



HAL
open science

Kernel density estimation on spaces of Gaussian distributions and symmetric positive definite matrices

Emmanuel Chevallier, Emmanuel Kalunga, Jesus Angulo

► **To cite this version:**

Emmanuel Chevallier, Emmanuel Kalunga, Jesus Angulo. Kernel density estimation on spaces of Gaussian distributions and symmetric positive definite matrices. 2015. hal-01245712

HAL Id: hal-01245712

<https://hal.science/hal-01245712>

Preprint submitted on 17 Dec 2015

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.

Kernel density estimation on spaces of Gaussian distributions and symmetric positive definite matrices

Emmanuel Chevallier^a, Emmanuel Kalunga^b, Jesús Angulo^a

^a CMM-Centre de Morphologie Mathématique, MINES ParisTech; France

^b Tshwane University of Technology; South Africa

`emmanuel.chevallier@mines-paristech.fr`

June 2014

Abstract

This paper analyses the kernel density estimation on spaces of Gaussian distributions endowed with different metrics. Explicit expressions of kernels are provided for the case of the 2-Wasserstein metric on multivariate Gaussian distributions and for the Fisher metric on multivariate centred distributions. Under the Fisher metric, the space of multivariate centred Gaussian distributions is isometric to the space of symmetric positive definite matrices under the affine-invariant metric and the space of univariate Gaussian distributions is isometric to the hyperbolic space. Thus kernel are also valid on these spaces. The density estimation is successfully applied to a classification problem of electro-encephalographic signals.

1 Introduction

Signal and image processing are dealing nowadays with data laying in more and more various spaces. Each type of data posses its own geometric and algebraic structures. Data laying in spaces of Gaussian laws or symmetric positive definite matrices $SPD(n)$ are particularly present in signal and image processing, hence the importance of the study of the geometry of these spaces.

The most common geometries of Gaussian laws are the Fisher information metric and the Wasserstein metric. The Fisher metric on multivariate centered Gaussians happens to be similar to the affine invariant metric on

SPD(n), that is to say the metric invariant under the action of invertible matrices $Gl(n)$, i.e.,

$$G \cdot \Sigma \mapsto G\Sigma G^t.$$

Furthermore the Fisher metric on univariate Gaussians is a hyperbolic metric of dimension two. Thus understanding the Fisher metric has consequences beyond Gaussian laws. The Wasserstein metric is another name for the earth mover's distance. Let S be a set of probability distributions on E . A fundamental difference between the Fisher metric and the Wasserstein metric on S , is that the Fisher metric do not relay on a geometry of E while the Wasserstein metric does. Indeed, the Wasserstein metric depends on the geometry of E through a transportation cost. The dependence on the geometry of E constitutes one of the main specificities and interests of the Wasserstein metric.

Median and means are fundamental quantities in signal and image processing. Several studies already addressed their definition and computation for the Fisher and the Wasserstein metrics, see for instance [20, 32, 1, 3, 5]. Although being a secondary problem, density estimation is an important tool of signal and image processing which is still little studied on spaces of Gaussian laws. It is in particular useful, for instance, for segmenting point clouds or for Bayesian classification. The study of distributions on SPD(n) is subject to recent studies. In [23], it is proposed for instance a generalization of the Gaussian distribution on SPD(n). To our knowledge, the existing literature dedicated to density estimation is mainly restricted to [13], [22] and [4]. These previous works have focused on techniques derived from the orthogonal series density estimation for the Fisher metric in the case of centered multivariate and univariate Gaussian laws. This paper considers several non parametric density estimators. Due to their drawbacks in the studied spaces, histograms and orthogonal series density estimators are not deeply analyzed. The paper is focused on the kernel density estimator, which often appears to be the easiest to use on Riemannian manifolds.

The paper is organized as follow. Section 2 introduces basics and notations on measure theory and Riemannian geometry. Main techniques of density estimation on Riemannian manifolds are reviewed in section 3. The quantities necessary to the density estimation are computed for different spaces of Gaussian distributions in sections 5-7.3. The computation of kernels for the Fisher and Wasserstein metric are provided. The expression of the different factors constituting the kernels in the case of the Fisher metric are already known in the literature, see [29, 31]. To our knowledge, the expression of kernels for the Wasserstein metric is new. Being an im-

portant problem for practical applications, the density estimation of partial quantities such as mean eigenvalues and rotation is discussed in section 9. Section 10 provides some experimental results.

2 Basics and notations on measure theory and Riemannian geometry

2.1 Measure theory and probability

Let $\Omega = \cup\{\omega\}$ be a set of outcome with a σ -algebra \mathcal{A}_Ω and a measure μ_Ω . The σ -algebra is the subset of $\mathcal{P}(\Omega)$ where it is possible to evaluate μ_Ω . Let V be a space with its own σ -algebra \mathcal{A}_V . A random variable X is a measurable function from Ω to V . A measurable function is a function such that:

$$A \in \mathcal{A}_V \Rightarrow X^{-1}(A) \in \mathcal{A}_\Omega.$$

The measurable hypothesis implies that the variable X can be seen as a transport of the measure of probability μ from the space Ω to the space V . Indeed, the function μ_X on \mathcal{A}_V defined by:

$$\mu_X(A) = \mu_\Omega(X^{-1}(A)),$$

is a probability measure on V .

Let E be a set and \mathcal{A}_E a σ -algebra of E . Let μ_1 and μ_2 be two measures on E . If the following holds,

$$\forall A \in \mathcal{A}_E, \mu_1(A) = \int_A f d\mu_2,$$

$f = \frac{d\mu_1}{d\mu_2}$ is called the density of μ_1 with respect to μ_2 . Let μ_V be a reference measure on V . If for some specific random variable X , the measure μ_X has a density f with respect to the reference measure of V , f is called the density of the random variable X . In a finite dimensional vector space there is a unique measure, up to a scaling factor, invariant under translations. The translation invariant measure that normalizes the unit hyper-cube is called the Lebesgue measure. In the case where V is a vector space the reference measure μ_V is often the Lebesgue measure. Fig. 1 illustrates a change of reference measure.

Let E and F be sets equipped with σ -algebras, and $f : E \rightarrow F$ an application. Any measure μ defined on E can be transformed into a measure $f^*(\mu)$ on F by:

$$f^*(\mu)(A) = \mu(f^{-1}(A)). \tag{1}$$

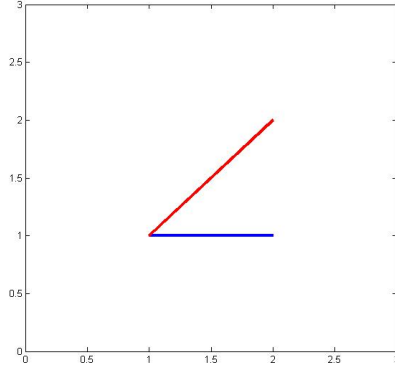


Figure 1: Let X be a random variable valued in $[1, 2]$. The blue curve is the density of μ_X with respect to the Lebesgue measure of \mathbb{R} restricted to $[1, 2]$. The red curve is the density of μ_X with respect to the measure $\mu([a, b]) = \log(b) - \log(a)$.

The smallest σ -algebra of $E \times F$ that contains all the products $U \times V, (U, V) \in \mathcal{A}_E \times \mathcal{A}_F$ is called the product σ -algebra of $E \times F$. A product measure μ on $E \times F$ is a measure such that there exist two measures μ_E and μ_F respectively on E and F with:

$$\forall (A, B) \in \mathcal{A}_E \times \mathcal{A}_F, \mu(A \times B) = \mu_E(A)\mu_F(B).$$

2.2 Riemannian geometry

Let \mathcal{M} be a topological space, homeomorphic to an open subset of \mathbb{R}^n . A homeomorphism is a continuous map whose converse is also continuous. Let ϕ be an homeomorphism from an open subset $U_\phi \subset \mathbb{R}^n$ to \mathcal{M} . ϕ is referred to as a parametrization of \mathcal{M} . A Riemannian metric is a smooth field of scalar product on U_ϕ . In other words, a Riemannian metric associates a positive definite matrix $G(x)$ to each point $x \in U_\phi$ depending smoothly on the point, see Fig. 2.

A smooth path is a map $\gamma : [a, b] \rightarrow \mathcal{M}$ such that $\phi^{-1} \circ \gamma$ is continuously smooth. Let γ be such a path. The Riemannian metric induces a notion of length on smooth paths as follows:

$$L(\gamma) = \int_a^b \sqrt{\langle (\phi^{-1} \circ \gamma)'(t), (\phi^{-1} \circ \gamma)'(t) \rangle_{(\phi^{-1} \circ \gamma)(t)}} dt, \quad (2)$$

where $\langle \cdot, \cdot \rangle_{(\phi^{-1} \circ \gamma)(t)}$ is the scalar product attached to the point $(\phi^{-1} \circ \gamma)(t)$. The notion of shortest path between two points induces a distance on \mathcal{M} .

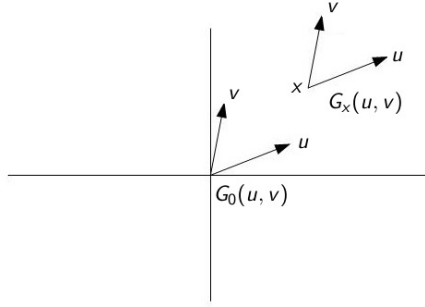


Figure 2: Riemannian space.

The distance $d(p, q)$ associated with the Riemannian metric is given by

$$d(p, q) = \inf_{\gamma} \{L(\gamma)\}, \quad (3)$$

where the infimum is taken over all the smooth paths from p to q . A path realizing this minimum is called a geodesic path. Geodesic paths can be seen as straight segments on \mathcal{M} . Geodesics are paths which are locally shortest paths.

Given $p \in \mathcal{M}$ with $\phi(x) = p$, the set of vectors $u \in \mathbb{R}^n$ attached to x is noted $T_p\mathcal{M}$ and is called the tangent space at p . Unlike the case of manifolds where the tangent space is defined as an equivalence class isomorphic to \mathbb{R}^n , we simply have here $T_p\mathcal{M} = \mathbb{R}^n$.

It can be shown that given $a < 0$, $0 < b$, vector $u \in \mathbb{R}^n$, there is only one geodesic $\gamma, [a, b] \rightarrow \mathcal{M}$ such that $\gamma(0) = p$ with tangent vector $(\phi^{-1} \circ \gamma)'(\gamma^{-1}(p)) = u$ such that $\|(\phi^{-1} \circ \gamma)'(t)\|_{(\phi^{-1} \circ \gamma)(t)} = \|u\|_x$ for all $t \in]a, b[$. It can be then shown that there exists a unique geodesic such that $\gamma(0) = p$, $(\phi^{-1} \circ \gamma)'(\gamma^{-1}(p)) = u$ and that its domain can not be extended. This geodesic is noted $geod_p(u)$. The exponential map, see Fig. 3, is then defined by

$$exp_p : \begin{cases} T_p\mathcal{M} & \rightarrow \mathcal{M} \\ u & \mapsto geod_p\left(\frac{1}{\|u\|_x}u\right)(\|u\|_x) \end{cases}$$

Note that this exponential has generally no link with the classical exponential application. However, they happen to coincide in specific cases.

The radius of injectivity r_p^{inj} of exp_p is the largest r such that exp_p restricted to the centered ball of radius r is injective. Inside this ball, $log_p(\cdot)$ denotes the inverse of $exp_p(\cdot)$.

Let \mathcal{S} be a subset of \mathcal{M} homeomorphic to an open subset of \mathbb{R}^m with $m \leq n$. The Riemannian metric on \mathcal{M} naturally induces a Riemannian

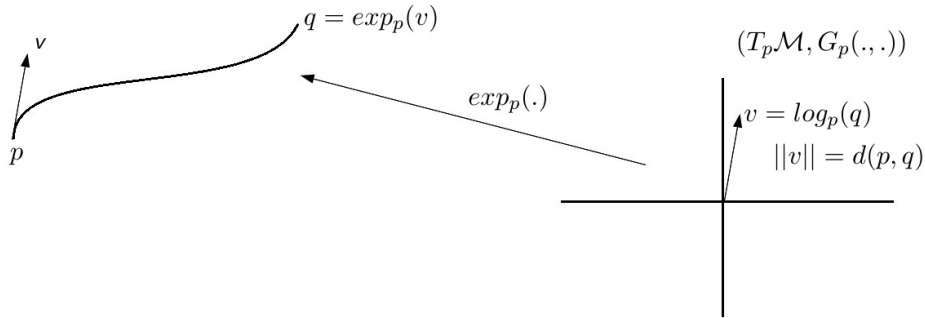


Figure 3: Exponential map.

metric on \mathcal{S} . The metric on \mathcal{S} is called the induced metric. In the expression of the geodesic distance $d_{\mathcal{S}}(p, q) = \inf_{\gamma} \{L(\gamma)\}$, the infimum is then taken over all the paths γ that stay in \mathcal{S} .

A Riemannian metric also induces a measure. The matrix of the metric $G(x)$ is symmetric and can be written $G(x) = AA^t$ where A is a $n \times n$ invertible matrix and represents a local linear change of coordinates that induces the new scalar product $G(x)$. The modification of a unitary volume associated with the linear change of coordinates is given by

$$\det(A) = \sqrt{\det(G(x))}. \quad (4)$$

The density of the Riemannian measure with respect to the Lebesgue measure of U_{ϕ} is given by $x \in U_{\phi} \mapsto \sqrt{\det(G(x))}$.

