \quad (20)$$

where $\log_z : H \rightarrow C$ is the inverse of $\exp_z : C \rightarrow H$. It follows from Equation (20) that,

$$\nabla f(z) = \frac{1}{N} \sum_{i=1}^N \log_z(z_i) \tag{21}$$

The analytic expression of \log_z , for any $z \in H$, will be given below (see Equation (23)).

Here, the gradient descent algorithm for computing \hat{z} is described. This algorithm uses a constant step size λ , which is fixed manually.

Once the maximum likelihood estimate \hat{z} has been computed, using the gradient descent algorithm, the maximum likelihood estimate of γ , say $\hat{\gamma}$, is found by solving the equation:

$$F(\gamma) = \frac{1}{N} \sum_{i=1}^N d^2(z_i, \hat{z}) \quad \text{where } F(\gamma) = \gamma^3 \times \frac{d}{d\gamma} \log\{Z(\gamma)\} \tag{22}$$

The gradient descent algorithm for computing \hat{z} is the following,

```

INPUT      {z1, ..., zN}           % N independent samples from G(z̄, γ)
           ẑ ∈ H                     % Initial guess
WHILE      ||∇f(ẑ)|| > ε           % ε ≈ 0 machine precision
           ẑ ← expẑ(-λ∇f(ẑ))    % ∇f(ẑ) given by Equation (21)
                                           % step size λ is constant
END WHILE
OUTPUT     ẑ                         % near Riemannian center of mass
    
```

Application of Formula (21) requires computation of $\log_z(z_i)$ for $i = 1, \dots, N$. Fortunately, this can be done analytically as follows. In general, for $\hat{z} = (\bar{x}, \bar{y})$,

$$\log_{\hat{z}}(z) = \bar{y} \log_i \left(\frac{z - \bar{x}}{\bar{y}} \right) \tag{23}$$

where \log_i is found by inverting Equation (6). Precisely,

$$\log_i(z) = r e^{i\varphi} \tag{24}$$

where, for $z = (x, y)$ with $x \neq 0$,

$$r = \text{acosh} \left(1 + \frac{x^2 + (y - 1)^2}{2y} \right)$$

and:

$$\cos(\varphi) = \frac{x}{y \sinh(r)} \quad \sin(\varphi) = \frac{\cosh(r) - y^{-1}}{\sinh(r)}$$

and, for $z = (0, y)$,

$$\log_i(z) = \ln(y)i$$

with \ln denoting the natural logarithm.

3.3. Significance of \bar{z} and γ

The parameters \bar{z} and γ of a Gaussian distribution $G(\bar{z}, \gamma)$ have been called the center of mass and the dispersion parameter. In the present paragraph, it is proven that,

$$\bar{z} = \operatorname{argmin}_{z \in H} \frac{1}{2} \int_H d^2(z', z) p(z' | \bar{z}, \gamma) dA(z') \tag{25}$$

and also that:

$$F(\gamma) = \int_H d^2(z', \bar{z}) p(z' | \bar{z}, \gamma) dA(z') \tag{26}$$

where $F(\gamma)$ was defined in Equation (22) and $p(z' | \bar{z}, \gamma)$ is the probability density function of $G(\bar{z}, \gamma)$, given in Equation (14).

Note that Equations (25) and (26) are asymptotic versions of Equations (18) and (22). Indeed, Equations (25) and (26) can be written:

$$\bar{z} = \operatorname{argmin}_{z \in H} \frac{1}{2} E_{\bar{z}, \gamma} d^2(z', z) \quad F(\gamma) = E_{\bar{z}, \gamma} d^2(z, \bar{z}) \tag{27}$$

where $E_{\bar{z}, \gamma}$ denotes the expectation with respect to $G(\bar{z}, \gamma)$, and the expectation is carried out on the variable z' in the first formula. Now, these two formulae are the same as Equations (18) and (22), but with expectation instead of empirical mean.

