Probability measures on infinite dimensional Stiefel manifolds

E. Bardelli [*], A. C. G. Mennucci [†]

November 23, 2015

Abstract

An interest in infinite dimensional manifolds has recently appeared in Shape Theory. One such example is the Stiefel manifold, that has been proposed as a model for the space of immersed curves in the plane. It may be useful to define probabilities on such manifolds; this has many applications, such as object recognition, estimation, tracking, etc. In the case of finite dimensional manifolds, there is a vast literature regarding the definition and perusal of probabilities on finite dimensional manifolds. Unfortunately less is know about the infinite dimensional case. In this paper we will present some negative and some positive results. We highlight the main results in this abstract. Suppose in the following that H is an infinite dimensional separable Hilbert space.

Let $S \subset H$ be the sphere, fix $p \in S$. Let μ be the probability that results when wrapping a Gaussian measure γ from T_pS onto S using the exponential map. Let $v \in T_pS$ be a Cameron–Martin vector for γ; let R be a rotation of S in the direction v, and $\nu = R_\# \mu$ be the rotated measure. Then μ, ν are mutually singular. This is counterintuitive, since when γ is a Gaussian measure on H and T is the translation in a Cameron–Martin direction, then $T_\# \gamma$ and γ are mutually absolutely continuous.

Suppose now that γ is a Gaussian measure on H; then there exists a smooth closed manifold $M \subset H$ such that the projection of H to the nearest point on M is not well defined for points in a set of positive γ measure. This is opposite to what is observed in finite dimensional spaces.

The situation is instead better for a special class of smooth manifolds, the Stiefel manifolds. Let $M = \mathrm{St}\,(n, H) \subset H^n$ be the Stiefel manifold. Let γ be a non-degenerate Gaussian measure in H^n; then the projection of $x \in H^n$ to the nearest point $z \in M$ is well defined for γ-almost all x. Consequently it is possible to project γ to M to define a probability on M. This has important applications for Shape Theory, since $\mathrm{St}\left(2, L^2\right)$ has been proposed as the model for the manifold of all immersed curves in

[*]Microsoft Development Center Copenhagen, Denmark. *The author contributed to this paper in her personal capacity. The views expressed in this paper are those of the authors and do not necessarily reflect the views of Microsoft Corporation*

[†]Scuola Normale Superiore, Pisa, Italy

the plane. The above procedure can be easily implemented numerically, and provides an effective family of probability models on the space of immersed curves in the plane.

Keywords. Probability, Gaussian measure, infinite dimensional manifold, Riemannian manifold, Hilbert space, Shape Space.

1 Introduction

Probability theory has been widely studied for almost four centuries. Large corpuses have been written on the theoretical aspects. A commonly studied subject was the theory of probability distributions defined on a countable set or on (an open subset of) a finite dimensional vector space. This setting though was insufficient for some important applications.

In 1935 Kolmogoroff [14] provided the first general definition of a Gaussian measure on an infinite dimensional space. [1] This setting was derived from, and formalized part of, the theory of stochastic processes. Subsequently the theory was expanded and refined in many works.

Another interesting branch of probability theory is the case of probabilities in finite dimensional manifolds. This has many important applications in Shape Theory. One example is the Kendall space [12, 13, 16]. Another example is the Lie group $SO(n)$ of rotations: probabilities on $SO(3)$ may be used *e.g.* for Bayesian estimation of motions of rigid bodies.

1.1 A Shape Space

Infinite dimensional manifolds appear often in Shape Theory. One example is the Stiefel manifold.

Definition 1.1. Let $p \in \mathbb{N}, p \geq 1$ and H be a Hilbert space. The Stiefel manifold $\mathrm{St}\,(p, H)$ is the subset of H^p consisting of orthonormal p-tuples of vectors. In symbols,

$$\mathrm{St}\,(p, H) = \{(v_1, \dots, v_p) \in H^p \ \mid \forall i, j \text{ with } i \neq j,\ \langle v_i, v_j \rangle = 0 \text{ and } \forall i,\ |v_i| = 1\,\} \quad .$$

The Stiefel manifold is a smooth embedded submanifold of H^p, hence it inherits its Riemannian structure.

We will use the above definition also in the case when H is finite dimensional; in that case we will always assume silently that $\dim(H) \geq p$ (otherwise $\mathrm{St}\,(p, H)$ is empty).

In [20] Younes studied the space $\mathcal{M}$ of smooth immersed closed planar curves. Those ideas were then revisited in Younes *et al* [21], where the authors proved that the quotient of $\mathcal{M}$ with respect to translations and scalings, when endowed with a particular Sobolev–type Riemannian metric, is isometric to a subset of the Stiefel manifold $\mathrm{St}\left(2, L^2\right)$, where $L^2 = L^2([0, 1])$ is the usual Hilbert space

[1] As reported in the bibliographical appendix of [4].

of real square integrable functions. Similarly the quotient of $\mathcal{M}$ with respect to rotations, translations and scalings is isometric to a subset of the Grassmann manifold of two-dimensional planes in L^2.

Sundaramoorthi *et al* [19] studied these Shape Spaces as well; they noted that there is a closed form formula for the geodesic starting from a given point with a given velocity in $\mathrm{St}\left(p, L^2\right)$ (adapting a method described in [9]); they proposed a novel method for tracking shapes bounded by curves, that is based on a model discrete time stochastic process on $\mathrm{St}\left(2, L^2\right)$.

Moreover Harms and Mennucci [11] proved that any two points in the Stiefel manifold (respectively in the Grassmann manifold) are connected by a minimal length geodesic.

Since the manifold $\mathrm{St}\left(2, L^2\right)$ enjoys all the above useful properties, and it can be identified with a Shape Space of curves, then it is a natural choice for Computer Vision tasks. Many such tasks require that a probability be defined on the Shape Space. Unfortunately little is known in this respect.

In this paper we will present some negative and some positive results.

1.2 Reference measure in finite dimensional manifolds

When M is finite dimensional, there are many ways to define a reference measure on M.

Suppose that M is an n-dimensional complete Riemannian manifold. Let $\mathcal{H}^n$ be the n-dimensional Hausdorff measure defined in M (using the distance induced by the Riemannian metric). Let $A \subset \mathbb{R}^n, V \subset M$ be open sets, and let $\varphi : A \to V$ be a local chart; then the push forward of $\mathcal{H}^n$ using φ^{-1} is equivalent[2] to the Lebesgue measure (when they are restricted to V and respectively to A); this is proved *e.g.* in Section 3.2.46 in [10], or in Section 5.5 in [6].

Another way to measure Borel subsets of finite dimensional differentiable manifold M is by using volume densities. All volume densities are equivalent. If M is an oriented manifold, then the volume density can be derived from a volume form; if moreover M is an oriented Riemannian manifold, then there is a natural volume form (derived from the Riemannian metric), and the associated volume density coincides again with $\mathcal{H}^n$. See again Sec. 5.5 in [6].

Other classical definitions of measures exist, such as the Haar measure in topological groups.

Each one of the above may be adopted as a "reference measure". Once a reference measure is fixed, it can be used to define other probabilities on M, by using densities.

Unfortunately, in the infinite dimensional case there is no canonical reference measure. In an infinite dimensional Hilbert space there is no equivalent to the Lebesgue measure; in particular any translation invariant measure is either identically 0 or it is $+\infty$ on all open sets.

[2] "Equivalent" means "mutually absolutely continuous". See Definition 1.2.

A well known and deeply studied family of probabilities on these spaces is the family of Gaussian probabilities (also called "Gaussian measures"). For this reason we will often make use of Gaussian measures. In Sec. 2 we will provide a brief compendium of the theory of Gaussian measure in separable Hilbert spaces.

We will then address two methods for defining a probability on a smooth complete Riemannian manifold M modeled on a Hilbert space.

1.3 Probabilities by exponential map

The first method uses the exponential map, and is discussed in detail in Sec. 3. We present here a short overview. Let us fix a point $p \in M$, then the tangent space T_pM is isomorphic to a subspace of H, so we can define a probability measure μ on T_pM, *e.g.* a Gaussian measure. At the same time, the exponential map $\exp_p : T_pM \to M$ is smooth, so we can push forward the probability μ to define the probability $\nu = \exp_p \#\mu$ on M. In some texts this procedure is known as *"wrapping"*. This method undoubtely works fine, even when M is infinite dimensional. There is though an important difference between the finite dimensional and the infinite dimensional case.

Suppose for a moment that M is finite dimensional. The tangent spaces are n-dimensional vector spaces, so (up to the choice of an orthonormal base) we identify them with $\mathbb{R}^n$; we can then define on them the Lebesgue measure $\mathscr{L}^n$. This measure does not depend on the choice of the base.