3 Non-parametric density estimation on Riemannian spaces

Let Ω be a space, endowed with a σ -algebra and a probability measure p . Let X be a random variable $\Omega \rightarrow \mathcal{M}$ with \mathcal{M} homeomorphic to U an open subset of \mathbb{R}^n . The homeomorphism is noted ϕ . \mathcal{M} is equipped with a Riemannian metric G . The Riemannian measure is called vol and the measure on \mathcal{M} induced by X is noted μ_X . We assume that μ has a density, noted f , with respect to vol , and that the support of X is a compact set noted $Supp$. Let $(x_1, \dots, x_k) \in \mathcal{M}^k$ be a set of draws of X .

The Dirac measure in point a is defined as:

$$\delta_a(U) : \begin{cases} 1 & \text{if } a \in U \\ 0 & \text{if } a \in \mathbb{C}U \end{cases}$$

Let $\mu_k = \frac{1}{k} \sum_i \delta_{x_i}$ denotes the empirical measure of the set of draws. This section presents three non-parametric techniques of estimation of f from the set of draws (x_1, \dots, x_k) . The estimated density in x in \mathcal{M} is noted $\hat{f}_k(x) = \hat{f}(x, x_1, \dots, x_k)$. Observe that $\hat{f}_k(x)$ can be seen as a random variable. The relevance of density estimation technique depends on several aspects. When the space allows it, the estimation technique should not privilege specific directions or locations. This results in an isotropy and a homogeneity condition. In the kernel method for instance, a kernel function K_{x_i} is placed at each observation x_i . Firstly, in order to treat directions equally, the function K_{x_i} should be invariant by the isotropy group of x_i . Secondly, for an other observation x_j , functions K_{x_i} and K_{x_j} should be similar up to the isometries that send x_i on x_j . These considerations strongly depend on the geometry of the space: if the space is not homogeneous and the isotropy group is empty, these indifference principles have no meaning. The convergence of the different estimation techniques is widely studied. Results were first obtained in the Euclidean case, and are gradually extended to the probability densities on manifold, see [11, 19, 13, 4]. The last relevant aspect, is computational. Each estimation technique has its own computational framework, which presents pro and cons given the different applications. For instance, the estimation by orthogonal series presents an initial pre-processing, but provides a fast evaluation of the estimated density in compact manifolds.

3.1 Estimation from the Euclidean structure of the parametrization

Assume that the term $\sqrt{\det(G(x))}$ known for all $x \in U$. According to section 2.2,

$$\frac{dvol}{dLeb}(x) = \sqrt{\det(G(x))}.$$

If we dispose of an estimator \hat{f}^{Eucl} of the density with respect to the Lebesgue measure of U , noted Leb_U , it is possible to obtain an estimator of the density with respect to the Riemannian measure vol :

$$\hat{f} = \hat{f}^{Eucl} \frac{dLeb_U}{dvol} = \frac{1}{\sqrt{\det(G(x))}} \hat{f}^{Eucl}. \quad (5)$$

The estimation \hat{f} is a probability density with respect to the Riemannian measure vol . However the estimation does not respect homogeneity and isotropy considerations. If \hat{f}^{Eucl} is constructed from an Euclidean kernel K , the ‘‘Riemannian shape’’ of K will differ from x_i to x_j in an uncontrolled way.

3.2 Histograms

Let $(A_j^k)_{0 \leq j \leq J_k}$ be a partition of U . Let α_j be the number of draws that lays in A_j^k , i.e., $\alpha_j = \text{Card}\{x_i \in A_j^k\}$. The histogram of the draws (x_i) is the function

$$\hat{f}_k = \frac{1}{k} \sum_j \frac{1}{\text{vol}(A_j^k)} \alpha_j \mathbf{1}_{A_j^k}, \quad (6)$$

where $\mathbf{1}_E$ is the indicator function of E . \hat{f}_k is an estimator of the density f .

Compared with other methods, the main advantage of histograms is often the low algorithmic complexity and the simplicity of use. In absence of prior information on the density, a good tiling for density estimation is a tiling which makes the estimation as invariant as possible under isometric transformations. Thus bins should have similar shapes, and should be as spherical as possible. Furthermore, it has to be possible to rescale the tiling in order to adapt the size of the bins to the number of data available for the estimation. In \mathbb{R}^2 , tiling the space with squares is reasonable solution, often used in practice. In Riemannian spaces, it is often difficult to find appropriate tilings of the space. Furthermore, when such tiling exists, they are not always rescalable. The hyperbolic space has many possible regular tilings, see Fig. 3.2. However there are no homothetic transforms of the hyperbolic space for every scaling factor $\lambda \in \mathbb{R}$. This means that changing the size of the bins often implies changing the tiling itself.

Except Euclidean spaces, none of the studied spaces in this paper have regular and scalable tilings. Thus histograms have not been investigated.

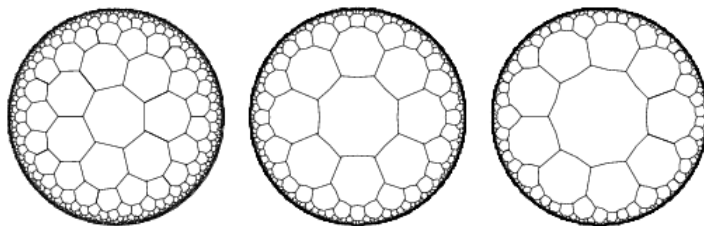


Figure 4: Hyperbolic tilings

3.3 Orthogonal series density estimator

Instead of directly studying the density f , the estimation is made from the estimation of the scalar product between f and a set of orthonormal

functions $\{e_j\}$. Recall that for g and h in $L^2(\mathcal{M})$:

$$\langle g, h \rangle = \int_{\mathcal{M}} g \bar{h} d\text{vol},$$

where \bar{x} denotes the complex conjugate of x . We have that:

$$\langle f, e_i \rangle = \int f \bar{e}_i d\text{vol} = \mathbb{E}(\bar{e}_i(X)).$$

Thus $\langle f, e_j \rangle$ can be approximated by an estimation of the expectation, i.e.,

$$\mathbb{E}(e_j(X)) \approx \frac{1}{k} \sum_{j=1}^k e_j(x_i)$$

Or in other words:

$$\int e_j f d\text{vol} \approx \int_{\text{Supp}} e_j d\mu_k.$$

Now, given an integer N , the orthogonal series estimator is defined as:

$$\hat{f} = \frac{1}{k} \sum_{j=-N}^N \left[\sum_{i=1}^k \bar{e}_j(x_i) \right] e_j. \quad (7)$$

The formula becomes an integral when the number of functions e_j is uncountable. First the base has to be ordered such that the norm of the rest of the decomposition, i.e., $\sum_{|j|>N} \langle f, e_j \rangle e_j$, decreases as fast as possible for regular functions. Second, in order to process locations and directions indifferently, the basis functions must present regularity properties regarding the metric. Eigenfunctions of the Laplacian operator fill all the criteria. Indeed the Fourier transform on \mathbb{R}^n , or the Fourier–Helgason transform on symmetric spaces, are highly related to the convolution by isotropic kernel. The orthogonal series density estimator is then equivalent to the kernel density estimator. When the e_j functions are Fourier functions, the estimation technique is also called the characteristic function method. In \mathbb{R} , the Fourier basis is uncountable, which gives place to a Fourier transform. However on $[a, b] \in \mathbb{R}$, the Fourier basis is countable, and gives place to the Fourier series. In the context of probability density estimation, it preferable for computational reasons to work with series rather than with integrals. The main issue is that in most Riemannian spaces, there are no explicit expressions of the eigenfunctions of the Laplacian operator. In several spaces, the eigenfunctions are known, but not on compact sub-domains. The space of univariate Gaussian distributions under the Fisher metric is isometric to the

2-dimensional hyperbolic space \mathcal{H} . On the hyperbolic space, eigenfunction of the Laplacian are known when they are defined on the entire space:

$$f : \mathcal{H} \rightarrow \mathbb{R}, \Delta f = \lambda f$$

but not for functions defined on compact domains $f : D \in \mathcal{H} \rightarrow \mathbb{R}$. The estimation of a density involves an integral, thus the kernel density estimator is preferred.

In each Riemannian space studied in this paper, the eigenfunctions of the Laplacian are either not known, or only known on non-compact domains.

3.4 Kernel density estimator

Let $K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a map which verifies the following properties:

- i) $\int_{\mathbb{R}^d} K(\|x\|)dx = 1$,
- ii) $\int_{\mathbb{R}^d} xK(\|x\|)dx = 0$,
- iii) $K(x > 1) = 0$, $\sup(K(x)) = K(0)$.

Let $p \in \mathcal{M}$ and $x \in U$ with $\phi(x) = p$. In the general case, given a point $p \in \mathcal{M}$, \exp_p defines an injective application only on a neighborhood of $0 \in T_p\mathcal{M}$. However in every studied situations of the present work, the parametrization induced by \exp_p is injective on the whole space. Recall that when $T_p\mathcal{M}$ is endowed with the scalar product $G(x)$, Euclidean distances to 0 corresponds to Riemannian distance to p . The Lebesgue measure of $T_p\mathcal{M}$ is noted Leb_p . The function θ_p defined by:

$$\theta_p : q \mapsto \theta_p(q) = \frac{dvol}{d\exp^*(Leb_p)}(q), \quad (8)$$

is the density of the Riemannian measure of \mathcal{M} with respect to the Lebesgue measure Leb_p after the identification of \mathcal{M} and $T_p\mathcal{M}$ induced by \exp_p , see Fig.5.

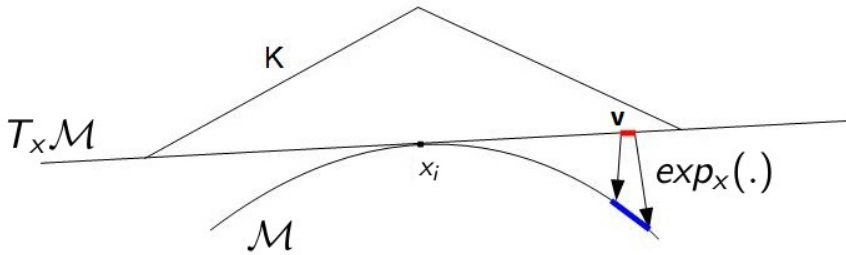


Figure 5: Volume change θ_{x_i} induced by the exponential map

Given K and a positive radius r , the estimator of f proposed by [19] is defined by:

$$\hat{f}_k = \frac{1}{k} \sum_i \frac{1}{r^n} \frac{1}{\theta_{x_i}(x)} K\left(\frac{d(x, x_i)}{r}\right). \quad (9)$$

The corrective factor $\theta_{x_i}(x)^{-1}$ is necessary since the kernel K originally integrates to one with respect to the Lebesgue measure, while we want it to integrate to one with respect to the Riemannian measure. It can be noted that this estimator is the usual kernel estimator in the case of Euclidean space.

Let $r^{inj} = \sup_{p \in \mathcal{M}}(r_p^{inj})$. Let $\delta = \sup \mathcal{K}$ the supremum of the sectional curvature in \mathcal{M} . For a definition of the sectional curvature, see. [7]. Then, [19] provides the following result.

Theorem 3.1 ([19]) *Let x_i be an arbitrary point of \mathcal{M} . Let μ be a measure on \mathcal{M} whose density with respect to the measure vol is*

$$\frac{1}{r^n} \frac{1}{\theta_{x_i}(\cdot)} K\left(\frac{d(\cdot, x_i)}{r}\right).$$

For $r < \min\{\frac{r^{inj}}{2}, \frac{\pi}{4\sqrt{\delta}}\}$ when $\delta > 0$ or $r < \frac{r^{inj}}{2}$ when $\delta \leq 0$, x_i is an intrinsic mean of μ .

Theorem. 3.1 ensures the kernel put on x_i is “centered” on x_i .

Despite convergence rates provided by [19] are obtained for compact manifolds without boundaries, they remain valid for non-compact manifolds when the support of X is compact. Indeed the double manifold, see [17], enables to see a compact manifold with boundaries as a submanifold of a compact manifold without boundaries. Assume f two-times differentiable with bounded second covariant derivative. See [18] for a definition of covariant derivative.

Theorem 3.2 ([19]) *For r satisfying conditions of theorem. 3.1, there exists a constant C_f such that:*

$$\int_{Supp} \mathbb{E}[(f(x) - \hat{f}_k(x))^2] d\mu \leq C_f \left(\frac{1}{kr^n} + r^4\right),$$

then for r equivalent to $k^{-\frac{1}{d+4}}$:

$$\int_{Supp} \mathbb{E}[(f(x) - \hat{f}_k(x))^2] d\mu = O(k^{-\frac{4}{n+4}}).$$

For the expression of the constants see details in [19].

It can be easily verified that for an isometry α we have:

$$\hat{f}_k(x, x_1, \dots, x_k) = \hat{f}_k(\alpha(x), \alpha(x_1), \dots, \alpha(x_k)).$$

Each location and direction are processed as similarly as possible. The only computational aspect that has to be taken into account is the cost of evaluation of the density at a point x in \mathcal{M} . The computational cost of $\hat{f}_k(x)$ is in $O(k)$.