Note, moreover, that Equations (25) and (26) can be interpreted as follows. If z' is distributed according to the Gaussian distribution $G(\bar{z}, \gamma)$, then Equation (25) states that \bar{z} is the unique point, out of all $z \in H$, which minimizes the expectation of squared Rao's distance to z' . Moreover, Equation (26) states that the expectation of squared Rao's distance between \bar{z} and z' is equal to $F(\gamma)$, so $F(\gamma)$ is the least possible expected squared Rao's distance between a point $z \in H$ and z' . This interpretation justifies calling \bar{z} the center of mass of $G(\bar{z}, \gamma)$ and shows that γ is uniquely related to the expected dispersion, as measured by squared Rao's distance, away from \bar{z} .

In order to prove Equation (25), consider the log-likelihood function,

$$\ell(\bar{z}, \gamma; z) = -\log\{Z(\gamma)\} - \frac{1}{2\gamma^2} d^2(z, \bar{z}) \tag{28}$$

Let $f_z(\bar{z}) = (1/2)d^2(z, \bar{z})$. The score function, with respect to \bar{z} is, by definition,

$$\nabla_{\bar{z}} \ell(\bar{z}, \gamma; z) = \nabla f_z(\bar{z}) \tag{29}$$

where $\nabla_{\bar{z}}$ indicates the Riemannian gradient (defined in Equation (13) of Section 2.3) is with respect to the variable \bar{z} . Under certain regularity conditions, which are here easily verified, the expectation of the score function is identically zero,

$$E_{\bar{z}, \gamma} \nabla f_z(\bar{z}) = 0 \tag{30}$$

Let $f(z)$ be defined by:

$$f(z) = E_{\bar{z}, \gamma} f_{z'}(z) = \frac{1}{2} E_{\bar{z}, \gamma} d^2(z', z)$$

with the expectation carried out on the variable z' . Clearly, $f(z)$ is the expression to be minimized in Equation (25) (or in the first formula in Equation (27), which is just the same). By interchanging Riemannian gradient and expectation,

$$\nabla f(\bar{z}) = E_{\bar{z},\gamma} \nabla f_z(\bar{z}) = 0$$

where the last equality follows from Equation (30).

It has just been proved that \bar{z} is a stationary point of f (a point where the gradient is zero). Theorem 2.1 in [16] states the function f has one and only one stationary point, which is moreover a global minimizer. This concludes the Proof (25).

The proof of Equation (26) follows exactly the same method, defining the score function with respect to γ and noting that its expectation is identically zero.

4. Classification of Univariate Normal Populations

The previous section studied Gaussian distributions on H , “as they stand”, focusing on the fundamental issue of maximum likelihood estimation of their parameters. The present Section considers the use of Gaussian distributions as prior distributions on the univariate normal model.

The main motivation behind the introduction of Gaussian distributions is that a Gaussian distribution $G(\bar{z}, \gamma)$ can be used to give a geometric representation of a cluster or class of univariate normal populations. Recall that each point $(x, y) \in H$ is identified with a univariate normal population with mean $\mu = \sqrt{2}x$ and standard deviation $\sigma = y$. The idea is that populations belonging to the same cluster, represented by $G(\bar{z}, \gamma)$, should be viewed as centered on \bar{z} and lying within a typical distance determined by γ .

In the remainder of this Section, it is shown how the maximum likelihood estimation algorithm of Section 3.2 can be used to fit the hyperparameters \bar{z} and γ to data, consisting in a class $\mathcal{S} = \{S_i; i = 1, \dots, K\}$ of univariate normal populations. This is then applied to the problem of the classification of univariate normal populations. The whole development is based on marginalized likelihood estimation, as follows.

Assume each population S_i contains N_i points, $S_i = \{s_j; j = 1, \dots, N_i\}$, and the points s_j , in any class, are drawn from a univariate normal distribution with mean μ and standard deviation σ . The focus will be on the asymptotic case where the number N_i of points in each population S_i is large.