Let $p \in M$. We recall that the exponential map is a local diffeomorphism near the origin of T_pM. Again, if we push forward $\mathscr{L}^n$ using this local diffeomorphism, we obtain a measure that is equivalent to $\mathcal{H}^n$ near p. (The n-dimensional Hausdorff measure $\mathcal{H}^n$ is defined in M using the distance induced by the Riemannian metric).

This local result can be extended to a global result, as we will show in Prop. 3.1 and Cor. 3.2. In short, let $p \in M$ a point and T_pM the tangent space to M in p. Let γ be a measure on T_pM, equivalent to the Lebesgue measure $\mathscr{L}^n$. Then the wrapped measure is equivalent to the Hausdorff measure on M. So any two measures on M built by the above procedure will be equivalent (*i.e.* mutually absolutely continuous).

In the infinite dimensional case it is well known that this is not true in general; it readily fails in the case of Gaussian measures defined on a separable Hilbert space. Suppose for a moment in the above example that M is itself a separable Hilbert space, so that we identify $T_pM = M$ for all $p \in M$; we view M as a (trivial) Riemannian manifold by associating the norm $\|\cdot\|_M$ to each tangent space T_pM. In this case the wrapping is trivial. Let $p_1 = 0$, $p_2 \neq 0$. Fix a non-degenerate Gaussian measure μ on M. Since $\exp_{p_1}$ is the identity map, then the wrapping $\nu_1 = \exp_{p_1} \#\mu$ of μ is μ itself. At the same time the wrapping $\nu_2 = \exp_{p_2} \#\mu$ of μ is the translation of μ, translated by the vector p_2. It is well known that ν_1 and ν_2 are equivalent if and only if p_2 lies in the Cameron-Martin space of μ_1, otherwise they are mutually singular. See next section 2 for detailed definitions and further results.

The matter becomes even more intricate in the case of an infinite dimensional manifold. Let S be the unit sphere in a infinite dimensional separable Hilbert space. We will show in Theorem 3.7 a result as follows. If we wrap a non degenerate Gaussian measure around the sphere S, and then we rotate it to obtain a second measure on S, then the two measures on the sphere are mutually singular. We can prove this fact for a class of rotations (*i.e.* unitary operators) that are intuitively analogous to the Cameron–Martin translations described in Prop. 2.8.

It is currently unknown to us if there exists any nontrivial rotation such that the two measures are equivalent. See also Remark 3.10.

1.4 Probabilities by projection

The second method can be used when M is a smooth embedded closed submanifold of a larger Hilbert space H. In this case we may define a probability on H, and then try to "project" it to M. This will be discussed in detail in Sec. 4.

For any such M consider the set $U_M \subset H$ of points $p \in H$ such that there is a unique point $z \in M$ of minimum distance from p; so we define the "projection" that is the map $\pi_M : U_M \to M$ such that $\pi_M(p) = z$.

Again, in the finite dimensional case this works fine. We will see in Prop. 4.2 that the set $H \setminus U_M$ has zero Lebesgue measure. So any probability on H that is defined by a density *wrt* the Lebesgue measure can be projected to M.

Instead in the infinite dimensional case this fails. We will show in Theorem 4.6 that for any Gaussian measure defined on ℓ^2 there exists a submanifold $M \subseteq \ell^2$ such that the "projection" fails to be defined almost everywhere, that is $\ell^2 \setminus U_M$ has positive measure.

We will though show in Section 4.3.3 that the projection method works fine in the case of the Stiefel manifold $\mathrm{St}\,(p, H)$. Indeed, for any non-degenerate Gaussian measure η on H^p, the projection from H^p to the nearest point in $\mathrm{St}\,(p, H)$ is defined for η-almost all points. So we can project η onto $\mathrm{St}\,(p, H)$ to define a "Gaussian-like" probability on it. This is again another point in favor of using the Stiefel manifold as a model in Shape Theory.

1.5 Notations and main definitions

In the following any Hilbert space H will be assumed to be a real separable Hilbert space, with norm $\|\cdot\|_H$ and scalar product $\langle\cdot,\cdot\rangle_H$.

Given $v \in H$, we will denote by v^* the continuous linear functional $v^*(x) = \langle v^*, x\rangle$.

By *"manifold"* we will mean a *"smooth connected second countable boundaryless Hausdorff differentiable manifold modeled on a Hilbert space"*.

If M is a Hilbert space, or a manifold modeled on a Hilbert space, we will associate to it the Borel sigma-algebra $\mathcal{B}(M)$.

By *"measure"* μ on M we will mean a countably additive map $\mu : \mathcal{B}(M) \to [0, \infty]$.

If N is another such set and $\psi : M \to N$ is a Borel-measurable transformation, then the *push forward* is the measure $\psi_{\#}\mu$ on N that is defined by $(\psi_{\#}\mu)(A) = \mu(\psi^{-1}(A))$ for all $A \in \mathcal{B}(N)$.

By *"probability measure"* μ on M (or more simply *"probability"*) we will mean a measure μ such that $\mu(M) = 1$.

When μ is a probability the *push forward* $\psi_{\#}\mu$ is a probability on N, and is usually called the *"distribution"* or the *"law"* of ψ on N.

Definition 1.2. Let μ and ν be measures on M.

- The measure ν is called *"absolutely continuous"* with respect to μ if $\nu(A) = 0$ for every set $A \in \mathcal{B}(M)$ with $\mu(A) = 0$. We will write $\nu \lll \mu$ in this case.
- The measures μ, ν are *"equivalent"* if they are mutually absolutely continuous. We will write $\mu \sim \nu$ in this case.
- The measures ν, μ are called *"mutually singular"* if there exists a set $\Omega \in \mathcal{B}(M)$ such that $\mu(\Omega) = 0$ and $\nu(M \setminus \Omega) = 0$.

2 Gaussian measures

The following is a short presentation of the theory of Gaussian Measures; more details may be found *e.g.* in [4] and [7].

2.1 Gaussian measures

We recall a few facts about Gaussian measures in Hilbert spaces.

Definition 2.1. A probability measure γ on $\mathbb{R}$ is said to be Gaussian if it is either a Dirac measure, or has density

$$x \to \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \tag{1}$$

with respect to the Lebesgue measure for some parameters σ, $m \in \mathbb{R}$. In the first case the measure is called degenerate.

Definition 2.2. Let H be a Hilbert space. A measure γ on H is said to be Gaussian if for all $x \in H$ the push forward measure ${x^*}_{\sharp}\gamma$ is a Gaussian measure on $\mathbb{R}$ (possibly degenerate).

This definition coincides with the usual definition when $H = \mathbb{R}^n$.

The characteristic functional (*i.e.* Fourier transform) of a probability μ on H is

$$\widehat{\mu} : H^* \to \mathbb{C} \quad , \quad \widehat{\mu}(f) := \int_H \exp\big(if(y)\big) \mathrm{d}\mu(y) \ .$$

Theorem 2.3 (Theorem 2.3.1 in [4], or Theorem 1.12 in [7])**.** *Let γ be a Gaussian measure on a Hilbert space H. Then there exist a vector $m \in H$ and a symmetric nonnegative nuclear (*i.e. *trace class) operator K such that*

$$\widehat{\gamma}(f) = \exp\Big(i \langle m, x\rangle_H - \langle Kx, x\rangle_H\Big) \ . \tag{2}$$

Viceversa, given any m and K as above, there exists a unique Gaussian measure γ satisfying (2)*.*

We recognize that m is **the mean** and K is **the covariance operator** of γ, in this sense. Given $v, w \in H$, we have that $v^*, w^* \in L^2(H, \gamma)$ and that the mean and covariance are

$$\begin{aligned} \mathbb{E}[v^*] \quad &:= \quad \int_H v^*(x)\, \mathrm{d}\gamma(x) \ = \ \langle m, v\rangle_H \quad , \\ \mathrm{Cov}[v^*, w^*] \quad &:= \quad \int_H v^*(x-m)\, w^*(x-m)\, \mathrm{d}\gamma(x) \ = \ \langle Kv, w\rangle_H \ . \end{aligned}$$

For this reason we will indicate γ with the usual notation $N(m, K)$. When m is zero, we say that γ is *centered*. When the kernel of K is $\{0\}$, we say that γ is *non degenerate*.

The following proposition is an intermediate step in the proof of the above theorem.

Proposition 2.4. *Every Gaussian measure $\gamma = N(m, K)$ on a Hilbert space H has second moment, more precisely*

$$\int_H \|x - m\|_H^2 \,\mathrm{d}\gamma(x) = \mathrm{Tr}(K) < \infty \ .$$

By choosing an appropriate Hilbertian base, a Gaussian measure can be seen as a process of independent real Gaussian random variables.