Due to the important drawbacks of histograms and orthogonal series, only the kernel density estimator is investigated in the studied cases. It must be noted that the kernel method requires the existence of explicit expression of distances and of the function θ_p , which renders it unusable in several situations. One of the main contributions of this paper is the computation of the function θ_p on the space of multivariate Gaussian distributions under the Wasserstein metric, and on multivariate centered Gaussians under the Fisher metric.

The k -nearest neighbor is an interesting variant of the standard kernel density estimation. It consists in making the parameter r depend on x_i , by setting r as the distance to the k -th nearest neighbor of x_i . This estimator has been studied on Riemannian manifolds in [12].

4 Classical metrics on Gaussian laws

4.1 The Fisher metric

We consider here a set \mathcal{M} of probability measures on a measure space (X, σ_X, μ) homeomorphic to an open subset of \mathbb{R}^n . Let $\theta = (\theta_1, \dots, \theta_n) \in U \subset \mathbb{R}^n$ be an homeomorphic parametrization of \mathcal{M} . The distributions are assumed to have a density $f(\cdot, \theta)$. The Fisher matrix $G(\theta)$ associated to the parametrization θ is defined as follows, see [2]:

$$g(\theta)_{i,j} = E \left[\frac{\partial \ln(f(x, \theta))}{\partial \theta_i} \frac{\partial \ln(f(x, \theta))}{\partial \theta_j} \right] = \int_x \frac{\partial \ln(f(x, \theta))}{\partial \theta_i} \frac{\partial \ln(f(x, \theta))}{\partial \theta_j} f(x, \theta) dx. \quad (10)$$

The matrix $G(\theta)$ is symmetric positive definite. The matrices $\{G(\theta), \theta \in U\}$ induce a smooth field of positive definite matrices on U , that is to say a Riemannian metric. It can be shown that the distances and the geodesic paths induced on \mathcal{M} by this Riemannian metric do not depend on the choice

of the parametrization. It is interesting to note that the Fisher metric do not rely on a metric on X .

Let P and Q be two probability measures. The relative entropy is defined as follows:

$$D(P, Q) = \int_x \ln \left(\frac{dP}{dQ}(x) \right) dP.$$

Function $D(., .)$ is not symmetric and is therefore not a distance. The relative entropy is also called the Kullback–Leibler divergence. The Kullback–Leibler divergence measures an informational difference between P and Q .

Let $D(\theta, \theta_0)$ be the relative entropy between the laws parametrized by θ and θ_0 . It can be shown that:

$$\frac{\partial D(\theta_0, \theta)}{\partial \theta_i}(\theta_0) = 0,$$

and

$$\frac{\partial^2 D(\theta_0, \theta)}{\partial \theta_i \partial \theta_j}(\theta_0) = g(\theta_0)_{i,j}.$$

Thus, the Fisher matrix is also the Hessian of the Kullback–Leibler divergence.

4.2 The Wasserstein metric/ Earth mover’s distance

The Wasserstein metric on probability distributions is the metric of optimal transport as founded by [16], see [30] for a modern and complete mathematical overview of the problem. If the probability distributions are seen as earth heaps, the Wasserstein distance between two distributions is the minimum cost that is needed to transform the first heap into the second one. The cost of transport of earth depends of the amount of earth and the traveled distance. Due to this interpretation, the Wasserstein metric is also called the earth mover’s distance.

More formally, let X and Y be two random variables valued in a space S endowed with a σ -algebra \mathcal{A}_S . Let μ_X and μ_Y be the measures induced on S by X and Y . Let $\Gamma(\mu_X, \mu_Y)$ be the set of measures on $S \times S$ such that μ_X and μ_Y are the first and the second marginals, i.e., for $\gamma \in \Gamma(\mu_X, \mu_Y)$, one has

$$\gamma(A, S) = \mu_X(A), \quad \gamma(S, B) = \mu_Y(B),$$

for all $A, B \in \mathcal{A}_S$. Thus $\Gamma(\mu_X, \mu_Y)$ represents the set of possible transports of the measure μ_X on the measure μ_Y . The p -Wasserstein distance between X and Y is then defined by:

$$d(X, Y) = \left[\inf_{\gamma \in \Gamma(\mu_X, \mu_Y)} \int_{S \times S} d(a, b)^p d\gamma \right]^{1/p}. \quad (11)$$

For most families probability distributions, the Wasserstein distance has no explicit expression. Fortunately, the expression of the 2-Wasserstein distance between Gaussian laws has a simple form. Given two Gaussian distributions $\mathcal{N}_1, \mathcal{N}_2$ of mean m_1, m_2 and covariances Σ_1, Σ_2 the distance becomes:

$$d(\mathcal{N}_1, \mathcal{N}_2)^2 = |m_1 - m_2|^2 + \text{tr}(\Sigma_1) + \text{tr}(\Sigma_2) - 2\text{tr}(\sqrt{\Sigma_2^{1/2}\Sigma_1\Sigma_2^{1/2}}).$$

[28] has showed that the 2-Wasserstein distance on Gaussian laws is a Riemannian metric distance. We restrict the present study to the case of the 2-Wasserstein distance.

5 The space of multivariate Gaussian laws

In this section we study the space of multivariate Gaussian laws under the two classical Riemannian metrics discussed in the previous section. A Gaussian law is determined by its mean vector m and its covariance matrix Σ and is noted $\mathcal{N}_{m,\Sigma}$. It is well known that the expression of its density $f_{m,\Sigma}$ is written as:

$$f_{m,\Sigma}(x) = \frac{1}{(2\pi)^{N/2}\det(\Sigma)^{1/2}} e^{-\frac{1}{2}(x-m)^t\Sigma^{-1}(x-m)}.$$

The space of n-dimensional Gaussian laws, noted $\mathcal{G}(n)$, can be identified with the product between the vector space \mathbb{R}^n and the space of positive definite matrices $\text{SPD}(n)$. The space $\text{SPD}(n)$ is an open cone included in the vector space of symmetrical matrices $\text{Sym}(n)$. Let:

$$\phi : \begin{cases} \mathbb{R}^n \times \text{SPD}(n) & \rightarrow \mathcal{G}(n) \\ (m, \Sigma) & \mapsto \mathcal{N}_{m,\Sigma} \end{cases}$$

and μ_ϕ be the product of Lebesgue measures:

$$\mu_\phi = \text{Leb}_{\mathbb{R}^n} \times \text{Leb}_{\text{SPD}(n)}$$

where $\text{Leb}_{\text{SPD}(n)}$ stand for the Lebesgue measures on coefficients of the $\text{SPD}(n)$ matrices.

The diagonalization of elements of $\text{SPD}(n)$ give place to the following parametrization:

$$\psi : \begin{cases} \mathbb{R}^n \times \mathbb{R}^n \times \text{SO}(n) & \rightarrow \mathbb{R}^n \times \text{SPD}(n) \\ (m, \lambda, R) & \mapsto (m, RD_\lambda R^t) \end{cases}$$

where $D_\lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$. Let

$$\mu_\psi = Leb_{\mathbb{R}^n} \times Leb_{\mathbb{R}^n} \times Haar_{SO(n)}$$

where $Haar_{SO(n)}$ stands for the Haar measure on rotations. The volume change $f_{\psi \rightarrow \phi}$ induced by ψ :

$$\mu_\phi(\psi(A)) = \int_A f_{\psi \rightarrow \phi} d\mu_\psi$$

has the following expression:

$$f_{\psi \rightarrow \phi}(m, \lambda, R) = \prod_{i < j \leq n} |\lambda_i - \lambda_j|, \quad (12)$$

This result can be found in Example 8.7, page 158, of [31].

5.1 The Fisher metric

Unlike multivariate centered and univariate Gaussian laws, the space of multivariate Gaussian laws under the Fisher metric is not a symmetric space. As in the large majority of the Riemannian manifolds, there is no explicit expression of distances. We only have access to the expression of the metric. An attempt was made by authors of [15] to symmetrize the space.

Since the expression of distances between arbitrary distributions is not known, the exponential map can not be computed. Thus, it is not possible to compute the volume change of Eq. 8 involved in the expression of kernels. It is however possible to compute the density of the volume measure induced by the Fisher metric with respect to μ_ϕ , which enables the use of Eq. 5.

Let u and v be vectors of $\mathbb{R}^n \oplus Sym(n)$. Let index m and Σ denote the components associated to \mathbb{R}^n and $Sym(n)$ respectively.

The scalar product of the Fisher metric at (m, Σ) can be rewritten, see [25], as

$$\langle u, v \rangle_{m, \Sigma}^{Fisher} = u_m^t \Sigma^{-1} v_m + \frac{1}{2} tr(\Sigma^{-1} u_\Sigma \Sigma^{-1} v_\Sigma). \quad (13)$$

The means and the covariances are orthogonal. Thus the expression of the volume measure associated to the Fisher metric is a product between a measure on the mean and a measure on the covariance.

$$\frac{dvol_{Fisher}}{d\mu_\phi} = f_m \cdot f_\Sigma$$

The measure on the mean is $f_m = \sqrt{\det(\Sigma^{-1})}$. Up to a normalizing factor, the measure on the covariance part can be found here,[29]:

$$f_\Sigma = 2^{-n^2} \det(\Sigma)^{-(n+1)/2} \quad (14)$$

Thus,

$$\frac{dvol_{Fisher}}{d\mu_\phi} = 2^{-n^2} \det(\Sigma)^{-(n+2)/2} = 2^{-n^2} \sqrt{\prod_i \frac{1}{\lambda_i^{n+2}}}. \quad (15)$$

In the parametrization induced by ψ ,

$$\begin{aligned} dvol_{Fisher} &= \frac{dvol_{Fisher}}{d\mu_\phi} \frac{d\mu_\phi}{d\mu_\psi} d\mu_\psi \\ &= 2^{-n^2} \sqrt{\prod_i \frac{1}{\lambda_i^{n+2}}} \prod_{i < j \leq n} |\lambda_i - \lambda_j| dLeb_{\mathbb{R}^n} dLeb_{\mathbb{R}_+^n} dHaar_{SO(n)}. \end{aligned} \quad (16)$$

There is no way of estimating densities based on a fully Riemannian approach. The only option here is to perform a density estimation in an Euclidean context, and to adapt the result to the Riemannian measure. If X is seen as a random variable in $\mathcal{G}(n)$ the estimation gives:

$$\hat{f}^{Fisher} = \hat{f}^{Eucl} \frac{d\mu_\phi}{dvol} = 2^{\frac{n(n-1)}{2}} \sqrt{\prod_i \lambda_i^{n+2}} \hat{f}^{Eucl}, \quad (17)$$

where \hat{f}^{Eucl} is an estimation of the density of X with respect to the measure induced by the scalar product $\langle A, B \rangle = tr(AB^t)$. However the estimation does not respect homogeneity and isotropy considerations.

5.2 The Wasserstein metric

The 2-Wasserstein metric is a product metric between the space of means and the space of covariance matrices:

$$d_{W_2}(\mathcal{N}_{m_1, \Sigma_1}, \mathcal{N}_{m_2, \Sigma_2})^2 = d_m(m_1, m_2)^2 + d_\Sigma(\Sigma_1, \Sigma_2)^2,$$

with

$$\begin{aligned} d_m(m_1, m_2) &= |m_1 - m_2|, \\ d_\Sigma(\Sigma_1, \Sigma_2)^2 &= tr(\Sigma_1) + tr(\Sigma_2) - 2tr(\sqrt{\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2}}). \end{aligned}$$

It was proved in [28] that this distance is induced by a Riemannian metric.

Unlike the case of the Fisher metric on multivariate Gaussian distributions, it is possible to obtain an explicit expression the volume change of Eq. 8 and thus to have an explicit expression of kernels.

Let vol_W be the measure associated with such Riemannian metric.