In order to fit the hyperparameters \bar{z} and γ to the data \mathcal{S} , assume moreover that the distribution of $z = (x, y)$, where $(x, y) = (\mu/\sqrt{2}, \sigma)$, is a Gaussian distribution $G(\bar{z}, \gamma)$. Then, the distribution of \mathcal{S} can be written in integral form:

$$p(\mathcal{S}|\bar{z}, \gamma) = \prod_{i=1}^K \int_H p(S_i|z)p(z|\bar{z}, \gamma)dA(z) \tag{31}$$

where $p(z|\bar{z}, \gamma)$ is the probability density of a Gaussian distribution $G(\bar{z}, \gamma)$, defined in Equation (14). Moreover, expressing $p(S_i|z)$ as a product of univariate normal distributions $p(s_j|z)$, it follows,

$$p(\mathcal{S}|\bar{z}, \gamma) = \prod_{i=1}^K \int_H \prod_{j=1}^{N_i} p(s_j|z)p(z|\bar{z}, \gamma)dA(z) \tag{32}$$

This expression, given the data \mathcal{S} , is to be maximized over (\bar{z}, γ) . Using the Laplace approximation, this task is reduced to the maximum likelihood estimation problem, addressed in Section 3.2.

The Laplace approximation will here be applied in its “basic form” [9]. That is, up to terms of order N_i^{-1} . To do so, write each of the integrals in Equation (32), using Equation (8) of Section 2.2. These integrals then take on the form:

$$\int_0^{+\infty} \int_{-\infty}^{+\infty} \prod_{j=1}^{N_i} |2\pi y^2|^{-1/2} \exp\left(\frac{-(s_j - \sqrt{2} x)^2}{2y^2}\right) \times p(z|\bar{z}, \gamma) \times \frac{1}{y^2} dx dy \tag{33}$$

where the univariate normal distribution $p(s_j|z)$ has been replaced by its full expression. Now, this expression can be written $p(s_j|z) = \exp[-N_i h(x, y)]$, where:

$$h(x, y) = -\frac{1}{2} \ln(2\pi y^2) - \frac{B_i^2 + V_i^2}{2y^2}$$

Here, B_i^2 and V_i^2 are the empirical bias and variance, within population S_i ,

$$B_i = \hat{S}_i - \sqrt{2} x \quad V_i^2 = N_i^{-1} \sum_{j=1}^{N_i} (\hat{S}_i - s_j)^2$$

where \hat{S}_i is the empirical mean of the population $\hat{S}_i = N_i^{-1} \sum_{j=1}^{N_i} s_j$.

The expression $h(x, y)$ is maximized when $x = \hat{x}_i$ and $y = \hat{y}_i$, where $\hat{z}_i = (\hat{x}_i, \hat{y}_i)$ is the couple of maximum likelihood estimates of the parameters (x, y) , based on the population S_i .

According to the Laplace approximation, the integral Equation (33) is equal to:

$$2\pi |\partial^2 h(\hat{x}_i, \hat{y}_i)|^{-1/2} \times \exp[-N_i h(\hat{x}_i, \hat{y}_i)] \times p(\hat{z}_i|\bar{z}, \gamma) \times \frac{1}{\hat{y}_i^2} + O(N_i^{-1})$$

where $\partial^2 h(\hat{x}_i, \hat{y}_i)$ is the matrix of second derivatives of h , and $|\cdot|$ denotes the determinant. Now, since h is essentially the logarithm of $p(s_j|z)$, a direct calculation shows that $\partial^2 h(\hat{x}_i, \hat{y}_i)$ is the same as the Fisher information matrix derived in Section 2.1 (where it was denoted $I(z)$). Thus, the first factor in the above expression is $2\pi \hat{y}_i^2$, and cancels out with the last factor.