Proposition 2.5. *Let $\gamma = N(m, K)$ be a Gaussian measure on a Hilbert space H. Consider an orthonormal complete basis $(e_n)_{n\in\mathbb{N}}$ of H that diagonalizes the operator K. Then the coordinate functions e_n^* are independent.*

In particular, if $m_n = \langle m, e_n\rangle \in \mathbb{R}$ and $\sigma_i \geq 0$ is the eigenvalue such that $Ke_n = \sigma_n e_n$, then $e^*_{n\#}\gamma \sim N(m_n, \sigma_n)$, that is, m_n is the mean and σ_n the variance of the real Gaussian random variable e_n^*. Moreover $\sum_{n=0}^{\infty} \sigma_n = \mathrm{Tr}(K) < \infty$. Obviously γ is *non degenerate* iff $\sigma_n > 0$ for all n.

2.2 Cameron–Martin theory

We now introduce the Cameron–Martin theory, using a simplified approach, as in Chap. 2 in [7].

Definition 2.6 (Cameron–Martin space)**.** Let γ be a centered Gaussian measure on a Hilbert space H, let K be its covariance. The *Cameron–Martin space* $\mathrm{CM}(\gamma)$ of γ is the range (*i.e.* the image) of $K^{1/2}$. In symbols,

$$\mathrm{CM}(\gamma) = K^{1/2}(H) \quad .$$

The above may be expressed as follows. Assume for simplicity that the measure is non degenerate. Let $(v_n)_n$ be the orthonormal basis of eigenvectors of K, so that the coordinate functions v_n^* are independent (as by Prop. 2.5). Let $a_n > 0$ be the variance of v_n^*, *i.e.* the eigenvalue associated to the eigenvector v_n.

In this case we have that

$$\sqrt{\sum_{n=0}^{\infty} \frac{1}{a_n} | \langle v_n, x \rangle_H |^2} = \|K^{-1/2} x\|_H \quad ;$$

moreover the left hand side is finite if and only if $x \in \mathrm{CM}(\gamma)$.

Note that $\mathrm{CM}(\gamma) = H$ if and only if H is finite dimensional. In the infinite dimensional case, $\mathrm{CM}(\gamma)$ is dense in H, but its γ-measure is zero.

Definition 2.7 (White noise mapping)**.** Consider the mapping

$$W : \mathrm{CM}(\gamma) \to L^2(H, \gamma) \quad , \quad z \mapsto W_z$$

where W_z is defined by $W_z(x) = \left\langle x, K^{-1/2} z \right\rangle_H$. This mapping is an isometry from $\mathrm{CM}(\gamma)$ (with the norm of H) to $L^2(H, \gamma)$, so it extends to an unique mapping $W : H \to L^2(H, \gamma)$, that is called the *white noise mapping.*

Proposition 2.8 (Cameron–Martin theorem – translation of Gaussian measures)**.** *Let γ be a centered non degenerate Gaussian measure on a Hilbert space H. Let $h \in H$, and $\mu = \gamma(\cdot - h)$ be the translation of γ.*

- *If $h \in CM(\gamma)$ then μ and γ are equivalent, and the Radon–Nicodým derivative is given by the Cameron–Martin formula*

  $$\frac{d\mu}{d\gamma}(x) = \exp\Big(-\frac{1}{2}\|a\|_H^2 + W_a(x)\Big)$$

  where $a = K^{-1/2}h$. Note that the term $W_a(x) = \left\langle K^{-1}h, x\right\rangle_H$ in the finite dimensional case.

- *Moreover the total variation distance is bounded by*

  $$\|\gamma - \mu\|_{TV} \le 2\sqrt{1 - \exp\Big(-\frac{1}{4}\|a\|_H\Big)} \quad .$$

- *If $h \notin CM(\gamma)$ then μ and γ are mutually singular.*

The above is a combination of results in Chap. 1 and 2 in [7], and in Chap. 2 Sect. 4 in [4].

3 Image of a probability measure under the exponential map

A possible way to define a probability measure on a Riemannian manifold M is to choose a point $p \in M$, define a probability measure γ on the tangent space T_pM in p and then push forward γ under the exponential map to define the desired probability on M.

We recall briefly the definition of the exponential map. More details may be found in [15]. The exponential map $\exp_p\colon T_pM \to M$ is defined as

$$\exp_p(v) = \sigma_v(1)$$

where σ_v is the geodesic starting from $\sigma_v(0) = p$ with tangent vector $\dot\sigma_p(0) = v$.

If M is a finite dimensional complete Riemannian manifold, then the exponential map from any point is surjective; this result is part of the Hopf-Rinow theorem (see Theorem 2.8, Chapter 7 of [8]).

If M is an infinite dimensional complete Riemannian manifold, then the exponential map may fail to be surjective [2]. This is a first problem in applying the above idea.

Moreover the resulting measure $\exp_{p\,\sharp}\gamma$ on M depends also on the point p and, since there is no natural way to compare the tangent spaces, it could be difficult to compare measures obtained starting from different points.

3.1 Finite dimensional manifolds

Proposition 3.1. *Let M be a complete n-dimensional Riemannian manifold. Let $p \in M$ a point and T_pM the tangent space to M in p. Let also γ be a measure on T_pM, absolutely continuous* wrt *the Lebesgue measure $\mathscr{L}^n$. Then its push forward $\mu = \exp_{p\#}\gamma$ under the exponential map is absolutely continuous* wrt *the Hausdorff measure. In symbols*

$$\gamma \ll \mathscr{L}^n \Rightarrow \exp_{p\#}\gamma \ll \mathcal{H}^n \ .$$

If moreover γ is equivalent to the Lebesgue measure, then μ is equivalent to the Hausdorff measure. In symbols

$$\gamma \sim \mathscr{L}^n \Rightarrow \exp_{p\#}\gamma \sim \mathcal{H}^n \ .$$

Proof. Suppose that $f : T_pM \to M$ is a C^1 map; let C_f be the set of critical points of f, that is the set of $x \in T_pM$ such that the differential Df is not invertible at x. We will use the *"change of variable"* Lemma 5.5.3 in [1]. The first point states that $f_\#\mathscr{L}^n$ is absolutely continuous *wrt* $\mathcal{H}^n$ if and only if $\mathscr{L}^n(C_f) = 0$. Let now $f = \exp_p$ so that $\mu = f_\#\mathscr{L}^n$. Let us divide $C_f = \bigcup_{i=0}^{n-1}\Gamma_i$ with

$$\Gamma_i = \{x \in T_pM \ \mid Df(x) \text{ has rank } i\} \quad .$$

The Theorem 4.4 of [18] proves that each of the above sets Γ_i is locally contained in a $(n-1)$-dimensional submanifold of T_pM. So $\mathscr{L}^n(C_f) = 0$.

Suppose now moreover that γ is equivalent to the Lebesgue measure, we want to prove that μ is equivalent to the Hausdorff measure $\mathcal{H}^n$. By the previous point, it is enough to prove that $\mathcal{H}^n$ is absolutely continuous *wrt* μ, *ie* $\mathcal{H}^n \ll \mu$. We will use some facts that are explained in [17]. Let K_p be the cutlocus of the point p, let $\Omega_p = M \setminus K_p$, that is an open set. It was proven in [17] that $\mathcal{H}^n(K_p) = 0$, so we will ignore K_p in the following. Let $E \subset \Omega_p$ be a Borel set such that $\mu(E) = 0$. Let $B = f^{-1}(E)$, then by definition of push forward $\mathscr{L}^n(B) = 0$. There exists an open star-shaped set $O \subseteq T_pM$ such that, calling g the exponential map $\exp_p$ restricted to O, the map $g : O \to \Omega_p$ is a diffeomorphism. Let now $D = B \cap O$, then g maps bijectively D onto E. Obviously $\mathscr{L}^n(D) = 0$. The second point in the above Lemma shows then that $\mathcal{H}^n(E) = 0$. $\square$

Corollary 3.2. *Let us consider, for $i = 1, 2$, a point $p_i \in M$, a probability measure μ_i defined on $T_{p_i}M$ that is equivalent to the Lebesgue measure on $T_{p_i}M$. Let $\nu_i = \exp_{p_i} \# \mu_i$ be the wrapping of μ_i on M. Then the two measures ν_1, ν_2 are equivalent.*

3.2 Infinite dimensional manifolds

If the manifold M is infinite dimensional, one can wonder if there could be a similar result. In the finite dimensional case, we compare measures on different tangent spaces by relating them with the Lebesgue measure, that can be defined in a standard way on all tangent spaces. The first question to be answered when trying to discuss Prop. 3.1 in the infinite dimensional setting, is how to compare measures on different tangent spaces.