5.2.1 The Riemannian volume density with respect to μ_ϕ

Let u and v be vectors of $\mathbb{R}^n \oplus Sym(n)$. Let index m and Σ denote the components associated to \mathbb{R}^n and $Sym(n)$ respectively. The scalar product at (m, Σ) according to [28] is given by

$$\langle u, v \rangle_{m, \Sigma}^W = u_m^t v_m + tr(\tilde{u}_\Sigma \Sigma \tilde{v}_\Sigma),$$

with

$$\begin{aligned} u_\Sigma &= \Sigma \tilde{u}_\Sigma + \tilde{u}_\Sigma \Sigma, \\ v_\Sigma &= \Sigma \tilde{v}_\Sigma + \tilde{v}_\Sigma \Sigma. \end{aligned} \tag{18}$$

Let $G_{\mathcal{N}_{m, \Sigma}}^W$ be the matrix of the scalar product in the canonical basis of $\mathbb{R}^n \oplus Sym(n)$. The density of the Wasserstein volume measure with respect to the Lebesgue measure of $\mathbb{R}^n \times SPD(n)$ is given by

$$\frac{dvol_W}{d\mu_\phi} = \sqrt{\det(G_{\mathcal{N}_{m, \Sigma}}^W)}.$$

The invariance of the Wasserstein metric under the action of rotations implies the invariance of the measure, i.e.,

$$\sqrt{\det(G_{\mathcal{N}_{m, RDR^t}}^W)} = \sqrt{\det(G_{\mathcal{N}_{m, D}}^W)}.$$

Thus the volume change only need to be computed for the case diagonal matrices. Let $D \in SPD(n)$ be a diagonal matrix of eigenvalues λ_i . Equations. (18) can now be solved. Since,

$$\begin{aligned} 2\lambda_i E'_{i,i} &= DE'_{i,i} + E'_{i,i}D, \\ (\lambda_i + \lambda_j)E'_{i,j} &= DE'_{i,j} + E'_{i,j}D, \end{aligned}$$

equations (18) are equivalent to

$$\begin{aligned} u_\Sigma &= P(\tilde{u}_\Sigma), \\ v_\Sigma &= P(\tilde{v}_\Sigma), \end{aligned}$$

where P is a linear operator of $Sym(n)$ whose matrix form is

$$P: \begin{pmatrix} & E'_{i,i} & E'_{i,j} & & \\ & & & & \\ 2\lambda_1 & & & & \\ & \ddots & & & \\ & & 2\lambda_n & & \\ & & & (\lambda_i + \lambda_j) & \end{pmatrix}$$

The different terms of the metric are given by:

$$\forall (i, j \leq k):$$

$$\langle e_{m_i}, E'_{j,k} \rangle_{\mathcal{N}_{m,D}}^W = 0,$$

$$\forall (i, j):$$

$$\langle e_{m_i}, e_{m_j} \rangle_{\mathcal{N}_{m,D}}^W = 1,$$

$$\forall (i, j):$$

$$\langle E'_{i,i}, E'_{j,j} \rangle_{\mathcal{N}_{m,D}}^W = \delta_{i,j} \frac{1}{4\lambda_i},$$

where again $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$,

$$\forall (i, j \neq k):$$

$$\langle E'_{i,i}, E'_{j,k} \rangle_{\mathcal{N}_{m,D}}^W = 0,$$

$$\forall i < j, k < l:$$

$$\langle E'_{i,j}, E'_{k,l} \rangle_{\mathcal{N}_{m,D}}^W = \delta_{i,k} \delta_{j,l} \frac{1}{2(\lambda_i + \lambda_j)}.$$

Thus, we obtain

$$G_{\mathcal{N}_{m,D}}^W = \begin{pmatrix} & e_{m_i} & E'_{i,i} & E'_{i,j} & & \\ & & & & & \\ 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \frac{1}{4\lambda_1} & & \\ & & & & \ddots & \\ & & & & & \frac{1}{4\lambda_n} \\ & & & & & & \frac{1}{2(\lambda_i + \lambda_j)} \end{pmatrix}$$

We can then compute the volume density,

$$\frac{dvol_W}{d\mu_\phi} = \sqrt{\det(G_{\mathcal{N}_{m,\Sigma}}^W)} = \frac{1}{2^{\frac{1}{2}n^2 + \frac{3}{2}n}} \sqrt{\prod_i \frac{1}{\lambda_i} \prod_{i < j} \frac{1}{(\lambda_i + \lambda_j)}}, \quad (19)$$

and in the parametrization induced by ϕ ,

$$\begin{aligned} dvol_W &= \frac{dvol_W}{d\mu_\phi} \frac{d\mu_\phi}{d\mu_\psi} d\mu_\psi \\ &= \frac{1}{2^{\frac{1}{2}n^2 + \frac{3}{2}n}} \prod_i \frac{1}{\sqrt{\lambda_i}} \prod_{i < j} \frac{|\lambda_i - \lambda_j|}{\sqrt{(\lambda_i + \lambda_j)}} dLeb_{\mathbb{R}^n} dLeb_{\mathbb{R}_+^n} dHaar_{SO(n)}. \end{aligned} \tag{20}$$

5.2.2 The Riemannian volume density in the exponential parametrization

We are interested here in the computation of the following quantity:

$$\theta_{\Sigma_1} : \Sigma_2 \mapsto \theta_{\Sigma_1}(\Sigma_2) = \frac{dvol_W}{dexp^*(Leb_{\Sigma_1})}(\Sigma_2),$$

that is to say the density of the Riemannian measure of $\mathcal{G}(n)$ with respect to the Lebesgue measure of the parametrization of $\mathcal{G}(n)$ induced by exp_{Σ_1} . As we have stated, Wasserstein metric is a product between the Euclidean metric on \mathbb{R}^n and a Riemannian metric on $SPD(n)$. The Euclidean part can be omitted in this section, since it does not affect the volume change.

It has been shown in [28] that the L^2 -Wasserstein distance on centered Gaussian measure is induced by a Riemannian metric on $SPD(n)$. Furthermore, the application

$$\Pi : \begin{cases} Gl_n & \rightarrow & SPD(n) \\ G & \mapsto & GG^t \end{cases}$$

is a Riemannian submersion when GL_n is endowed with the scalar product $\langle A, B \rangle = tr(AB^t)$ and $SPD(n)$ is embedded with the Riemannian metric. For $A \in Gl(n)$ the kernel of $d\Pi$ is called the vertical space, and its orthogonal the horizontal space. For $\Sigma = RDR^t \in SPD(n)$, let $\Sigma^{1/2} = RD_{(\sqrt{\lambda_i})}R^t$, such that $\Pi(\Sigma^{1/2}) = \Sigma$. The horizontal space at $\Sigma^{1/2}$, see [28], is given by:

$$\mathcal{H}_{\Sigma^{1/2}} = \{X\Sigma^{1/2}, X \in Sym\}.$$

$\mathcal{H}_{\Sigma^{1/2}}$ can be identified with $T_\Sigma M$ via $d\Pi$. Let $\Pi_A(Z) = \Pi(A + Z)$. For simplicity reasons, the exponential map will be expressed in $\mathcal{H}_{\Sigma^{1/2}}$. For $Z \in \mathcal{H}_{\Sigma^{1/2}}$ classical results on Riemannian submersions give that the image by Π of the geodesic $\Sigma^{1/2} + tZ$ is the geodesic $exp_\Sigma(tZ)$. Hence, one has

$$exp_\Sigma(Z) = \Pi_{\Sigma^{1/2}}(Z).$$

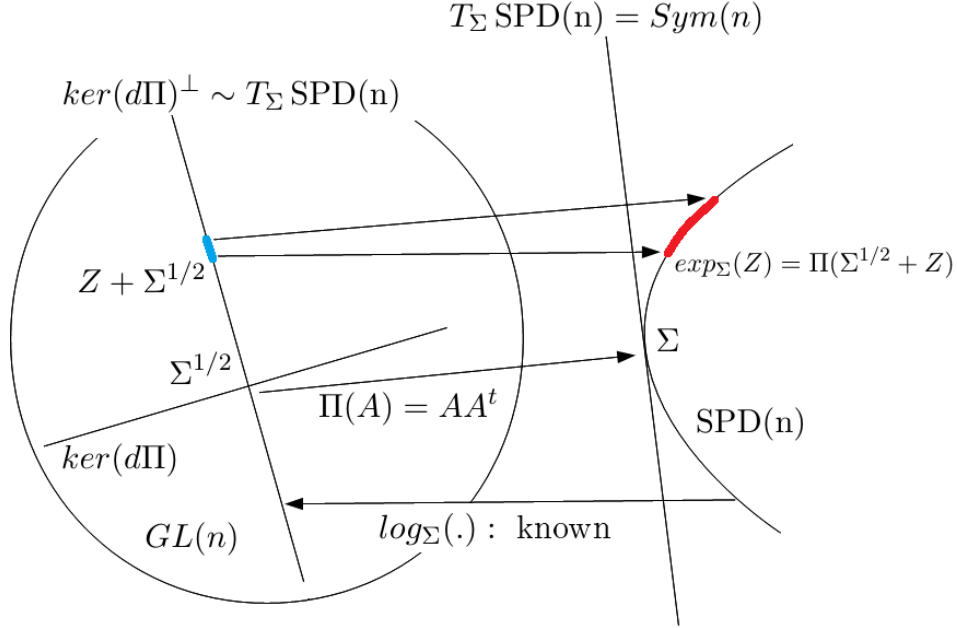


Figure 6: The application Π is a Riemannian submersion for the flat metric on $GL(n)$ and the Wasserstein metric on $SPD(n)$. The volume change between the blue and the red area appears in the expression of kernels.

Fig. 6 sums up the situation. Let $\tilde{\Pi}_A$ be the restriction of Π_A to \mathcal{H}_A . The density of interest is a product of two factors, i.e.,

$$\frac{dvol_W}{dexp^*(Leb_{\mathcal{H}_{\Sigma^{1/2}}})} = \frac{dvol_W}{dLeb_{SPD(n)}} \frac{dLeb_{SPD(n)}}{dexp^*(Leb_{\mathcal{H}_{\Sigma^{1/2}}})},$$

where $\frac{dvol_W}{dLeb_{SPD(n)}}$ was computed in Eq. (19). The computation of $\frac{dLeb_{SPD(n)}}{dexp^*(Leb_{\mathcal{H}_{\Sigma^{1/2}}})}$ can be made through the Jacobian of the exponential application. The computation of the differential of $\tilde{\Pi}_A$ gives:

$$\begin{aligned} d\tilde{\Pi}_A(XA, X \in Sym(n))(YA, Y \in Sym(n)) &= (A + XA)(YA)^t + YA(A + XA)^t \\ &= AA^tY + XAA^tY + YAA^t + YAA^tX \\ &= AA^tY + YAA^t + XAA^tY + YAA^tX. \end{aligned}$$

The invariance of the Wasserstein metric under the action of rotations enables to restrict the computation of θ_{Σ_1} to θ_D with D a diagonal matrix.

So,

$$\begin{aligned} d\tilde{\Pi}_{D^{1/2}}(XD^{1/2})(YD^{1/2}) &= DY + YD + XDY + (XDY)^t \\ &= (I + X)DY + Y((I + X)D)^t. \end{aligned}$$

The following result enables the computation of the volume change of the application $Y \mapsto d\tilde{\Pi}_{D^{1/2}}(XD^{1/2})(YD^{1/2})$.

Theorem 5.1 *Let A be a square matrix of order n and*

$$S_A : \begin{cases} \text{Sym}(n) & \rightarrow \text{Sym}(n) \\ Y & \mapsto YA + A^tY \end{cases}$$

Then

$$\det(S_A) = \prod_{i \leq j} (\lambda_i + \lambda_j) = 2^n \det(A) \prod_{i < j} (\lambda_i + \lambda_j),$$

where the λ_i are the complex eigenvalues of A . Furthermore, $\det(S_A)$ depends only on the coefficients of the characteristic polynomial of A

$$\chi_A(\lambda) = \det(\lambda I - A) = \prod_{i=1}^n (\lambda - \lambda_i) = \lambda^n + \prod_{k=1}^n (-1)^k \sigma_{n-k} \lambda^k.$$

1. If $n = 2$, $\det(S_A) = 4(\det A)(\text{trace} A)$;
2. If $n = 3$, $\det(S_A) = 2^3 \det A (\sigma_1 \sigma_2 - \sigma_3)$;
3. If $n = 4$, $\det(S_A) = 2^4 \det A (\sigma_1 \sigma_2 - \sigma_3 - \sigma_4 \sigma_1^2 - \sigma_3^2)$;
4. If $n \geq 4$, $\det(S_A)$ can be obtained with the help of standard libraries of mathematical programming languages using the decomposition on the elementary symmetric polynomials.

Proof 1. Suppose A is diagonalisable and $(X_i)_{1 \leq i \leq n}$ is a basis of eigenvectors of A . Then

$$S_A(X_i X_j^t + X_j X_i^t) = (\lambda_i + \lambda_j)(X_i X_j^t + X_j X_i^t),$$

and $(X_i X_j^t + X_j X_i^t)_{1 \leq i \leq j \leq n}$ is a basis of $\text{Sym}(n)$. Consequently

$$\det(S_A) = \prod_{i \leq j} (\lambda_i + \lambda_j) = 2^n \det(A) \prod_{i < j} (\lambda_i + \lambda_j)$$

2. $\det(S_A)$ is a symmetric polynomial of the λ_i . The fundamental theorem of symmetric polynomial states that any symmetric polynomial can be expressed as a polynomial function of the elementary symmetric polynomials. Since the coefficients of the characteristic polynomial of A are the elementary symmetric polynomials in λ_i , $\det(S_A)$ is a polynomial function of the coefficient of the characteristic polynomial of A .
3. In the general case, any matrix A can be seen as a limit of diagonalizable matrices A_n . By continuity we have $\det(S_A) = \lim \det(S_{A_n})$ and the coefficients of the polynomial χ_A are the limits of the coefficients of the polynomials χ_{A_n} . Therefore the formula can be extended to any matrix.