Finally, the Laplace approximation of the integral Equation (33) reads:

$$2\pi \times \exp[-N_i h(\hat{x}_i, \hat{y}_i)] \times p(\hat{z}_i|\bar{z}, \gamma) + O(N_i^{-1})$$

and the resulting approximation of the distribution of \mathcal{S} , as given by Equation (32), can be written:

$$p(\mathcal{S}|\bar{z}, \gamma) \approx \prod_{i=1}^K \alpha \times p(\hat{z}_i|\bar{z}, \gamma) \tag{34}$$

where α is a constant, which does not depend either on the data or on the parameters, and $p(\hat{z}_i|\bar{z}, \gamma)$ has the expression (14).

Accepting this expression to give the distribution of the data \mathcal{S} , conditionally on the hyperparameters (\bar{z}, γ) , the task of estimating these hyperparameters becomes the same as the maximum likelihood estimation problem, described in Section 3.2.

In conclusion, if one assumes the populations S_i belong to a single cluster or class \mathcal{S} and wishes to fit the hyperparameters \bar{z} and γ of a Gaussian distribution representing this cluster, it is enough to start

by computing the maximum likelihood estimates \hat{x}_i and \hat{y}_i for each population S_i and then to consider these as input to the maximum likelihood estimation algorithm described in Section 3.2.

The same reasoning just carried out, using the Laplace approximation, can be generalized to the problem of classification of univariate normal populations. Indeed, assume that classes $\{S_L, L = 1, \dots, C\}$, each containing some number K_L of univariate normal populations, have been identified based on some training sequence. Using the Laplace approximation and the maximum likelihood estimation approach of Section 3.2, to each one of these classes, it is possible to fit hyperparameters (\bar{z}_L, γ_L) of a Gaussian distribution $G(\bar{z}_L, \gamma_L)$ on H .

For a test population S_t , the maximum likelihood rule, for deciding which of the classes S_L this test population S_t belongs to, requires finding the following maximum:

$$L^* = \operatorname{argmax}_L p(S_t | \bar{z}_L, \gamma_L) \quad (35)$$

and assigning the test population S_t to the class with label L^* . If the number of points N_t in the population S_t is large, the Laplace approximation, in the same way used above, approximates the maximum in Equation (35) by:

$$L^* = \operatorname{argmax}_L p(\hat{z}_t | \bar{z}_L, \gamma_L) \quad (36)$$

where $\hat{z}_t = (\hat{x}_t, \hat{y}_t)$ is the couple of maximum likelihood estimates computed based on the test population S_t and where $p(\hat{z}_t | \bar{z}_L, \gamma_L)$ is given by Equation (14). Now, writing out Equation (14), the decision rule becomes:

$$L^* = \operatorname{argmax}_L \left(-\log \{Z(\gamma_L)\} - \frac{1}{2\gamma_L^2} d^2(\hat{z}_t, \bar{z}_L) \right) \quad (37)$$

Under the homoscedasticity assumption, that all of the γ_L are equal, this decision rule essentially becomes the same as the one proposed in [5], which requires S_t to be assigned to the “nearest” cluster, in terms of Rao’s distance. Indeed, if all the γ_L are equal, then Equation (37) is the same as,

$$L^* = \operatorname{argmin}_L d^2(\hat{z}_t, \bar{z}_L) \quad (38)$$

This decision rule is expected to be less efficient than the one proposed in Equation (37), which also takes into account the uncertainty associated with each cluster, as measured by its dispersion parameter γ_L .