One tool to address the problem is to connect points using a geodesic, and push forward the measure on the tangent space using the parallel transport. This was the method proposed in [19] when devising a discrete stochastic process on the Stiefel manifold $\mathrm{St}\left(2, L^2\right)$, to be used as a model for tracking shapes enclosed by curves. In that case, the geodesic was provided by the model itself. In general, though, this method has two drawbacks. One is that there may be no geodesic connecting two points (even if the manifold is metrically complete [2]). The opposite drawback is that there may be multiple geodesics connecting a pair of points, and so there may be no canonical choice.

Another possible tool to address this problem is a group of transformations that acts transitively on M, if one is available. Again, a drawback is that there may be multiple transformations moving a point to another. (Unless the manifold is also a Lie group, of course).

To simplify utterly the matter, we will study the case of $M = S$, where S is the unit sphere in an infinite dimensional Hilbert space. We associate to S the group of unitary transformations, that we call “rotations” for simplicity. In this case the parallel transport coincides with the tangent map of a suitable rotation.

We will in the following show in Theorem 3.7 that, if we wrap a Gaussian measure around the sphere S, and then we rotate it, then the two measures on the sphere are mutually singular. We can prove this fact for a class of rotation, that are described in the statement of Theorem 3.7.

We will prove the following results assuming that Gaussian measures are non degenerate. These results hold more in general for the case of Gaussian measures that are not concentrated on finite dimensional spaces. Indeed if a Gaussian measure is supported on an infinite dimensional space, then we may restrict the following analysis to that space, and obtain a non degenerated Gaussian measures on an infinite dimensional Hilbert space.

We first state a few results and observations, which are useful to prove the following results.

Lemma 3.3 (Law of large numbers)**.** *Let H be a Hilbert space, and γ be a non degenerate Gaussian measure on it. Let v_n be the eigenvectors of the covariance operator K, and σ_n^2 be the corresponding eigenvalues. Let $f_n = v_n^*/\sigma_n$, that is, $f_n(x) = \langle v_n, x\rangle/\sigma_n$ so that the random variable f_n has standard Gaussian distribution $N(0,1)$. Since the joint distributions of $(f_1, \dots f_n)$ is centered Gaussian, then orthogonality implies independence. So their squares f_i^2 are a sequence of independent, identically distributed random variables each with chi-squared distribution (with 1 degree of freedom) and having mean 1 and variance 2. By the law of large numbers (Theorem 3.27 in [5]), γ is concentrated on the Borel set*

$$C = \left\{ x \in T \;\middle|\; \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_i^2(x) \;=\; 1 \right\} \quad ,$$

(a point x such that the above limit does not exists is not in C). This set C has some peculiar properties.

- *For every vector $x \in H$ there exist either two or no values $\lambda \in \mathbb{R}$ such that $\lambda x \in C$; if there are two values, they have opposite sign. So this set is quite "thin" in the radial directions.*
- *At the same time, for any r in the Cameron–Martin space $CM(\gamma)$ of γ, and for any $v \in C$, then $v + r \in C$. In symbols,*

  $$C + CM(\gamma) = C \quad .$$

  So the set C is quite "large" in many linear directions.

Proof. We prove the second point. Suppose for simplicity that $H = \ell^2$, and that K is diagonal, so when $x = (x_n)_{n\in\mathbb{N}}$ we identify $f_n(x) = x_n/\sigma_n$. Let $\tilde{x} = x + r$, so forall i

$$\tilde{x}_i = x_i + r_i \ , \quad |\tilde{x}_i|^2 = |x_i|^2 + |r_i|^2 + 2 r_i x_i \ ,$$

We have to deal with the three terms in the right hand side.

Since r is in the Cameron–Martin space, by definition $\sum_{k=0}^{\infty} \frac{|r_k|^2}{\sigma_k^2} < \infty$, then $\lim_{k\to\infty} \frac{|r_k|^2}{\sigma_k^2} = 0$ so by Cesaro's lemma

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{|r_i|^2}{\sigma_i^2} = 0 \quad . \tag{3}$$

We know that the variables x_i are independent *wrt* γ. Note that $r_i x_i/\sigma_i^2 \sim N(0, r_i^2/\sigma_i^2)$; since $\sum_{i=0}^{\infty} r_i^2/\sigma_i^2 < \infty$ then the sequence r_i^2/σ_i^2 is bounded, so by the law of large numbers

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{r_i x_i}{\sigma_i^2} = 0$$

for γ-almost any x. Similarly, again by the law of large numbers, for γ-almost every x,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{|x_i|^2}{\sigma_i^2} = 1 \quad .$$

Summing up we obtain the desired result. □

Lemma 3.4. *Let γ be a centered non degenerate Gaussian measure on a separable Hilbert space H. Then every sphere has measure zero.*

Proof. Let $S_r = \{x \in H \mid |x|_H = r\}$ be a sphere of radius r and $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis of H such that the coordinates functions are independent. Such a basis exists by Corollary 2.5. Consider the orthogonal decomposition

$$H = \mathrm{Span}(e_1) \times H',$$

where $H' = \mathrm{Span}(e_2, e_3, \dots)$ and let π be the orthogonal projection on H'. By the independence of the coordinate functions, γ can be decomposed as

$$\gamma = e^*_{1\sharp}\gamma \otimes \pi_\sharp\gamma.$$

We compute the measure of the sphere using Fubini's theorem for the product measure $e^*_{1\sharp}\gamma \otimes \pi_\sharp\gamma$. For every $x \in H'$, there are at most two x_1 such that $(x_1, x') \in S_r$. Since $e^*_{1\sharp}\gamma$ is a Gaussian measures on $\mathbb{R}$, finite sets are negligible with respect to it. It follows that S_r is negligible for γ, since every slice at $x' \in H'$ fixed is negligible with respect to $e^*_{1\sharp}\gamma$. □

It is worth nothing this fact.

Lemma 3.5. *Let H be a Hilbert space and $S \subseteq H$ the unit sphere in H. Fix $p \in S$, and $\exp_p$ the exponential map. Let A be a Borel subset of the tangent space T_pS. Then the image $\exp_p(A)$ is a Borel subset of S.*

The proof is based on the very simple structure of the exponential map of the sphere (see Equation (4)), we omit it.

We now provide a simpler case of the following Theorem 3.7. This case can help understanding the spirit of the proof of the theorem.

Proposition 3.6. *Let H be a separable infinite dimensional Hilbert space and $S \subseteq H$ the unit sphere in H. Consider a pair of a points $p \in S$ and $-p$. The tangent spaces can be seen as subsets of H, and both can be identified with the subspace*

$$T_pS = T_{-p}S = T = \{x \in H \mid \langle x, p\rangle = 0\} \ .$$

Consider a centered non degenerate Gaussian measure γ on T. Then the measures $\exp_{p\,\sharp}\gamma$ and $\exp_{-p\,\sharp}\gamma$ are mutually singular.

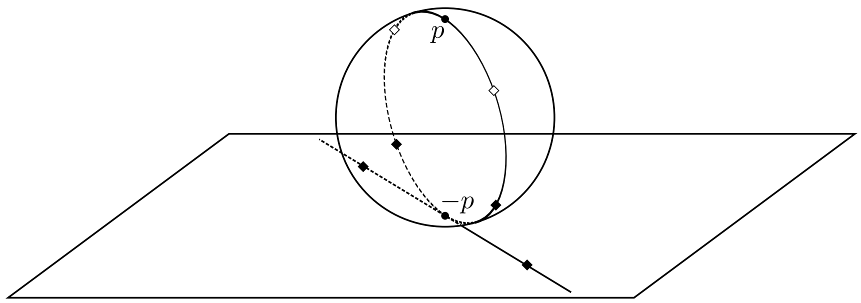

Figure 1: Proof of Proposition 3.6. Two points on the ellipsoid C, their images under $\exp_{-p}$ and, in white, their images under $\exp_p$. The black diamonds on the sphere can coincide with the white diamonds only if they all lay on the equator.

Proof. By Proposition 2.5 in the Hilbert space $L^2(\gamma)$ there exists an orthonormal sequence $\{f_i\}_{i\in\mathbb{N}}$ of continuous linear functional on H.

Reasoning as in Lemma 3.3, we obtain that γ is concentrated on the set

$$C = \left\{ x \in T \;\middle|\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_i^2(x) \;=\; 1 \right\} .$$

We also know that, for every direction $x \in T$ there exist either two or no values $\lambda \in \mathbb{R}$ such that $\lambda x \in C$ and, if there are two, they have opposite sign.

Call μ_1, μ_2 the push forward of γ under $\exp_p$ and $\exp_{-p}$

$$\mu_1 = \exp_{p\,\sharp}\gamma \qquad \mu_2 = \exp_{-p\,\sharp}\gamma\,.$$

Call C_1, $C_2 \subseteq S$ the images of C under $\exp_p$ and $\exp_{-p}$. By the previous Lemma the sets C_1, C_2 are Borel sets. Clearly, μ_1 is concentrated on C_1 and μ_2 is concentrated on C_2.