The basis $\mathcal{B} = (E'_{i \leq j} D^{1/2})$ of $\mathcal{H}_{D^{1/2}}$ is orthogonal for the scalar product $\langle A, B \rangle = \text{tr}(AB^t)$, with $\langle E'_{ii} D^{1/2}, E'_{ii} D^{1/2} \rangle = \lambda_i$ and $\langle E'_{i < j} D^{1/2}, E'_{i < j} D^{1/2} \rangle = \frac{\lambda_i + \lambda_j}{2}$. Thus in an orthonormal basis of $\mathcal{H}_{D^{1/2}}$, we have

$$\det(\mathcal{B}) = \sqrt{\prod \lambda_i \prod_{i < j} \frac{\lambda_i + \lambda_j}{2}}.$$

Let consider the map

$$f : \begin{cases} \mathcal{H}_{D^{1/2}} & \rightarrow \text{Sym}(n) \\ Y D^{1/2} & \mapsto Y \end{cases}$$

Expressed in orthonormal basis, $\det(f) = \det(\mathcal{B})^{-1}$. Since $d\tilde{\Pi}_{D^{1/2}}(X D^{1/2}) = S_{(I+X)D} \circ f$, the determinant of $d\tilde{\Pi}_{D^{1/2}}$ expressed in orthonormal basis is given by

$$\det(d\tilde{\Pi}_{D^{1/2}}) = \det(S_{(I+X)D}) \det(f) = \frac{\det(S_{(I+X)D})}{\det(\mathcal{B})}.$$

Hence the expression of the volume change is given by,

$$\frac{d\text{Leb}_{\text{SPD}(n)}}{d\text{exp}^*(\text{Leb}_{\mathcal{H}_{\Sigma^{1/2}}})} = \det(d\tilde{\Pi}_{D^{1/2}}) = \frac{\det(S_{(I+X)D})}{\det(\mathcal{B})}. \quad (21)$$

Given two matrices $\Sigma_0, \Sigma_1 \in \text{SPD}(n)$, [28] provides the expression of the vector $Z \in \mathcal{H}_{\Sigma_0^{1/2}}$:

$$Z = (\Sigma_1^{1/2} (\Sigma_1^{1/2} \Sigma_0 \Sigma_1^{1/2})^{-1/2} \Sigma_1^{1/2} - I) \Sigma_0^{1/2},$$

such that $\Pi_{\Sigma_0^{1/2}}(Z) = \Sigma_1$. Thus,

$$\begin{aligned}\theta_D(\Sigma) &= \frac{dvol_W}{dLeb_{\text{SPD}(n)}} \frac{dLeb_{\text{SPD}(n)}}{dexp^*(Leb_{\mathcal{H}_{D^{1/2}}})} \\ &= \frac{1}{2^{n(n+1)}} \frac{det(S_A)}{\sqrt{\prod \lambda_i^\Sigma \prod_{i<j} (\lambda_i^\Sigma + \lambda_j^\Sigma)} \sqrt{\prod \lambda_i^D \prod_{i<j} (\lambda_i^D + \lambda_j^D)}},\end{aligned}$$

where (λ_i^D) and (λ_i^Σ) are the eigenvalues of D and Σ respectively, and where

$$A = \Sigma^{1/2}(\Sigma^{1/2}D\Sigma^{1/2})^{-1/2}\Sigma^{1/2}D^{1/2}.$$

We recall that A corresponds to the matrix $(I + X)D$ of Eq. (21).

Finally, using the invariance of the Wasserstein metric under the action of rotations, we have for $\Sigma_0 = RDR^t$:

$$\theta_{\Sigma_0}(\Sigma_1) = \frac{1}{2^{n(n+1)}} \frac{det(S_A)}{\sqrt{\prod \lambda_i^{\Sigma_1} \prod_{i<j} (\lambda_i^{\Sigma_1} + \lambda_j^{\Sigma_1})} \sqrt{\prod \lambda_i^{\Sigma_0} \prod_{i<j} (\lambda_i^{\Sigma_0} + \lambda_j^{\Sigma_0})}}, \quad (22)$$

where $(\lambda_i^{\Sigma_0})$ and $(\lambda_i^{\Sigma_1})$ are the eigenvalues of Σ_0 and $R^t\Sigma_1R$ respectively,

$$A = (R^t\Sigma_1R)^{1/2}((R^t\Sigma_1R)^{1/2}D(R^t\Sigma_1R)^{1/2})^{-1/2}(R^t\Sigma_1R)^{1/2}D^{1/2},$$

and $det(S_A)$ is given in Theorem 5.1. Given a set of Gaussian distributions $x_i = (m_i, \Sigma_i)$ and a scaling parameter $r > 0$, the kernel density estimator becomes:

$$\hat{f}_k^r((m, \Sigma)) = \frac{1}{k} \sum_i \frac{1}{r^n} \frac{1}{\theta_{\Sigma_i}(\Sigma)} K\left(\frac{d_{W_2}((m, \Sigma), x_i)}{r}\right). \quad (23)$$

6 The space of multivariate centred Gaussian laws

We consider specifically the study of multivariate Gaussian laws, after adding the property $m = 0$. The set of multivariate centered Gaussian laws is noted $\mathcal{G}_{m=0}(n)$. In $\mathcal{G}_{m=0}(n)$, a Gaussian law is determined by its covariance matrix Σ only. Its elements are noted \mathcal{N}_Σ . The expression of its density f_Σ is:

$$f_{m,\Sigma}(x) = \frac{1}{(2\pi)^{N/2} det(\Sigma)^{1/2}} e^{-\frac{1}{2}(x)^t \Sigma^{-1}(x)}.$$

Technical considerations provided in section 5 remain valid after removal of the mean. In particular, let us redefine ϕ, ψ, μ_ϕ and μ_ψ :

$$\phi : \begin{cases} \text{SPD}(n) & \rightarrow \mathcal{G}(n) \\ \Sigma & \mapsto \mathcal{N}_\Sigma \end{cases}$$

$$\begin{aligned} \mu_\phi &= Leb_{SPD(n)} \\ \psi : \begin{cases} \mathbb{R}^n \times SO(n) & \rightarrow SPD(n) \\ (\lambda, R) & \mapsto RD_\lambda R^t \end{cases} \\ Leb_{\mathbb{R}^n} \times \mu_\psi &= Haar_{SO(n)}. \end{aligned}$$

The different expressions associated to the Wasserstein metric computed in section 5.2 do not change for the case of centered Gaussian laws. Furthermore, space $\mathcal{G}(n)$ does not gain any property after imposing $m = 0$ in the Wasserstein metric. Thus this section focuses on the Fisher metric.

The Fisher metric on $\mathcal{G}_{m=0}(n)$ happens to be the metric induced by the Fisher metric on $\mathcal{G}(n)$. On the contrary of $\mathcal{G}(n)$, $\mathcal{G}_{m=0}(n)$ is symmetric and is then easier to study. Let $GL(n)$ be the set of invertible matrices of size n . For any G in $GL(n)$, the application $\Sigma \mapsto G\Sigma G^t$ is an isometry of $\mathcal{G}_{m=0}(n)$. The action of $GL(n)$ on $\mathcal{G}_{m=0}(n)$ induced by this application is transitive. Indeed, for any $\Sigma \in SPD(n)$, let $\Sigma^{1/2} \in GL(n)$. The orbit of the identity I is $SPD(n)$: $\Sigma = \Sigma^{1/2} I \Sigma^{1/2}$. The stabilizer of I are the orthogonal matrices. Thus $\mathcal{G}_{m=0}(n)$ can be identified to $GL(n)/O(n)$.

There is an explicit expression of the distance between two laws [21]:

$$d_{Fisher}(\mathcal{N}_{\Sigma_1}, \mathcal{N}_{\Sigma_2}) = \|\log(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2})\|,$$

where $\|\cdot\|$ is the norm associated with the scalar product $tr(AB^t)$.

6.1 The Riemannian volume density with respect to μ_ϕ

For the centered case Eq. 13 simply becomes:

$$\langle u, v \rangle_{m, \Sigma}^{Fisher} = \frac{1}{2} tr(\Sigma^{-1} u_\Sigma \Sigma^{-1} v_\Sigma). \quad (24)$$

The expression of the volume measure is well known, see [29] and is similar to Eq. 14:

$$\frac{dvol_{Fisher}}{d\mu_\phi} = 2^{-n^2} \sqrt{\prod_i \frac{1}{\lambda_i^{n+1}}} = 2^{-n^2} det(\Sigma)^{-(n+1)/2}. \quad (25)$$

In the parametrization induced by ϕ , the volume measure becomes,

$$\begin{aligned} dvol_{Fisher} &= \frac{dvol_{Fisher}}{d\mu_\phi} \frac{d\mu_\phi}{d\mu_\psi} d\mu_\psi \\ &= 2^{-n^2} \prod_{i < j \leq n} |\lambda_i - \lambda_j| \sqrt{\prod_i \frac{1}{\lambda_i^{n+1}}} dLeb_{\mathbb{R}_+^n} dHaar_{SO(n)}. \end{aligned} \quad (26)$$

6.2 The Riemannian volume density in the exponential parametrization

We are interested here in the computation of the following quantity:

$$\theta_{\Sigma_0} : \Sigma_1 \mapsto \theta_{\Sigma_0}(\Sigma_1) = \frac{dvol_{Fisher}}{dexp_{\Sigma_0}^*(\mu_\phi)}(\Sigma_1), \quad (27)$$

that is to say the density of the Riemannian measure of $\mathcal{G}_{m=0}(n)$ with respect to the Lebesgue measure of the parametrization induced by exp_Σ . In other words, θ is the Riemannian volume element in the exponential parametrization. Recall that $T_\Sigma \mathcal{M} = Sym(n)$, where $Sym(n)$ stands for the symmetric matrices. The computation of Eq. 27 involves two factors known in the literature. In the parametrization of $T_\Sigma \mathcal{M} = Sym(n)$ induced by ψ , the volume element of the Fisher metric is known, see [29]:

$$dvol_{Fisher} = 2^{-n(n+1)/2} \prod_{i < j} \sinh\left(\frac{|\lambda_i - \lambda_j|}{2}\right) d\mu_\psi.$$

The scaling factor corresponds to the Fisher metric, which differs from the conventions of [29]. Note that the scaling factor have little influence on density estimation. The volume change induced by the parametrization is given in Eq. 12:

$$f_{\psi \rightarrow \phi}(m, \lambda, R) = \prod_{i < j \leq n} |\lambda_i - \lambda_j|,$$

Thus,

$$\theta_{\Sigma_0}(\Sigma_1) = 2^{-n(n+1)/2} \prod \frac{\sinh\left(\frac{|\lambda_i - \lambda_j|}{2}\right)}{|\lambda_i - \lambda_j|}, \quad (28)$$

where the λ_i are the eigenvalues of $\log_{\Sigma_0}(\Sigma_1) = \log(\Sigma_0^{1/2} \Sigma_1 \Sigma_0^{1/2})$. The explicit expression of the change of volume $\theta_p(q)$ induced by the exponential map enables us to use the kernel density estimation. Given a set of covariance matrices Σ_i and a scaling parameter $r > 0$, the estimator becomes:

$$\hat{f}_k^r(\Sigma) = \frac{1}{k} \sum_i \frac{1}{r^n} \frac{1}{\theta_{\Sigma_i}(\Sigma)} K\left(\frac{d_{Fisher}(\Sigma, \Sigma_i)}{r}\right). \quad (29)$$

Thus we have the following expression for the Pelletier kernel density estimator

$$\hat{f}_k^r = \frac{1}{k} \sum_i \frac{2^{-n(n+1)/2}}{r^n} \prod_{p < q} \frac{|\lambda_p - \lambda_q|}{\sinh\left(\frac{|\lambda_p - \lambda_q|}{2}\right)} K\left(\frac{\| \log(\Sigma_i^{-1/2} \Sigma \Sigma_i^{-1/2}) \|}{r}\right), \quad (30)$$

where the λ_i are the eigenvalues of $\log(\Sigma_i^{-1/2} \Sigma \Sigma_i^{-1/2})$.

7 Euclidean subspaces

7.1 The space of Gaussian laws with fixed rotation

Given a rotation matrix R , we study here the specificity of the space of Gaussian laws of the form $\mathcal{N}_{(m, RDR^t)}$ where D is diagonal matrix, under the Wasserstein metric. We choose here to represent the Gaussian law $\mathcal{N}_{(m, RDR^t)}$ by its mean $m \in \mathbb{R}^n$ and a set of standard deviation $\sigma \in \mathbb{R}^n$, with σ_i^2 an eigenvalue of D . In the (m, σ) parametrization, the expression of the metric computed using Eq. 13 takes the following form:

$$G_{\mathcal{N}_{m, \sigma}}^W = \begin{pmatrix} & e_{m_i} & & & & e_{\sigma_i} \\ 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}$$

Thus in the (m, σ) parametrization, the metric is Euclidean. The density estimation can then be achieved using standard techniques.

7.2 The space of Gaussian laws with fixed mean and rotation

Given a mean m and a rotation R , we study here the specificity of the space of Gaussian laws of the form $\mathcal{G}_{m, R}(n) = \{\mathcal{N}_{(m, RDR^t)}, \text{ with } D \text{ diagonal}\}$, under the Fisher metric. The matrix of the metric takes the following form:

$$G_{\mathcal{N}_{m, RD_\lambda R^t}}^{Fisher} = \begin{pmatrix} & e_{\lambda_i} \\ \frac{1}{\lambda_1^2} & \\ & \cdot \\ & & \frac{1}{\lambda_n^2} \end{pmatrix}$$

After the change of coordinates $\gamma_i = \log(\lambda_i)$, the metric becomes:

$$G_{\mathcal{N}_{m, RD_\lambda R^t}}^{Fisher} = \begin{pmatrix} & e_{\gamma_i} \\ 1 & \\ & \cdot \\ & & 1 \end{pmatrix}$$

Thus in the $\gamma = \log(\lambda)$ parametrization, the metric is Euclidean. The density estimation can then be achieved using standard techniques.