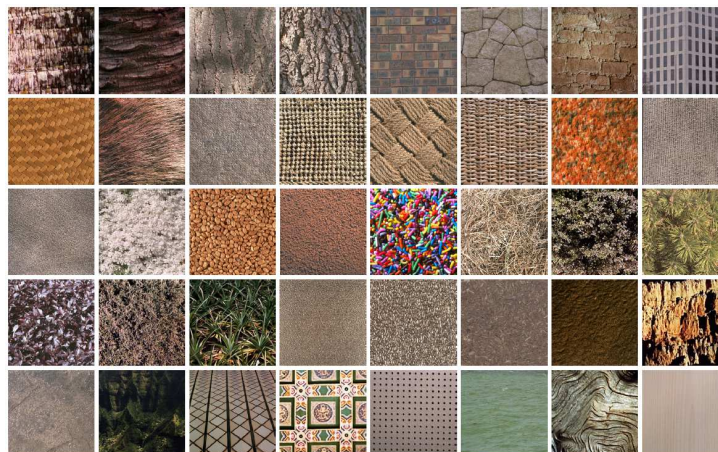
5. Application to Image Classification

In this section, the framework proposed in Section 4, for classification of univariate normal populations, is applied to texture image classification using Gabor filters. Several authors have found that Gabor energy features are well-suited texture descriptors. In the following, consider 24 Gabor energy sub-bands that are the result of three scales and eight orientations. Hence, each texture image can be decomposed as the collection of those 24 sub-bands. For more information concerning the implementation, the interested reader is referred to [22].

Starting from the VisTeX database of 40 images [10] (these are displayed in Figure 1), each image was divided into 16 non-overlapping subimages of 128×128 pixels each. A training sequence was formed by choosing randomly eight subimages out of each image. To each subimage in the training sequence, a bank of 24 Gabor filters was applied. The result of applying a Gabor filter with scale s and orientation

o to a subimage i belonging to an image L is a univariate normal population $S_{i,s,o}$ of 128×128 points (one point for each pixel, after the filter is applied).

Figure 1. Forty images of the VisTex database.



These populations $S_{i,s,o}$ (called sub-bands) are considered independent, each one of them univariate normal with mean $\mu_{i,s,o} = \sqrt{2}x_{i,s,o}$, standard deviation $\sigma_{i,s,o} = y_{i,s,o}$ and with $z_{i,s,o} = (x_{i,s,o}, y_{i,s,o})$. The couple of maximum likelihood estimates for these parameters is denoted $\hat{z}_{i,s,o} = (\hat{x}_{i,s,o}, \hat{y}_{i,s,o})$. An image L (recall, there are 40 images) contains, in each sub-band, eight populations $S_{i,s,o}$, with which hyperparameters $\bar{z}_{L,s,o}$ and $\gamma_{L,s,o}$ are associated, by applying the maximum likelihood estimation algorithm of Section 3.2 to the inputs $\hat{z}_{i,s,o}$.

If S_t is a test subimage, then one should begin by applying the 24 Gabor filters to it, obtaining independent univariate normal populations $S_{t,s,o}$, and then compute for each population the couple of maximum likelihood estimates $\hat{z}_{t,s,o} = (\hat{x}_{t,s,o}, \hat{y}_{t,s,o})$. The decision rule Equation (37) of Section 4 requires that S_t should be assigned to the image L^* , which realizes the maximum:

$$L^* = \operatorname{argmax}_L \sum_{s,o} -\log\{Z(\gamma_{L,s,o})\} - \frac{1}{2\gamma_{L,s,o}^2} d^2(\hat{z}_{t,s,o}, \bar{z}_{L,s,o}) \tag{39}$$

When considering the homoscedasticity assumption, *i.e.*, $\gamma_{L,s,o} = \gamma_{s,o}$ for all L , this decision rule becomes:

$$L^* = \operatorname{argmin}_L \sum_{s,o} d^2(\hat{z}_{t,s,o}, \bar{z}_{L,s,o}) \tag{40}$$

For this concrete application, to the VisTex database, it is pertinent to compare the rate of successful classification (or overall accuracy) obtained using the Riemannian prior, based on the framework of Section 4, to that obtained using a more classical conjugate prior, *i.e.*, a normal-inverse gamma distribution of the mean $\mu = \sqrt{2}x$ and the standard deviation $\sigma = y$. This conjugate prior is given by:

$$p(\mu|\sigma, \mu_p, \kappa_p) = \frac{\sqrt{\kappa_p}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\kappa_p}{2\sigma^2}(\mu - \mu_p)^2\right)$$

with an inverse gamma prior, on σ^2 ,

$$p(\sigma^2|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right) \tag{41}$$