To prove that μ_1 and μ_2 are mutually singular it is sufficient to show that $C_1 \cap C_2$ is negligible for one of them.

The exponential maps from the points p and $-p$, defined $T \to S$, could be written as

$$\begin{aligned} \exp_p(x) &= \cos(|x|)p + \sin(|x|)\frac{x}{|x|} \\ \exp_{-p}(x) &= -\cos(|x|)p + \sin(|x|)\frac{x}{|x|} \end{aligned} \tag{4}$$

and are symmetric with respect to the reflection through T. From this symmetry and the fact that for each line through the origin in T, if there is one point in C on that line then there are exactly two opposite in sign, it follows that $C_1 \cap C_2$ is contained in $T \cap S$ (see also Figure 1).

The equator $T \cap S$ is negligible for μ_1 (and also for μ_2), indeed, denoted by S_r the sphere of radius r in T,

$$\mu_1(T \cap S) = \gamma(\exp_p^{-1}(T \cap S)) = \gamma\left(\bigcup_{k=1}^{+\infty} S_{k\pi+\frac{\pi}{2}}\right) = \sum_{k=0}^{+\infty} \gamma(S_{k_\pi+\frac{\pi}{2}})$$

and all those spheres are negligible by Lemma 3.4. □

We now come to the general result.

Theorem 3.7. *Suppose that H is a separable Hilbert space; let S be the unit sphere. Let $p \in S$. Let γ be a centered non degenerate Gaussian measure on T_pS. Let $\mu = \exp_{p\#}\gamma$ the wrapping of γ on S. Let $r \in T_pS, r \neq 0$ be a vector that is in the Cameron–Martin space of γ. Let $\widehat{pr}$ be the plane spanned by p, r. We define a rotation $R : H \to H$ in this way: R rotates any vector in the plane $\widehat{pr}$ by a fixed angle, whereas R keeps fixed any vector orthogonal to $\widehat{pr}$. Suppose that R is not the identity map. Let $\nu = R_\#\mu$. Then μ, ν are mutually singular.*

The rotation R rotates p in the direction r, as to say. So we may think of R as a "Cameron–Martin rotation". This would mislead us into thinking that ν and μ be equivalent. Instead they are mutually singular.

We remark this fact.

Remark 3.8. Let $q = Rp$, suppose for simplicity that $q \neq -p$. Let ξ be the unique minimal geodesic connecting p to q. Define the tangent map $\tilde{R} = D_pR : T_pS \to T_qS$, then $\tilde{R}$ coincides with the parallel transport along ξ. Let $\tilde{\gamma} = \tilde{R}_\#\gamma$, let then $\tilde{\mu} = \exp_{q\#}\tilde{\gamma}$ the wrapping of $\tilde{\gamma}$ on S. Then $\tilde{\mu} = \nu$. So the probability ν is also obtained by identifying $T_pS \to T_qS$ using parallel transport, and then wrapping.

We will need the following Lemma.

Lemma 3.9. *Suppose that E is a Hilbert space, γ is a non degenerate Gaussian measure on E, V is a finite dimensional subspace of the Cameron–Martin space of γ. Let $H = V \oplus V^\perp$ be the standard decomposition. We decompose any $x \in H$ as $x = y + z$ with $y \in V, z \in V^\perp$. Let $\pi_{V^\perp}$ the orthogonal projection on $V^\perp$; let $\tilde{\gamma} = \pi_{V^\perp\#}\gamma$ be the projection of γ on $V^\perp$. In this setting there is a family ν_z, for $z \in V^\perp$, with the following properties: each ν_z is a non degenerate Gaussian measure on V; the family ν_z is the conditional distribution of y knowing z, that is, for any continuous bounded f,*

$$\int_E f(x)\,\mathrm{d}\gamma(x) = \int_{V^\perp}\left(\int_V f(y+z)\,\mathrm{d}\nu_z(y)\right)\mathrm{d}\tilde{\gamma}(z) \quad .$$

The above results are proved in Section 3.10 in [4]. It is interesting to note this fact: if V is not contained in the Cameron–Martin space of γ, then the conditional measures ν_z exist, but they are concentrated on single points.

We now provide the proof of Theorem 3.7.

By using an appropriate choice of Hilbertian base for the space H, we can rewrite the hypotheses of the theorem as follows. Let $H = \ell^2$ for simplicity. We will denote by e_n the canonical coordinate vectors. Let S be the unit sphere in H. We assume that $\|r\|_H = 1$ for simplicity. Let $\theta \in (0, \pi/2)$ be fixed. (We exclude the case $\theta = 0$, when R is the identity map; we also exclude for simplicity the case $\theta = \pi/2$, in this case R is the antipodal map $Rx = -x$, and this case is equivalent to the case discussed in Proposition 3.6 — note anyway that the result may be proved using the following analysis, paying attention to some details).

Let $p, q \in S$ be given by

$$\begin{array}{rcl} p &=& e_1 \cos\theta + r \sin\theta \quad , \\ q &=& e_1 \cos\theta - r \sin\theta \quad . \end{array}$$

These are the endpoints of the geodesic

$$\exp_{e_1}(tr) = e_1 \cos t + r \sin t$$

for times $t = \pm\theta$. We also define

$$\tilde{p} = -e_1 \sin\theta + r \cos\theta \quad , \quad \tilde{q} = -e_1 \sin\theta - r \cos\theta \quad ;$$

these are the speeds of the above geodesic at $t = \pm\theta$. Note that the plane spanned by p, q is also the plane spanned by e_1, r; we call this plane V. Moreover the spaces $V \cap T_p M$, $V \cap T_{e_1} M$ and $V \cap T_q M$ are one-dimensional, and are spanned by the vectors $\tilde{p}$, r and $\tilde{q}$ respectively.

We define the rotation R by stating that R is the identity for any vector in $V^\perp$, whereas it rotates vectors in the plane V by the angle θ (so that $Re_1 = p$ and $Rq = e_1$). Let $\tilde{R}_p : T_{e_1} S \to T_p S$ and $\tilde{R}_q : T_q S \to T_{e_1} S$ be the tangent maps.

We assume that γ is a probability measure on $T_{e_1} S$, and that the covariance operator K is diagonal in the standard base $\{e_2, e_3, \ldots\}$ of $T_{e_1} S$. Let σ_k^2 be the eigenvalue of K in direction e_k.

We push forward γ to γ_p using $\tilde{R}_p$, and pull it back to γ_q using the inverse of $\tilde{R}_q$. Note that $\tilde{R}_p r = \tilde{p}$ while $\tilde{R}_q \tilde{q} = r$, so that $\tilde{p}$ is in the Cameron–Martin space of γ_p and $\tilde{q}$ is in the Cameron–Martin space of γ_q.

This setting mimics the hypotheses of the theorem, only in a more symmetric fashion. Indeed if $\mu = (\exp_q)_\# \gamma_q$ is the wrapping of γ_q and $\nu = (\exp_p)_\# \gamma_p$ is the wrapping of γ_p, then $(R^2)_\# \mu = \nu$.

In this setting we have a very powerful situation. Indeed $V^\perp \subset T_p S$, $V^\perp \subset T_q S$ and $V^\perp \subset T_{e_1} S$; moreover the projections of the three measures $\gamma_q, \gamma, \gamma_p$ on $V^\perp$ are identical.

We now consider two generic vectors $v \in T_p S$ and $w \in T_q S$. We decompose them (in an unique way) as

$$v = a\tilde{p} + \tilde{v} \quad , \quad w = b\tilde{q} + \tilde{w}$$

with $a,b \in \mathbb{R}$ and $\tilde{v}, \tilde{w} \in V^\perp$. (Obviously $a = \langle \tilde{p}, v\rangle_H$ and $b = \langle \tilde{q}, w\rangle_H$). The joint distribution of $(a, \tilde{v})$ according to γ_p is the same as the joint distribution of $(b, \tilde{w})$ according to γ_q. In particular, by what we said above, $\tilde{v}, \tilde{w}$ are identically distributed. Similarly a, b are real-valued marginals and have the same non-degenerate centered Gaussian distribution on $\mathbb{R}$.

For simplicity we will abbreviate $t = \|v\|_H, s = \|w\|_H$.

The above quantities are related by $a \in [-t,t], b \in [-s,s]$ and

$$t^2 - a^2 = \|\tilde{v}\|_H^2 \quad , \quad s^2 - b^2 = \|\tilde{w}\|_H^2 \quad . \tag{5}$$

We can assume $t > 0, s > 0, a \notin \{-t, 0, t\}, b \notin \{-s, 0, s\}$ (the complementary choices correspond to negligible sets in the following reasoning).

We define $\operatorname{sinc}(x) = \sin(x)/x$. This function is analytic.