7.3 The space of Gaussian laws with fixed covariance

Given a covariance matrix Σ , we study now the specificity of the space of Gaussian laws of the form $\mathcal{N}_{(m,\Sigma)}$, under the Fisher metric. Note that the mean under the Wasserstein metric follows the canonical Euclidean scalar product. From Eq. (13) in $\mathcal{G}(n)$:

$$\langle u, v \rangle_{m,\Sigma}^{Fisher} = u_m^t \Sigma^{-1} v_m + \frac{1}{2} \text{tr}(\Sigma^{-1} u_\Sigma \Sigma^{-1} v_\Sigma),$$

and we have that, at a fixed Σ , it is given by

$$\langle u_m, v_m \rangle_m^{Fisher} = u_m^t \Sigma^{-1} v_m.$$

The metric on the mean is Euclidean, of scalar product Σ^{-1} . The density estimation can then be achieved using standard techniques.

8 The space of univariate Gaussian laws

This section addresses the case of $\mathcal{G}(n = 1)$. Space $\mathcal{G}(n = 1)$ is identified with the half plane $\mathbb{R} \times \mathbb{R}_+$. That means that each normal law \mathcal{N} is described by its mean $m \in \mathbb{R}$ and its (positive) standard deviation $\sigma \in \mathbb{R}_+$. Note that, as in section 5.2, there is a slight change of convention with respect to the previous sections since the equivalent of the covariance matrix Σ would be the square of the standard deviation σ^2 . The study is focused on the Fisher metric since the case of the Wasserstein metric has been analyzed in section 7.1.

8.1 Fisher metric and the Poincaré upper half plane

The Poincaré upper half plane of curvature R is a half plane $\{x \in \mathbb{R}, y \in \mathbb{R}_+\}$ endowed with the following Riemannian metric:

$$G_{(x,y)} = a^2 \begin{pmatrix} 1/y^2 & 0 \\ 0 & 1/y^2 \end{pmatrix}, \quad (31)$$

with $R = -\frac{1}{a^2}$. The Poincaré upper half plane is a model of hyperbolic geometry, see [6]. The distance between (x_1, y_1) and (x_2, y_2) is given by:

$$d((x_1, y_1), (x_2, y_2)) = a \cosh^{-1} \left(1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2y_1 y_2} \right). \quad (32)$$

The Poincaré upper half plane is isotropic: each location and directions are equivalent. After having considered the change of convention between Σ

and σ , the matrix of the Fisher metric on $\mathcal{G}(n = 1)$ can be deduced from Eq. (13), i.e.,

$$G_{(m,\sigma)}^{Fisher} = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix} \quad (33)$$

Consider now the application $s : (m, \sigma) \mapsto (x, y) = (m, \sqrt{2}\sigma)$. Solving the equation:

$$(ds)^t \cdot G_{(x,y)}^{Fisher} \cdot ds = G_{(m,\sigma)}^{Fisher} \quad (34)$$

where ds is the Jacobian matrix of s , the expression of the metric is obtained in the new parametrization:

$$G_{(x,y)}^{Fisher} = 2 \begin{pmatrix} 1/y^2 & 0 \\ 0 & 1/y^2 \end{pmatrix} \quad (35)$$

Thus, under the Fisher metric, the space $\mathcal{G}(n = 1)$ is isometric to a Poincaré upper half plane of curvature $-\frac{1}{2}$. Let \mathbb{H} denotes just the Poincaré upper half plane of curvature $-\frac{1}{2}$.

8.2 The Riemannian volume element in the (m, σ) parametrization

In the (x, y) parametrization, the Riemannian volume measure vol has the following expression:

$$\frac{dvol}{dLeb_{(x,y)}} = \sqrt{\det(G_{(x,y)}^{Fisher})} = \frac{2}{y^2}. \quad (36)$$

8.3 The Riemannian volume element in the exponential parametrization

We proceed similarly to section 6.2. In polar coordinates $(r \in \mathbb{R}_+, \alpha \in [0, 2\pi[)$, the expression of the Riemannian volume element is known, see [10]:

$$dvol = 2 \sinh(r) dr d\alpha$$

By switching to an Euclidean parametrization of the tangent plane, we obtain:

$$\frac{dvol}{dexp^*(Leb_p)}(q_{(r,\alpha)}) = 2 \frac{\sinh(r)}{r}. \quad (37)$$

We obtain then

$$\theta_p(q) = 2 \frac{\sinh(d(p, q))}{d(p, q)}, \quad (38)$$

where θ is the function used in Eq. (8). The explicit expression of the change of volume $\theta_p(q)$ induced by the exponential map enables us to use the kernel density estimation. Given a set of univariate Gaussian distributions $\mathcal{N}_i = (m_i, \sigma_i)$ and a scaling parameter $r > 0$, the estimator becomes:

$$\hat{f}_k^r(\mathcal{N}) = \frac{1}{k} \sum_i \frac{1}{r^2} \frac{d_{Fisher}(\mathcal{N}, \mathcal{N}_i)}{2 \sinh(d_{Fisher}(\mathcal{N}, \mathcal{N}_i))} K\left(\frac{d_{Fisher}(\mathcal{N}, \mathcal{N}_i)}{r}\right), \quad (39)$$

where here $d_{Fisher}(\mathcal{N}, \mathcal{N}_i) = \sqrt{2} \cosh^{-1}\left(1 + \frac{(m-m_i)^2 + 2(\sigma-\sigma_i)^2}{2\sigma\sigma_i}\right)$.

9 Partial quantities: mean, eigenvalues and rotation

This section addresses the study of the standard partial quantities that are the mean, the eigenvalues and the rotation. On the one hand, depending on the application, the interesting information is sometimes carried by only one or two partial quantities. On the other hand, the curse of dimensionality pushes us to reduce the dimension of the studied objects.

We start the discussion with an example. Let X be a random variable valued in \mathbb{R}^2 of density f . Consider the distribution of the radius of the polar coordinates (r, θ) . Firstly, the Lebesgue measure of \mathbb{R}^2 can be written as a product $dLeb = r d\theta dr$. Thus the Lebesgue measure of \mathbb{R}^2 induces a measure on radiuses $r dr$. The density computed with respect to $r dr$ can be interpreted as an average of f over a slice of constant radius. Secondly, the metric on \mathbb{R}^2 induces a natural metric on the space of radius seen as the quotient space $\mathbb{R} \sim \mathbb{R}^2/\theta$. The natural quotient metric is the Euclidean metric on \mathbb{R} , the associated measure being the Lebesgue measure dr . Thus this example shows that there is not a unique way of addressing the density estimation of partial quantities.

Since the parametrization ψ is not injective, the partial quantities of a Gaussian distribution are not clearly defined. Indeed given $\Sigma \in \text{SPD}(n)$ there exists several $(\lambda, R) \in \mathbb{R}_+^n \times \text{SO}(n)$ such that $\Sigma = R D_\lambda R^t$. If this is not a problem in most of theoretical works, it is when one comes to applications. Indeed it is important to always use the same representation of each object. Let $\tilde{\mathcal{G}}(n)$ be the set of Gaussian laws $\mathcal{N}_{m, \Sigma}$ such that Σ has distinct eigenvalues. Note that the difference between $\tilde{\mathcal{G}}(n)$ and $\mathcal{G}(n)$ has a null measure for standard measures. Let define

$$E = \{\lambda \in \mathbb{R}_+^n \mid \forall 0 \leq i < j \leq n, \lambda_i < \lambda_j\},$$

and

$$H = \{D_\lambda \mid \lambda \in \{+1, -1\}^n, \det(D_\lambda) = 1\}.$$

Note that H is a group. The quotient $SO(n)/H = \{RH, R \in SO(n)\}$ can be identified to the fundamental domain $\mathcal{D} \subset SO(n)$ where for $R \in \mathcal{D}$, the maximal element with the smallest index element of each column $1 \leq j < n$ is positive. In other words

$$D = \{R \in SO(n) \mid \forall j < n, R_{\min\{\arg\max_i \{R_{i,j}\}, j\}} > 0\}.$$

Then, parametrization

$$\tilde{\psi} : \begin{cases} \mathbb{R}^n \times E \times \mathcal{D} & \rightarrow \tilde{\mathcal{G}}(n) \\ (m, \lambda, R) & \mapsto \mathcal{N}_{m, (RD_\lambda R^t)} \end{cases}$$

is an homeomorphism.

Remind the context of section 3. Let Ω be a space, endowed of with a σ -algebra and a probability measure p . Space $\mathcal{G}(n)$ is equipped with a Riemannian metric G , the associated Riemannian measure is called vol . Let X be a random variable, $X, \Omega \mapsto \mathcal{G}(n)$. The measure on $\mathcal{G}(n)$ induced by X is noted μ_X . We assume that μ_X has a density, noted f , with respect to vol . $\tilde{\mathcal{G}}(n)$ can be identified to the product $\mathbb{R}^n \times E \times \mathcal{D}$, recall that $vol(\mathcal{G}(n) \setminus \tilde{\mathcal{G}}(n)) = 0$. Let X_{mean} , X_λ and X_R be the random variables naturally induced by the identification between $\tilde{\mathcal{G}}(n)$ and $\mathbb{R}^n \times E \times \mathcal{D}$. Note that X_R is not always properly defined. This problem can be neglected given that $vol(\mathcal{G}(n) \setminus \tilde{\mathcal{G}}(n)) = 0$ and that the law of X has a density with respect to vol . Let $\mu_{X_{mean}}$, μ_{X_λ} , μ_{X_R} be the associated measures.

9.1 Average over slices

We start by showing that in many classical cases, measures on multivariate Gaussian laws can be decomposed as a product between a measure on the mean space, a measure on the eigenvalue space, and a measure on the rotation space. Observe that rotations naturally act on $\mathcal{G}(n)$, i.e.,

$$R_0 \cdot \mathcal{N}_{m, \lambda, R_1} = \mathcal{N}_{m, \lambda, R_0 R_1}.$$

Theorem 9.1 *Let μ be a measure on $\tilde{\mathcal{G}}(n)$, invariant under the action of rotations. Measure μ is equivalent to a product measure,*

$$\mu = \mu_{mean} \times \mu_\lambda \times \mu_R,$$

where μ_{mean} is a measure on \mathbb{R}^n , μ_λ is a measure on E and μ_R is a measure on $SO(n)$.

Proof This is a direct consequence of Theorem 8.2, page 155, of [31].

Eq. 16 and Eq. 20 are consistent with this result. In most situations, measures μ on $\mathcal{G}(n)$ present the desired invariances, and $\mu(\mathcal{G}(n) \setminus \tilde{\mathcal{G}}(n)) = 0$. Writing μ as a product measure enables to define densities for the mean, the eigenvalues and the rotations separately. This is in particular the case for the measures induced by the Fisher metric, the Wasserstein metric, and the Lebesgue measure of $\mathbb{R}^n \times \text{SPD}(n)$. The measure vol is thus described by a product $vol = \mu_{mean} \times \mu_\lambda \times \mu_R$

The measure $\mu_{X_{mean}}(A)$ is the probability that $x_{mean} \in A$ when x follows μ_X . Thus,

$$\mu_{X_{mean}}(A) = \mu_X(A \times E \times \mathcal{D}).$$

Let f_{mean} be the density of $\mu_{X_{mean}}$ with respect to μ_{mean} . Since

$$\mu_X(A \times E \times \mathcal{D}) = \int_A \int_E \int_{\mathcal{D}} f d\mu_R d\mu_\lambda d\mu_{mean},$$

we have,

$$f_{mean}(m) = \int_E \int_{\mathcal{D}} f(m, \lambda, R) d\mu_R d\mu_\lambda.$$

Let $\alpha = (\mu_\lambda(E)\mu_R(\mathcal{D}))^{-1}$. Quantity αf_{mean} can be interpreted as an average value of the original density f over a slice $E \times \mathcal{D}$. The same hold for μ_{X_λ} and μ_{X_R} . Up to a scaling factor, we recall here the expression of the different measures:

$$\begin{aligned} \mu_{\phi, mean} &= Leb, \\ d\mu_{\phi, \lambda} &= \prod_{i < j \leq n} |\lambda_i - \lambda_j| dLeb(\lambda), \\ \mu_{\phi, R} &= Haar_{SO(n)}, \\ \mu_{mean}^{Fisher} &= Leb, \\ \mu_R^{Fisher} &= Haar_{SO(n)}, \\ \mu_{mean}^W &= Leb, \\ d\mu_\lambda^W &= \prod_i \frac{1}{\sqrt{\lambda_i}} \prod_{i < j} \frac{|\lambda_i - \lambda_j|}{\sqrt{(\lambda_i + \lambda_j)}} dLeb, \\ \mu_R^W &= Haar_{SO(n)}, \end{aligned}$$

where $Haar_{SO(n)}$ is restricted to \mathcal{D} . The expression of μ_λ^{Fisher} varies between $\mathcal{G}(n)$ and $\mathcal{G}_{m=0}(n)$. When $\mathcal{G}(n)$ is the underlying space, we have:

$$d\mu_\lambda^{Fisher} = \prod_{i < j \leq n} |\lambda_i - \lambda_j| \sqrt{\prod_i \frac{1}{\lambda_i^{n+2}}} dLeb, \quad (40)$$

and when $\mathcal{G}_{m=0}(n)$ is the underlying space the expression becomes:

$$d\mu_{\lambda}^{Fisher} = \prod_{i < j \leq n} |\lambda_i - \lambda_j| \sqrt{\prod_i \frac{1}{\lambda_i^{n+1}}} dLeb \quad (41)$$

Fig. 7 shows several cases of density change for $n = 2$.