Using this conjugate prior, instead of a Riemannian prior, and following the procedure of applying the Laplace approximation, a different decision rule is obtained, where L^* is taken to be the maximum of the following expression:

$$\sum_{s,o} \frac{\ln \kappa_{pL,s,o}}{2} - \frac{\kappa_{pL,s,o}}{2\hat{y}_{t,s,o}^2} \left(\sqrt{2}\hat{x}_{t,s,o} - \mu_{pL,s,o} \right)^2 + \alpha_{L,s,o} \ln \beta_{L,s,o} - \ln \Gamma(\alpha_{L,s,o}) - 2(\alpha_{L,s,o} + 1) \ln \hat{y}_{t,s,o} - \frac{\beta_{L,s,o}}{\hat{y}_t^2} \tag{42}$$

where, as in Equation (39), $\hat{x}_{t,s,o}$ and $\hat{y}_{t,s,o}$ are the maximum likelihood estimates computed for the population $S_{t,s,o}$.

Both the Riemannian and conjugate priors have been applied to the VisTex database, with half of the database used for training and half for testing. In the course of 100 Monte Carlo runs, a significant gain of about 3% is observed with the Riemannian prior compared to the conjugate prior. This is summarized in the following table.

Prior Model	Overall Accuracy
Riemannian prior Equation (39)	71.88% ± 2.16%
Riemannian prior, homoscedasticity assumption Equation (40)	69.06% ± 1.96%
Conjugate prior Equation (42)	68.73% ± 2.92%

Recall that the overall accuracy is the ratio of the number of successfully classified subimages to the total number of subimages. The table shows that the use of a Riemannian prior, even under a homoscedasticity assumption, yields significant improvement upon the use of a conjugate prior.

6. Conclusions

Motivated by the problem of the classification of univariate normal populations, this paper introduced a new class of prior distributions on the univariate normal model. With the univariate normal model viewed as the Poincaré half plane H , these new prior distributions, called Gaussian distributions, were meant to reflect the geometric picture (in terms of Rao’s distance) that a cluster or class of univariate normal populations can be represented as having a center $\bar{z} \in H$ and a “variance” or dispersion γ^2 . Precisely, a Gaussian distribution $G(\bar{z}, \gamma)$ has a probability density function $p(z)$, with respect to Riemannian volume of the Poincaré half plane, which is proportional to $\exp\left(-\frac{d^2(z, \bar{z})}{2\gamma^2}\right)$. Using Gaussian distributions as prior distributions in the problem of the classification of univariate normal populations was shown to lead to a new, more general and efficient decision rule. This decision rule was implemented in a real-world application to texture image classification, where it led to significant improvement in performance, in comparison to decision rules obtained by using conjugate priors.

The general approach proposed in this paper contains several simplifications and approximations, which could be improved upon in future work. First, it is possible to use different prior distributions, which are more geometrically rich than Gaussian distributions, to represent classes of univariate normal populations. For example, it may be helpful to replace Gaussian distributions that are “isotropic”, in the

sense of having a scalar dispersion parameter γ , by non-isotropic distributions, with a dispersion matrix Γ (a 2×2 symmetric positive definite matrix). Another possibility would be to represent each class of univariate normal populations by a finite mixture of Gaussian distributions, instead of representing it by a single Gaussian distribution.

These variants, which would allow classes with a more complex geometric structure to be taken into account, can be integrated in the general framework proposed in the paper, based on: (i) fitting each class to a prior distribution (Gaussian non-isotropic, mixture of Gaussians); and (ii) choosing, for a test population, the most adequate class, based on a decision rule. These two steps can be realized as above, through the Laplace approximation and maximum likelihood estimation, or through alternative techniques, based on Markov chain Monte Carlo stochastic optimization.