The exponential map from p (resp. q) in direction v (resp. w) is

$$\exp_p(v) = p\cos(t) + v\operatorname{sinc}(t) \quad \text{resp.} \quad \exp_q(w) = q\cos(s) + w\operatorname{sinc}(s) \quad .$$

We can express them in the decomposition $V \oplus V^\perp$

$$\begin{aligned} \exp_p(v) &= \Big(p\cos(t) + a\tilde{p}\operatorname{sinc}(t)\Big) + \tilde{v}\operatorname{sinc}(t) \ , \\ \exp_q(w) &= \Big(q\cos(s) + b\tilde{q}\operatorname{sinc}(s)\Big) + \tilde{w}\operatorname{sinc}(s) \ . \end{aligned}$$

If these maps reach the same point, then

$$p\cos(t) + a\tilde{p}\operatorname{sinc}(t) = q\cos(s) + b\tilde{q}\operatorname{sinc}(s) \tag{6}$$

$$\tilde{v}\operatorname{sinc}(t) = \tilde{w}\operatorname{sinc}(s) \quad . \tag{7}$$

We know that $\tilde{v} = v - a\tilde{p}$ and that $\tilde{p}$ is in the Cameron–Martin space of γ_p. By applying Lemma 3.3 we assert that

$$\lim_{j\to\infty} \frac{1}{j}\sum_{i=1}^{j} \frac{|\tilde{v}_i|^2}{\sigma_i^2} = 1 \quad .$$

for γ_p-almost any v. The same holds for w as well, *mutatis mutandis*. By Lemma 3.4 we can assume that $\operatorname{sinc}(s) \neq 0$ and $\operatorname{sinc}(t) \neq 0$. So for γ_p-almost all v and γ_q-almost all w,

$$\lim_{j\to\infty} \frac{1}{j}\sum_{i=1}^{j} \frac{(\tilde{v}_i)^2}{\sigma_i^2} = 1 = \lim_{j\to\infty} \frac{1}{j}\sum_{i=1}^{j} \frac{(\tilde{w}_i)^2}{\sigma_i^2} = \frac{\operatorname{sinc}^2 t}{\operatorname{sinc}^2 s} \lim_{j\to\infty} \frac{1}{j}\sum_{i=1}^{j} \frac{(\tilde{v}_i)^2}{\sigma_i^2} \quad , \tag{8}$$

where the last equality comes from eqn. (7). So we obtain $\operatorname{sinc}(t) = \pm\operatorname{sinc}(s)$, so $\tilde{v} = \pm\tilde{w}$ by equation (7).

We will now use this fact to the best. Foremost, we elaborate on the equation (6). We know that the frame $p, \tilde{p}$ is obtained by $q, \tilde{q}$ by rotating by an angle 2θ. So

$$\begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) \end{pmatrix} \begin{pmatrix} \cos t \\ a\operatorname{sinc} t \end{pmatrix} = \begin{pmatrix} \cos s \\ b\operatorname{sinc} s \end{pmatrix} \quad . \tag{9}$$

Let

$$E = E_0 \cup E_1 \quad , \quad E_0 = \{k\pi : k \in \mathbb{N}\} \quad E_1 = \{x > 0 : x = \tan x\} \quad .$$

The points in E_0 are all the positive zeros of sinc, while the points in E_1 are all the positive zeros of its derivative sinc$'$. Let $(I_n)_{n\in\mathbb{N}}$ be an enumeration of all open intervals that constitute the complement of E in $(0, \infty)$. We have that $I_0 = (0, \pi)$; when $n \geq 1$, I_n has one endpoint in the set E_0 while the other endpoint is in the set E_1. On these intervals $\operatorname{sinc}(s)$ is either always positive or always negative, and is monotonic.

We fix $n, k \in \mathbb{N}$. We restrict our attention to the case $t \in I_n$ and $s \in I_k$. Then there is a function $\varphi = \varphi_{n,k}$ with these characteristics: φ is a homeomorphism between maximal subintervals of I_n, I_k; each one of these subintervals has a zero of sinc as one of its endpoints; φ and its inverse are analytic; when $t \in I_n$ and $s \in I_k$ the relation $\operatorname{sinc}(t) = \pm \operatorname{sinc}(s)$ holds if and only if $s = \varphi(t)$.

Recall that v is distributed according to γ_p. By Lemma 3.9, for almost any $\tilde{v}$, the conditional distribution of a is Gaussian and non degenerate. Let us fix such a $\tilde{v}$.

From (9) we extract the identity

$$\cos(2\theta)\cos(t) + a\sin(2\theta)\operatorname{sinc}(t) - \cos(\varphi(t)) = 0 \quad ,$$

where $t = \sqrt{a^2 + \|\tilde{v}\|^2}$. The left hand side is an analytic function of a. If we move a so that t converges to a zero of $\operatorname{sinc} t$, then $s = \varphi(t)$ has to converge to a zero of $\operatorname{sinc}(s)$, so both converge to an integer multiple of π: hence the above left hand side converges

$$\pm \cos(2\theta) \pm 1$$

that is never zero. So that function is not identically zero, hence it has at most countably many zeros. Then the probability of this event is null.

This ends the proof.

Remark 3.10. Suppose in the above proof that r is not in the Cameron–Martin space. As we remarked that Lemma 3.9, in this case conditional measures ν_z exist, but they may be concentrated on single points. So the above proof cannot be easily adapted to the case when r is not in the Cameron–Martin space.

4 Push forward of a probability measure under a projection

A simple way to define a probability measure on a manifold M is to choose a probability space $(X, \mathcal{F}_X, \mathbb{P})$, a measurable map $f: X \to M$ and endow M with the push forward measure $f_\sharp\mathbb{P}$.

Example 4.1. Let $S^n \subseteq \mathbb{R}^{n+1}$ be the n-dimensional unit sphere and γ a Gaussian measure on $\mathbb{R}^{n+1}$ with mean 0 and covariance operator the identity. Consider the projection

$$\begin{array}{rccc} \pi: & \mathbb{R}^{n+1} \setminus \{0\} & \to & S^n \\ & x & \mapsto & \frac{x}{|x|} \end{array} ,$$

which is defined γ almost everywhere. Then the measure $\pi_\sharp\gamma$ on S^n coincides with the Hausdorff measure $\mathcal{H}^n$ restricted to the sphere and normalized.

4.1 Finite dimensional manifolds

The above example can be properly generalized, provided that we define a "projection". One easy way to define the projection is by looking at a point of minimum distance. To this end, in this section we consider a closed subset M of a complete finite dimensional Riemannian manifold N. Let d be the Riemannian distance on N and $d_M : N \to \mathbb{R}$ the distance from the set M, defined by

$$d_M(x) = \inf_{y\in M} d(x, y)\,. \tag{10}$$

Since M is closed and N is locally compact, the infimum is a minimum, and then for all $x \in N$ there exists a point $y \in M$ such that $d(x, y) = d_M(x)$. However there may be more than one such point. For those point x such that the closest point y in M is unique, we denote this point by $\pi(x) = y$, so that

$$d(x, \pi(x)) = d_M(x)\,.$$

Proposition 4.2. *Let M be a closed set in a complete m-dimensional Riemannian manifold N. Then for almost every x there exists a unique point $\pi(x) \in M$ that realizes the minimum of the distance from x.*

So, given a measure γ which is locally absolutely continuous with respect to the Lebesgue measure, the measure $\pi_\sharp\gamma$ is well defined on M.

Proof. Here is a sketch of the proof, the detailed arguments may be found in [17] and references therein. The distance function d_M is Lipschitz. At all points where d_M is differentiable, the projection point is unique. Let Σ be the set where d_M is not differentiable. By Rademacher Theorem Σ is negiligible. □

In the case when M is a smooth submanifold, moreover, Σ and its closure both have Hausdorff dimension at most $m-1$; see [17]. So the projection is well defined (and smooth) on an open set with negligible complement.

4.2 Infinite dimensional manifolds

In the following we will only consider the case when M is embedded in an infinite dimensional Hilbert space H, for simplicity.

As in the finite dimensional case, the minimum point is almost surely unique when it exists.

Proposition 4.3. *Let $M \subset H$ be a closed subset. Let d_M be defined as in equation (10) (by setting $d(x, y) = |x - y|_H$ as is usual). Let γ be a Gaussian measure on H. Then for γ-almost any x there is at most one point $y \in M$ at minimum distance from x.*

Proof. By Theorem 5.11.1 in [4], the set Σ where d_M is not Gâteaux differentiable has measure $\gamma(S) = 0$. The rest of the proof works as in the finite dimensional case. $\square$

If we now consider an infinite dimensional manifold, M embedded in a Hilbert space H, the projection on the manifold does not necessarily exist. An infinite dimensional Hilbert space is not locally compact, so there could be many points $x \in H$ for which there is no point on the manifold at minimal distance.