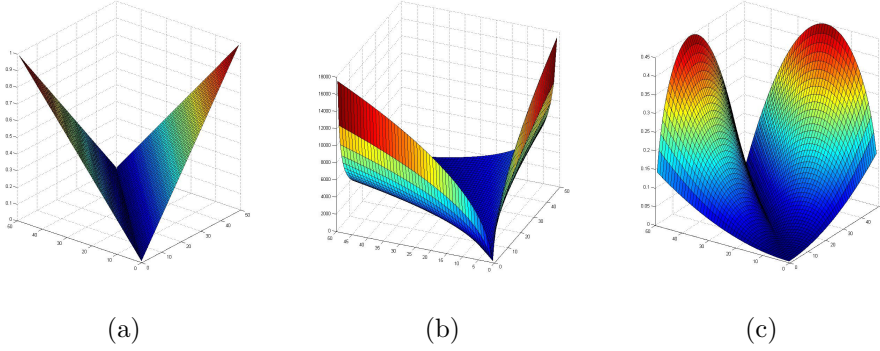


Figure 7: For $n = 2$, (a),(b) and (c) are visualizations of $d\mu_{\phi,\lambda}$, $d\mu_{\lambda}^{Wass}$ and $d\mu_{\lambda}^{Fisher}$ (for $\mathcal{G}_{m=0}(n)$) respectively.

The question of the underlying metric associated to these measures has not yet been clarified. Thus the estimation can be achieved using the approach described section 3.1. Despite the absence of clear argument, the Lebesgue measure and the Haar measure push to the use of an underlying translation invariant metric.

9.2 Quotient metric on partial quantities

The space associated to each partial quantity can be seen as a quotient of $\mathcal{G}(n)$ by the rest of the partial quantities. The quotient $\mathcal{G}(n)/(\mathbb{R}_+^n \times SO(n))$ denotes the identification elements with all those which differ only in the eigenvalues or the rotation. Thus $\mathcal{G}(n)/(\mathbb{R}_+^n \times SO(n))$ can be interpreted as the space of means. In several cases, the Riemannian metric on $\mathcal{G}(n)$ induces a canonical metric on such quotient space. This metric induces a measure on the partial quantity, which enable the estimation of density of the associated random variable $X_{(\cdot)}$.

9.2.1 Mean

For the Wasserstein metric and the Euclidean metric of $\mathbb{R}^n \times Sym(n)$, the quotient space $\mathbb{R}^n \sim \mathcal{G}(n)/(\mathbb{R}_+^n \times SO(n))$ inherits naturally of the canonical Euclidean metric. Indeed, each couple (λ, R) induces the same metric on

m . The density estimation can thus be achieved using standard techniques. Note that the induced Lebesgue measure is similar to μ_{mean} .

Recall the expression of the scalar product of the Fisher metric (13):

$$\langle u, v \rangle_{m, \Sigma}^{Fisher} = u_m^t \Sigma^{-1} v_m + \frac{1}{2} \text{tr}(\Sigma^{-1} u_\Sigma \Sigma^{-1} v_\Sigma).$$

Each covariance matrix Σ induces a metric on the space of means. However, the metric varies when Σ varies. Thus the Fisher metric do not induces a canonical metric on $\mathbb{R}^n \sim \mathcal{G}(n)/(\mathbb{R}_+^n \times SO(n))$.

9.2.2 Eigenvalues

For $\lambda \in \mathbb{R}^n$, let D_λ be the associated diagonal matrix. Fixing a mean m and a rotation R induces a metric on eigenvalues:

$$d_{m,R}(\lambda_1, \lambda_2) = d(\mathcal{N}_{m, D_{\lambda_1}, R}, \mathcal{N}_{m, D_{\lambda_2}, R}). \quad (42)$$

The group of rotations naturally acts on Gaussian laws by $R_{a \cdot \text{Left}} \mathcal{N}_{m, \lambda, R_b} = \mathcal{N}_{m, \lambda, R_a R_b}$. Furthermore for the Euclidean metric of $\mathbb{R}^n \times \text{Sym}(n)$, the Fisher and the Wasserstein metric, we have that for all $R \in SO(n)$:

$$d(\mathcal{N}_{m_1, \lambda_1, R_1}, \mathcal{N}_{m_2, \lambda_2, R_2}) = d(R_{\cdot \text{Left}} \mathcal{N}_{m, \lambda, R_1}, R_{\cdot \text{Left}} \mathcal{N}_{m, \lambda, R_1}). \quad (43)$$

Thus since $d_{m,R}$ is independent of m and R , the three metrics induce a canonical metric on the quotient $\mathbb{R}_+^n \sim \mathcal{G}(n)/(\mathbb{R}^n \times SO(n))$.

On the one hand according to section 7.1, the Wasserstein metric on eigenvalues at fixed rotation is the canonical Euclidean metric after the change of coordinates $\sigma = \sqrt{\lambda}$. On the other hand, according to section 7.2 the Fisher metric on eigenvalues at fixed mean and rotation is the canonical Euclidean metric after the change of coordinates $\gamma = \log(\lambda)$. Finally, it is easy to see that the quotient metric induced by the Euclidean metric of $\mathbb{R}^n \times \text{Sym}(n)$, is the Euclidean metric on \mathbb{R}_+ . Note that the associated measures differs from μ_λ .

9.2.3 Rotation

Given a Gaussian law $\mathcal{N}_{m, \lambda, R}$, the Euclidean metric of $\mathbb{R}^n \times \text{Sym}(n)$, the Fisher and the Wasserstein metric induce a left-invariant metric on Q :

$$d_{\mathcal{N}_{m, \lambda, R}}(R_1 \in Q, R_2 \in Q) = d(R_{1 \cdot \text{Left}} \mathcal{N}_{m, \lambda, R}, R_{2 \cdot \text{Left}} \mathcal{N}_{m, \lambda, R}). \quad (44)$$

However it can be verified that the left-invariant metric is dependent of the choice of $\mathcal{N}_{m, \lambda, R}$. Thus neither the Euclidean metric of $\mathbb{R}^n \times \text{Sym}(n)$, nor the Fisher nor the Wasserstein metric induce a canonical quotient metric on $SO(n) \sim \mathcal{G}(n)/(\mathbb{R}^n \times \mathbb{R}_+^n)$.

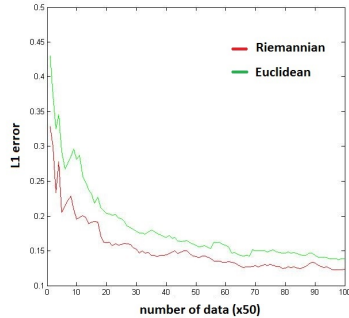


Figure 8: Convergence of the kernel density estimations of the law defined in Eq. 45 using an Euclidean and a Riemannian Kernel.

10 Experimental section

Every standard density estimation technique involves a scaling parameter. This scaling factor controls the influence of the observation x_i on the estimated density at x , depending on the distance between x and x_i . In the experiments, the scaling factor has been chosen following the framework proposed in [8]: a cross validation of the likelihood of the estimator.

This section start with an illustration, see Fig. 8(a), of the importance of the underlying metric on the density estimation method. Points are drawn in the Poincaré upper half plane according to the following density:

$$f(z) = \sqrt{d_{Fisher}(z, (0, 1))^2 - 1} \mathbf{1}_{d_{Fisher}(z, (0, 1)) < 1}. \quad (45)$$

From the draws, firstly, the density is estimated using an Euclidean kernel method, followed by an adaptation to the Riemannian measure, as described in section 3.1. Secondly, the density is estimated using a Riemannian kernel, see section 3.4. The base kernel is a quartic kernel $K(\|x\|) = \frac{3}{\pi}(1 - \|x\|^2)^2 \mathbf{1}_{\|x\| < 1}$. Fig. 8(a) shows the convergence of the two estimations to the true density.

10.1 Histograms of multiple grey-scale image acquisition

The studied example is a time lapse sequence of grey-scale images from a retina. At each pixel, we dispose of 20 successive fast acquisitions. By assuming a Gaussian distribution on the time series, we obtain a Gaussian valued image represented in Fig. 9.

Each pixel of the image contains an univariate Gaussian law. We recall that for univariate Gaussian laws, the Wasserstein metric is the Euclidean metric of the (m, σ) plane and that the Fisher metric is the Poincaré metric.

Histograms of the Gaussian-valued image are computed with respect the Fisher and the Wasserstein metric, using the appropriate kernel density estimation.

10.2 Density estimation in structure tensor images

The structure tensor image from a grayscale image is a field of symmetric positive definite matrices. Despite the fact that these matrices are not directly covariance matrices of Gaussian laws, they can be studied under the Fisher and the Wasserstein metric of $\mathcal{G}_{m=0}(n)$, typically $n = 2$ for 2D images and $n = 3$ for 3D ones. Fig. 10(b)-(g) shows the densities of the two eigenvalues $\lambda_1 \leq \lambda_2$ of the structure tensor field computed from Fig. 10(a). For each metric, the densities of eigenvalues are computed following considerations of section 9. For the induced measures μ_λ^{Eucl} , μ_λ^{Fisher} and $\mu_\lambda^{Wasserstein}$, the densities are firstly estimated using an Euclidean kernel density estimation, followed by an adaptation to the reference measures, given in section 9.1. According to section 9.2.2, each quotient metric is Euclidean in the right parametrization. In the adapted parametrizations, the densities are obtained using an Euclidean kernel density estimation.

Given a set of samples, the watershed transform of the complement of its associated density provides a non parametric clustering [26], similar to that of mean-shift algorithm. When the set of samples are the values taken by an image, the clustering of the samples can be interpreted as a segmentation of the image. Fig. 11 presents an example of texture segmentation according to the watershed transform of the density of the structure tensor field using measures.

10.3 EEG signals classification

This section presents results of density estimation for classification of Electroencephalogram (EEG) signals. The EEG is recorded during a asynchronous steady state visually evoked potential (SSVEP) brain computer interface (BCI) experiment. The SSVEP is stimulated by 3 sets of LEDs flickering at different frequencies: 13Hz, 17Hz or 21Hz. A trial consists of the subject gazing at one of four targets made of either of the 3 sets of LEDs or a non-flickering screen while the EEG is recorded on 8 electrodes at a sampling rate of 256Hz. For experimental purposes, only 4 electrodes are kept: Oz, POz, PO3, and PO4. The classification problem is also simplified to 2 classes: 21Hz-SSVEP versus no-SSVEP, and 13Hz-SSVEP versus 17Hz-SSVEP. For each subject and each class, 16 trials are recorded, with

a trial length of $t = 4s$. From the signal recorded at each trial, a covariance matrix $\mathbf{C} \in \mathbb{R}^{4 \times 4}$ is estimated using the Schäfer's shrinkage method [24].

Given an un-labeled covariance matrix, the classification experiment consists in determining the visual target attended by the subject. Two methods are tested, on two types of data. The first method is the one proposed in [14]. A barycenter is computed for each group of labeled points. An un-labeled point is affected to the class with the closest barycenter. The second method consists in estimating a probability density for each group of points associated to a given frequency. The un-labeled point is affected to the class which presents the highest density. The two methods are evaluated on the covariance matrices and on their eigenvalues. Each class has the same size and our experiments involve only two classes. The figures presented in tables 1 are the classification accuracy rate. The classifications are performed using different metric. "Euclidean" denotes the Euclidean metric on the coefficient of the matrices. "A.i." stands for affine-nvariant metric. The affine-invariant metric on $\text{SPD}(n)$ is isometric to the Fisher metric on centered multivariate Gaussian distributions. "Log-Euclidean" is the Euclidean metric on coefficients after taking the matrix logarithm of the covariance matrices. We recall that the log-Euclidean metric is the canonical Euclidean metric of the tangent space of $\text{SPD}(n)$ at the identity, under the affine-invariant metric. "Wasserstein" stands for the 2-Wasserstein distance on $\text{SPD}(n)$ induced by centered Gaussian distributions. The metrics used on eigenvalues are the quotient metrics described in section 9.2.2.

The results shows the importance of the choice of the underlying metric. Results obtained for the Affine invariant metric are superior to the ones obtained with the Euclidean and Wasserstein metric. The superiority of the results are apparently not due to the curvature of the space. Indeed, the Euclidean metric in the tangent space at the identity matrix after taking the logarithm provides similar results.