In addition to generalizing the approach of this paper and improving its performance, a further important objective for future work will be to extend it to other parametric models, beyond univariate normal models. Indeed, there is an increasing number of parametric models (generalized Gaussian, elliptical models, *etc.*), whose Riemannian geometry is becoming well understood and where the present approach may be helpful.

Author Contributions

Salem Said carry out the mathematical development, and specify the algorithms, appearing in Sections 2, 3 and 4. Lionel Bombrun carry out all numerical simulations, and to propose the theoretical development of Section 4. Yannick Berthoumieu devise the main idea of the paper. That is, use of Riemannian priors as geometric representation of a class or cluster of univariate normal population. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Amari, S.I; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2000.
2. Rao, C.R. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.
3. Kass, R.E. The geometry of asymptotic inference. *Stat. Sci.* **1989**, *4*, 188–234.
4. Amari, S.I. Natural gradient works efficiently in learning. *Neur. Comput.* **1998**, *10*, 251–276.
5. Nielsen, F; Nock, R. Hyperbolic Voronoi diagrams made easy. **2009**, arXiv:0903.3287.
6. Lenglet, C.; Rousson, M.; Deriche, R.; Fougeras, O. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *J. Math. Imaging Vis.* **2006**, *25*, 423–444.
7. Verdoolaege, G.; Scheunders, P. On the geometry of multivariate generalized Gaussian models. *J. Math. Imaging Vis.* **2012**, *43*, 180–193.

8. Berkane, M.; Oden, K. Geodesic estimation in elliptical distributions. *J. Multival. Anal.* **1997**, *63*, 35–46.
9. Erdélyi, A. Asymptotic Expansions; Dover Books: Mineola, New York, NY, USA, 2010.
10. MIT Vision and Modeling Group. Vision Texture. Available online: <http://vismod.media.mit.edu/pub/VisTex> (accessed on 10 June 2014).
11. Pennec, X. Intrinsic statistics on Riemannian manifold: Basic tools for geometric measurements. *J. Math. Imaging Vis.* **2006**, *25*, 127–154.
12. Atkinson, C.; Mitchell, A.F.S. Rao's distance measure. *Sankhya Ser. A* **1981**, *43*, 345–365.
13. Gallot, S.; Hulin, D.; Lafontaine, J. *Riemannian Geometry*; Springer-Verlag: Berlin, Germany, 2004.
14. Absil, P.A.; Mahony, R.; Sepulchre, R. Optimization Algorithms on Matrix Manifolds; Princeton University Press: Cambridge, MA, USA, 2006.
15. Fréchet, M. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'I.H.P.* **1948**, *10*, 215–310. (In French)
16. Afsari, B. Riemannian L^p center of mass: Existence, Uniqueness and convexity. *Proc. Am. Math. Soc.* **2011**, *139*, 655–673.
17. Manton, J.H. A centroid (Karcher mean) approach to the joint approximate diagonalisation problem: The real symmetric case. *Digit. Sign. Process.* **2006**, *16*, 468–478.
18. Arnaudon, M.; Barbaresco, F. Riemannian medians and means with applications to RADAR signal processing. *IEEE J. Sel. Top. Sign. Process.* **2013**, *7*, 595–604.
19. Le, H. On the consistency of procrustean mean shapes. *Adv. Appl. Prob.* **1998**, *30*, 53–63.
20. Turaga, P.; Veeraraghavan, A.; Chellappa, R. Statistical Analysis on Stiefel and Grassmann Manifolds with Applications in Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; doi: 10.1109/CVPR.2008.4587733.
21. Chavel, I. Riemannian geometry: A modern introduction; Cambridge University Press: Princeton, MA, USA, 2008.
22. Grigorescu, S.E.; Petkov, N.; Kruijzinga, P. Comparison of texture features based on Gabor filter. *IEEE Trans. Image Process.* **2002**, *11*, 1160–1167.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).