We first discuss a counterexample; in the next sections we will show some cases in which the projection can be defined.

Let H be a separable Hilbert space. Up to the choice of an orthonormal basis of H, we suppose (without loss of generality) that $H = \ell^2$.

Given a submanifold of H, we will denote by $d_M : H \to \mathbb{R}$ the distance from the manifold, defined as in the finite dimensional case by

$$d_M(x) = \inf_{y \in M} \|x - y\|_H \,.$$

Lemma 4.4. *Consider in $H = \ell^2$ the ellipsoid S defined by*

$$S = \left\{ x \in H \;\middle|\; \sum_{i=1}^{+\infty} a_i^2\, x_i^2 = c^2 \right\}$$

where $c \in \mathbb{R}$ is a positive number and $\{a_i\}_{i \in \mathbb{N}} \subseteq \mathbb{R}$ is a sequence of positive numbers increasing to 1

$$c > 0\,, \quad a_i \nearrow 1\,, \quad a_i > 0\,.$$

Then

1. *the set S is a closed submanifold of H,*
2. *the distance of the origin from S is $d_S(0) = c$,*
3. *there is no point on the ellipsoid at distance c from the origin.*

Proof. Define the continuous linear function $T : H \to H$ as

$$T : x \mapsto (a_i\, x_i)_{i \in \mathbb{N}}$$

and $f : H \to \mathbb{R}$ as $f(x) = |T(x)|^2$. The function f is continuous and differentiable with gradient

$$\nabla f(x) = 2\, T \circ T(x) = 2\, (a_i^2\, x_i)_{i \in \mathbb{N}}\,.$$

Note that the set S is the inverse image of c under the function f and so, since f is continuous, S is closed. To see that S is a submanifold of H, we can use the implicit function theorem, see [15] for a proof of the theorem in infinite dimension. Indeed, the gradient of f is null only in the origin and the origin does not belong to the ellipsoid S, since $c \neq 0$.

For every point $x \in H$, using that $a_i < 1$, we get

$$f(x) \;=\; \sum_{i=1}^{+\infty} a_i^2\, x_i^2 \;<\; \sum_{i=1}^{+\infty} x_i^2 \;=\; |x|^2$$

and so for all $x \in S$,

$$|x| > c\,.$$

This says that there are no points on S at distance c from the origin and gives the bound

$$d_S(0) \geq c\,.$$

To get the other inequality, consider the points $ca_n^{-1}\, e_n$ for $n \in \mathbb{N}$,

$$d_S(0) \leq \inf_{n\in\mathbb{N}} \left|ca_n^{-1}\, e_n\right| = \inf_{n\in\mathbb{N}} ca_n^{-1} = c$$

since $a_i \nearrow 1$. □

Lemma 4.4 shows that, in a separable Hilbert space H, there exists a submanifold for which the distance from the origin does not have a minimum on the manifold. However this is not yet the desired counterexample, because a single point will usually be negligible for a measure and so the projection could still exist almost everywhere.

We now show that there are "many" other points for which there is no point on the manifold at minimal distance.

Lemma 4.5. *Let $\{a_i\}_{i\in\mathbb{N}}$, c and S be an ellipsoid and its parameters, satisfying the hypotheses of Lemma 4.4. Then for each x in the set*

$$E_S = \left\{ x \in H \;\middle|\; \sum_{i=1}^{+\infty} \left(\frac{1}{1-a_i^2}\right)^2 x_i^2 \;<\; c^2 \right\}$$

there is no point on S at minimal distance.

The idea of the proof is the following. Consider a point on one of the ellipsoid's axes, i.e. of the form $\lambda\, e_n$. Then there is only one reasonable point that could be at minimal distance from it, the point $ca_n^{-1}\, e_n$ (or $-ca_n^{-1}\, e_n$, if λ is negative). If λ is small, that point would be too far and it would be convenient to "go to infinity". A similar argument works for points that are linear combinations of the $e_1, \ldots, e_n$ for some $n \in \mathbb{N}$, by reasoning that the point at minimum distance, if it exists, should be a linear combination of $e_1, \ldots, e_n$ as well. For the other points, we show that there are no "reasonable" minima, meaning that the function to minimize has no stationary points on the ellipsoid.

Proof. First of all, observe that E_S is inside S, i.e.

$$\sum a_i^2\, x_i^2 \;<\; c^2 \quad \text{for all } x \in E_S$$

because $a_i < 1 < (1 - a_i^2)^{-1}$ for all $i \in \mathbb{N}$.

By symmetry, it sufficient to prove the lemma when x is such that $x_i \geq 0$ for all $i \in \mathbb{N}$. Fix one such x. It is enough to consider only points $y \in S$ such that $y_i \geq 0$ for all $i \in \mathbb{N}$.

Let $f\colon H \to \mathbb{R}$ be the function $f(y) = \sum a_i^2\, y_i^2$. As noted in Lemma 4.4, $S = f^{-1}(\{c^2\})$, f is differentiable and

$$\nabla f(y) = (a_i^2\, y_i)_{i \in \mathbb{N}} \,.$$

Let also $g\colon H \to \mathbb{R}$ be the square of the function we want to minimize on S, i.e. $g(y) = |y - x|^2$. The function g is differentiable as well,

$$\nabla g(y) = 2\,(y_i - x_i)_{i \in \mathbb{N}}$$

and the distance from x attains minimum on S if and only if g has minimum on S.

From differential calculus we know that, if z is a minimum for g on S, then $\nabla f(z)$ and $\nabla g(z)$ should be linearly dependent, namely there exists $\lambda \in \mathbb{R}$ such that

$$\lambda\, a_i^2\, z_i = z_i - x_i \quad \text{for all } i \in \mathbb{N}$$

or equivalently

$$x_i = \left(1 - \lambda\, a_i^2\right) z_i \,. \tag{11}$$

This equation gives us some information about λ. Since x_i and z_i are non negative

$$\lambda < \frac{1}{a_i^2} \quad \text{for all } i \text{ such that } x_i \neq 0 \,. \tag{12}$$

Suppose that the point x has infinitely many coordinates different from 0. Then, passing to the limit in Equation 12,

$$\lambda \leq 1 \,.$$

Compute $f(z)$ using Equation 11 to substitute the coordinates of z_i:

$$f(z) = \sum a_i^2 \left(\frac{1}{1 - \lambda\, a_i^2}\right)^2 x_i^2 \leq \sum \left(\frac{1}{1 - a_i^2}\right)^2 x_i^2 < c^2$$

since $a_i < 1$, $\lambda \leq 1$ and $x \in E_S$. On the other side z is on the ellipsoid, and so it should hold

$$f(z) = c^2$$

but this is not possible, and we can conclude that z does not exist.

It remains to consider the case where the coordinates of x are eventually null. Let $n \in \mathbb{N}$ be such that $x_m = 0$ for all $m > n$ and decompose every point $y \in H$ as $y = \bar{y} + \hat{y}$, where $\bar{y} \in \operatorname{Span}(e_1, \dots, e_n)$ and $\hat{y} \in \operatorname{Span}(e_1, \dots, e_n)^\perp$. A point y belongs to S if and only if

$$f(\bar{y}) \leq c^2 \quad \text{and} \quad \hat{y} \in S(\bar{y})$$

where $S(\bar{y})$ is an ellipsoid defined by parameters $\{a_{n+1}, a_{n+2}, \dots\}$ and $\sqrt{c^2 - f(\bar{y})}$. To simplify notation, call $c_{\bar{y}}$ the number $\sqrt{c^2 - f(\bar{y})}$.

Compute the infimum of g on S minimizing first in $\hat{y}$ and then in $\bar{y}$:

$$\inf_{y \in S} g(y) = \inf_{f(\bar{y}) \leq c^2} \inf_{\hat{y} \in S(\bar{y})} \sum_{i=1}^{n} (y_i - x_i)^2 + \sum_{i=n+1}^{+\infty} y_i^2 =$$
$$= \inf_{f(\bar{y}) \leq c^2} \left(\sum_{i=1}^{n} (y_i - x_i)^2 + \inf_{\hat{y} \in S(\bar{y})} \sum_{i=n+1}^{+\infty} y_i^2 \right) .$$

The innermost inf is minimizing the square of distance from the origin on a ellipsoid if $c_{\bar{y}} > 0$ and is 0 if $c_{\bar{y}} = 0$. By Lemma 4.4 the infimum is equal to

$$\inf_{f(\bar{y}) \leq c^2} \sum_{i=1}^{n} (y_i - x_i)^2 + c_{\bar{y}}^2 = \inf_{f(\bar{y}) \leq c^2} \sum_{i=1}^{n} (y_i - x_i)^2 + c^2 - \sum_{i=1}^{n} a_i^2 \, y_i^2 \, . \tag{13}$$

The function in the above equation has a global minimum at the point $\bar{z}$ of coordinates

$$\bar{z}_i = \frac{x_i}{1 - a_i^2} \quad \text{for } i = 1, \dots, n \, .$$

Since $x \in E_S$, the equation of E_S gives that $\bar{z}$ is such that

$$f(\bar{z}) < c^2$$

and so $\bar{z}$ realizes the infimum in Equation 13.