11 Conclusions and perspectives

Several density estimator on Riemannian manifolds have been considered. Due to their drawbacks on spaces of Gaussian distributions, histograms and orthogonal series have not been deeply studied. It is always possible to perform a classical estimations in the Euclidean context of the parameters, multiplied by a density ratio to obtain a density with respect to the desired measure. This type of density estimation is possible in each studied space and metric, but do not take into account the geometry of the space: if two metrics induce the same volume measure then the estimated density

Classification on eigenvalues:

0Hz vs 21Hz	Euclidean	A.i. (log)	Wasserstein (square root)
barycenter	0.6	0.71	0.64
density	0.61	0.73	0.63
13Hz vs 17Hz	Euclidean	A.i. (log)	Wasserstein (square root)
barycenter	0.52	0.58	0.51
density	0.54	0.58	0.54

Classification on matrices:

0Hz vs 21Hz	Euclidean	log-Euclidean	A. i.	Wasserstein
barycenter	0.61	0.71	0.71	0.67
density	0.63	0.76	0.76	0.7
13Hz vs 17Hz	Euclidean	log-Euclidean	A. i.	Wasserstein
barycenter	0.53	0.58	0.56	0.56
density	0.53	0.63	0.63	0.56

Table 1: Classification accuracy rate of EEG signals.

does not depend on the choice of one the metrics. Thus, the kernel density estimation seems to be the most adapted in most cases. It presents a reasonable computational complexity, and expressions of kernels are ready to use in each studied situations, except the space of multivariate Gaussian laws under the Fisher metric.

To our knowledge, the most original contribution of this work are the expression of kernels for the case of the Wasserstein metric, and the study of partial quantities. The latter being particularly useful in practical cases. The quantities useful for density estimation computed in the paper are summed up in tables 2-5. Section 10 presents results of the use of density estimation under Riemannian metrics for EEG signals classifications.

	multivariate	multivariate centered	univariate
adaptation from the Lebesgue measure	Eq. 15, Eq. 16	Eq. 25, Eq. 26	Eq. 36
kernels	-	Eq. 30	Eq. 39

Table 2: Density estimation on Gaussian laws under the Fisher metric.

	mean	eigenvalues	rotations (Q)
quotient metric	-	Euclidean on $\gamma = \log(\lambda)$	-
product measure	Lebesgue	Eq. 40 or Eq. 41	$Haar_{SO(n)}$

Table 3: Structures on partial quantities under the Fisher metric

	multivariate	multivariate centered	univariate
adaptation from the Lebesgue measure	Eq. 19, Eq. 20	Eq. 19, Eq. 20	Euclidean on $\sigma = \sqrt{\lambda}$
kernels	Eq. 23	Eq. 23	Euclidean on $\sigma = \sqrt{\lambda}$

Table 4: Density estimation on Gaussian laws under the Wasserstein metric

Future work will focus mainly on the application of our results to different problems in image and signal processing. Densities are useful objects in segmentation and classification. The mean shift algorithm is a standard approach that search local maxima and provides a segmentation of a density associated with a point cloud. This algorithm, proposed by [9] has been widely studied in \mathbb{R}^n . It is mainly used in point cloud segmentation, but also in tracking and smoothing. The algorithm has been studied on Riemannian manifolds in [27]. However in order to gain generality authors chose to work with “pseudo” kernels in the sense that they do not take into account the volume change induced by the exponential map. Indeed, this term is generally unknown. The study can thus be resumed for spaces Gaussian laws, except multivariate Gaussians under the Fisher metric. Probability densities are also important tools in Bayesian classification. The classification decision consists in maximizing the probability of a class given an observation. The Affine invariant metric on $SPD(n)$ has proven useful in the case of EEG signal. We will now search for situations where the Wasserstein metric provides better results than Euclidean metrics.

References

- [1] M. Agueh and G. Carlier. Barycenters in the Wasserstein Space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, January 2011.
- [2] Shun-ichi Amari, Hiroshi Nagaoka, and Daishi Harada. *Methods of information geometry*. Number 191 in Translations of mathematical

	mean	eigenvalues	rotations (Q)
quotient metric	Euclidean	Euclidean on $\sigma = \sqrt{\lambda}$	-
product measure	Lebesgue	$\frac{d\mu_\lambda^{Wass}}{dLeb(\lambda)} = \prod_i \frac{1}{\sqrt{\lambda_i}} \prod_{i < j} \frac{ \lambda_i - \lambda_j }{\sqrt{(\lambda_i + \lambda_j)}}$	$Haar_{SO(n)}$

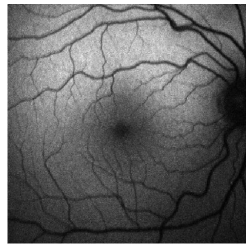
Table 5: Structures on partial quantities under the Wasserstein metric

monographs. American Math. Soc. [u.a.], Providence, RI, nachdr. edition, 2007.

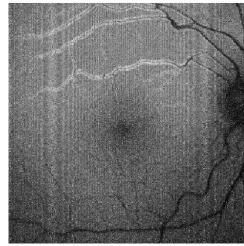
- [3] M. Arnaudon, F. Barbaresco, and Le Yang. Riemannian Medians and Means With Applications to Radar Signal Processing. *IEEE Journal of Selected Topics in Signal Processing*, 7(4):595–604, August 2013.
- [4] Dena Marie Asta. Kernel Density Estimation on Symmetric Spaces. *arXiv:1411.4040 [math, stat]*, November 2014. arXiv: 1411.4040.
- [5] Dario A. Bini and Bruno Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. *Linear Algebra and its Applications*, 438(4):1700–1710, February 2013.
- [6] J. W. Cannon, W. J. Floyd, R. Kenyon, and W. R. Parry. *Hyperbolic Geometry*. 1997.
- [7] Manfredo Perdigão do Carmo. *Riemannian Geometry*. Birkhuser, 1992.
- [8] R. P. W. Duin. On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions. *IEEE Transactions on Computers*, 25(11):1175–1179, 1976.
- [9] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, January 1975.
- [10] Alexander Grigoryan. *Heat Kernel and Analysis on Manifolds*. American Mathematical Soc., 2009.
- [11] Harrie Hendriks. Nonparametric Estimation of a Probability Density on a Riemannian Manifold Using Fourier Expansions. *The Annals of Statistics*, 18(2), 1990.
- [12] Guillermo Henry, Andrés Muñoz, and Daniela Rodríguez. k-Nearest neighbor density estimation on Riemannian Manifolds. *arXiv:1106.4763 [math, stat]*, June 2011. arXiv: 1106.4763.

- [13] Stephan F. Huckemann, Peter T. Kim, Ja-Yong Koo, and Axel Munk. Mbius deconvolution on the hyperbolic plane with application to impedance density estimation. *The Annals of Statistics*, 38(4):2465–2498, August 2010.
- [14] Emmanuel Kalunga, Sylvain Chevallier, Quentin Barthlemy, Karim Djouani, Yskandar Hamam, and Eric Monacelli. From euclidean to riemannian means: Information geometry for ssvep classification. In Frank Nielsen and Frdric Barbaresco, editors, *Geometric Science of Information*, volume 9389 of *Lecture Notes in Computer Science*, pages 595–604. Springer International Publishing, 2015.
- [15] Miroslav Lovri, Maung Min-Oo, and Ernst A Ruh. Multivariate Normal Distributions Parametrized as a Riemannian Symmetric Space. *Journal of Multivariate Analysis*, 74(1):36–48, July 2000.
- [16] Gaspard Monge. *Mmoire sur la thorie des dblais et des remblais*. De l’Imprimerie Royale, 1781.
- [17] James R. Munkres. *Elementary Differential Topology*. Princeton University Press, 1966.
- [18] Thomas H Parker. *Geometry Primer*, January 2015.
- [19] Bruno Pelletier. Kernel density estimation on Riemannian manifolds. *Statistics & Probability Letters*, (3):297–304, 2005.
- [20] Xavier Pennec. Probabilities and Statistics on Riemannian Manifolds : A Geometric approach. report, January 2004.
- [21] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66, January 2006.
- [22] Peter T. Kim and Donald St. P. Richards. Deconvolution Density Estimation on the Space of Positive Definite Symmetric Matrices. In *Nonparametric Statistics and Mixture Models*, pages 147–168. WORLD SCIENTIFIC, January 2011.
- [23] Salem Said, Lionel Bombrun, Yannick Berthoumieu, and Jonathan Manton. Riemannian Gaussian Distributions on the Space of Symmetric Positive Definite Matrices. *arXiv:1507.01760 [math, stat]*, July 2015. arXiv: 1507.01760.

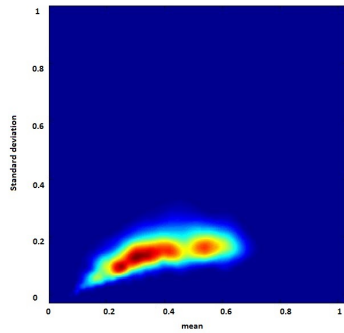
- [24] Juliane Schfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4:Article32, 2005.
- [25] Lene Theil Skovgaard. A Riemannian Geometry of the Multivariate Normal Model. *Scandinavian Journal of Statistics*, 11(4):211–223, January 1984.
- [26] Pierre Soille. *Morphological Image Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- [27] Raghav Subbarao and Peter Meer. Nonlinear Mean Shift over Riemannian Manifolds. *International Journal of Computer Vision*, 84(1):1–20, March 2009.
- [28] Asuka Takatsu. Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics - OSAKA J MATH*, 48(2011), 2011.
- [29] Audrey Terras. *Harmonic Analysis on Symmetric Spaces and Applications II*. Springer New York, New York, NY, 1988.
- [30] Cdric Villani. *Optimal Transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [31] Robert A. Wijsman. *Invariant measures on groups and their use in statistics*. Number v. 14 in Lecture notes-monograph series. Institute of Mathematical Statistics, Hayward, Calif, 1990.
- [32] Le Yang. Riemannian Median and Its Estimation. *LMS Journal of Computation and Mathematics*, 13, 2009.



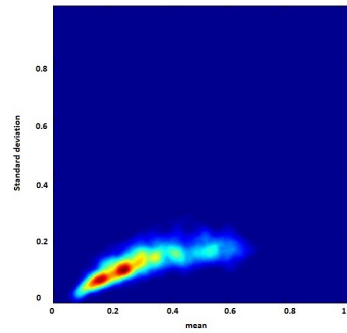
(a)



(b)



(c)

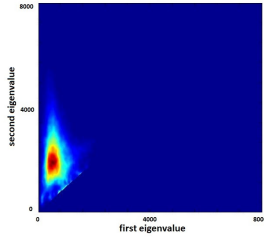


(d)

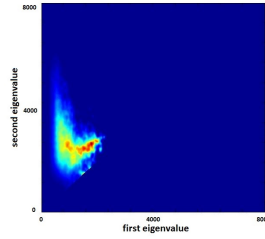
Figure 9: Time lapse sequence of images from the retina: (a) mean image, (b) standard deviation image. Histogram using the Fisher metric in (c) and the Wasserstein metric in (d)



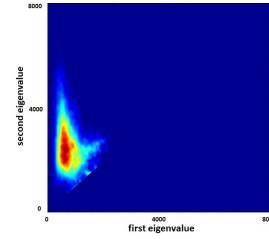
(a)



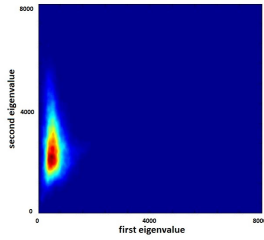
(b)



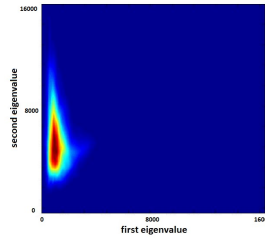
(c)



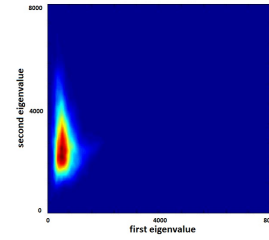
(d)



(e)



(f)



(g)

Figure 10: Densities estimated from the structure tensor field computed from image (a). Top row, densities computed from the induced measures μ_λ : (b) Euclidean metric of $\mathbb{R}^n \times Sym(n)$, (c) Fisher metric, (d) Wasserstein metric. Bottom row, densities computed using the quotient metrics: (e) Euclidean metric of $\mathbb{R}^n \times Sym(n)$, (f) Fisher metric (note the change of scale), (g) Wasserstein metric.

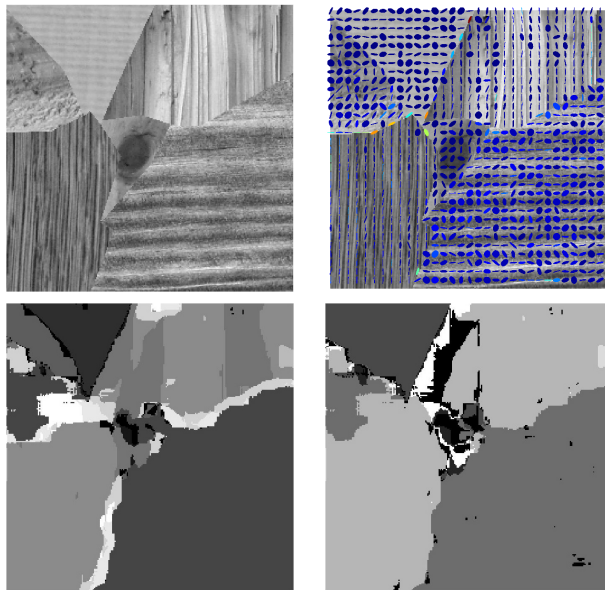


Figure 11: Texture image segmentation: (a) and (b) are respectively a texture image and its associated structure tensor field. The density of the eigenvalues of the structure tensors is estimated with respect to the two measures discussed in section 9 induced by the Wasserstein metric: (c) segmentation with respect to the Lebesgue measure on eigenvalues, and (d) segmentation with respect to μ_λ^{Wass} .