Now we are nearly done, because if g has a minimum z on S then its first component in the decomposition should be $\bar{z}$. The second component should be not null and minimize the distance from the origin on a ellipsoid. This contradicts Lemma 4.4 and so g has no minimum. □

Now we state and prove that Theorem 4.2 is false in the case of an infinite dimensional Hilbert space with a Gaussian measure.

Theorem 4.6. *Let H be an infinite dimensional separable Hilbert space and γ a Gaussian measure on it. Then there exists a manifold S embedded and closed in H and a set E_S of positive γ-measure such that for every $x \in E_S$ the distance from x has no minimum on S.*

Proof. We assume that γ is centered and non degenerate. Choose an orthonormal basis $\{e_i\}_{i \in \mathbb{N}}$ of H that diagonalizes the covariance K of γ. The coordinate functions $x \mapsto x_i = \langle x, e_i \rangle_H$ are independent Gaussian random variables; we denote by σ_i^2 their variances; we know that $\sum_{i=1}^{+\infty} \sigma_i^2 < +\infty$ since K is trace class (see Proposition 2.4).

We look for an ellipsoid S that satisfies the thesis of the theorem. Consider an ellipsoid S depending on parameters $\{a_i\}_{i \in \mathbb{N}}$ and c that satisfy the hypothesis of Lemma 4.4. By the same lemma S is a manifold embedded and closed in H.

By Lemma 4.5 there exists a set E_S of points for which the minimum does not exist and, if $f\colon H \to \mathbb{R} \cup \{+\infty\}$ is the function

$$f(x) = \sum_{i=1}^{+\infty} \left(\frac{1}{1-a_i^2}\right)^2 x_i^2 \,,$$

the set E_S is defined by the equation

$$f(x) < c^2 \,.$$

The function f is positive and so its integral is

$$\int_H f(x)\, \mathrm{d}\gamma = \sum_{i=1}^{+\infty} \left(\frac{1}{1-a_i^2}\right)^2 \int_H x_i^2 \, \mathrm{d}\gamma = \sum_{i=1}^{+\infty} \left(\frac{1}{1-a_i^2}\right)^2 \sigma_i^2 \,.$$

Since $\sum \sigma_i^2$ is convergent, it is possible to choose a_i so that the above integral is finite. For this choice of a_i, the function $f(x)$ is finite γ almost everywhere and, up to negligible sets,

$$H = \bigcup_{n \in \mathbb{N}} \left\{ f(x) < n \right\} \quad ;$$

we choose c large enough so that E_S is not negligible for γ. $\square$

4.3 Stiefel manifolds

We have seen that, in general, we cannot “project” a Gaussian probability measure on a submanifold of an infinite dimensional Hilbert space. In this section though we will show that the projection onto Stiefel manifolds is almost everywhere well defined, for all possible choices of non degenerate Gaussian measures. So the “projection method” of considering the projection of a Gaussian measure from the ambient space to the manifold of interest is well defined when the manifold is an infinite dimensional Stiefel manifold. The simplest case of Stiefel manifold is the unit sphere.

Example 4.7. Let H be a separable Hilbert space and $S \subseteq H$ the unit sphere. Then the function

$$\pi\colon x \mapsto \frac{x}{|x|}$$

is defined in H minus the origin and it is the projection on the nearest point of the sphere S. If γ is a Gaussian measure on H and γ is not the Dirac delta centered in the origin, then the projection π is defined γ-almost everywhere.

4.3.1 The projection map

Let H be a separable Hilbert space and $\mathrm{St}\,(p, H)$ a Stiefel manifold embedded in H^p, as defined in Definition 1.1. We first characterize the points in H^p that admit projection on the Stiefel manifold and then prove that this set has full measure for every non degenerate Gaussian measure on H^p.

Given a point $x \in H^p$, we will denote by x_i his components, namely

$$x = (x_1, \ldots, x_p)$$

with $x_i \in H$.

Proposition 4.8. *Let H be a Hilbert space, $p \in \mathbb{N}, p \geq 1$, $\mathrm{St}\,(p, H) \subseteq H^p$ the Stiefel manifold and $x \in H^p$. Then*

1. *if the components $x_1, \ldots, x_p$ are linearly independent, there exists an unique point $\pi(x) \in \mathrm{St}\,(p, H)$ that realizes the minimum of the distance from x;*
2. *if $x_1, \ldots, x_p$ are linearly dependent, there still exists a point that realizes the minimum of the distance from x, but it is not unique.*

This result holds also when H is not separable.

Proof. We should minimize the function $\mathrm{St}\,(p, H) \to \mathbb{R}$

$$v \mapsto \|x - v\|_{H^p}^2 = \sum_{i=1}^{p} \|x_i - v_i\|_H^2$$

Since x is fixed and $|v_i| = 1$ for all $i = 1, \ldots, p$, this is the same as maximizing the linear function $g \colon \mathrm{St}\,(p, H) \to \mathbb{R}$,

$$g(v) = \sum_{i=1}^{p} \langle x_i, v_i \rangle_H \ .$$

To see if the minimum of the distance exists, it is sufficient to see if the maximum of g exists.

First of all we show that g has maximum. If H is finite dimensional, this is clear, because the Stiefel manifold is compact and g is continuous.

In the case where H is infinite dimensional, let $X = \mathrm{Span}(x_1, \ldots, x_p) \subset H$ and q be the dimension of X. Without loss of generality we can suppose that $x_1 \ldots x_q$ are a basis of X. We now consider the $p + q$ dimensional subspaces of H containing X and call them "nice" subspaces. Let Y be a "nice" subspace and $y_1 \ldots y_p$ an orthonormal basis of the orthogonal to X in Y. The vectors $x_1 \ldots x_q, y_1 \ldots y_p$ are a basis of Y.

Consider the function g restricted to $Y^p \cap \mathrm{St}\,(p, H)$. Using the above basis, this intersection can be written as

$$Y^p \cap \mathrm{St}\,(p, H) = \left\{ v \in H^p \;\middle|\; \exists a \in S : \forall j,\ v_j = \sum_{i=1}^{q} a_{j,i} x_i + \sum_{i=1}^{p} a_{j,i+q} y_i \right\}$$

where

$$S = \left\{ a \in \mathbb{R}^{p \times (p+q)} \;\middle|\; \substack{\forall j \ \sum_{i,l=1}^{q} a_{j,i} a_{j,l} \langle x_i, x_l \rangle + \sum_{i=1}^{p} a_{j,i+q}^2 = 1 \\ \forall j,k, j \neq k \ \sum_{i,l=1}^{q} a_{j,i} a_{k,l} \langle x_i, x_l \rangle + \sum_{i=1}^{p} a_{j,i+q} a_{k,i+q} = 0} \right\} .$$

Note that S does not depend on Y, but it is the same for all "nice" subspaces.

In the above basis, the supremum of g in $Y^p \cap \mathrm{St}\,(p, H)$ is

$$\sup_{v \in Y^p \cap \mathrm{St}(p,H)} g(v) = \sup_{a \in S} \sum_{j=1}^{p} \sum_{i=1}^{q} a_{j,i} \langle x_j, x_i \rangle \ .$$

The right hand side does not depend on Y. This means that the supremum is the same in all finite dimensional subspaces of the form Y^p for some "nice" subspace Y.

Moreover for each v in $\mathrm{St}\,(p, H)$ there exists a "nice" subspace $Y \subseteq H$ such that $v \in Y^p$ and so the global supremum in $\mathrm{St}\,(p, H)$ is equal to the supremum attained in any subspace of the form Y^p for some "nice" subspace Y. But subspaces of that form are finite dimensional, and there the supremum is clearly achieved.

We can now talk about uniqueness. To show that in case *1* there is uniqueness, we explicitly compute the minimum, choosing a suitable basis of H^p. The explicit computation shows also that a point at minimal distance exists, so the above proof is not really necessary in case *1*.

First of all, note that if the components of x are orthogonal, it is easy to find the minimum. Consider the point

$$v_{\min} = \left(\frac{x_1}{|x_1|}, \ldots, \frac{x_p}{|x_p|} \right) . \tag{14}$$

It minimizes the distance from x between all vectors whose components have unit norms. Thanks to the fact that the components of x are orthogonal, $v_{\min}$ belongs to $\mathrm{St}\,(p, H)$ and then it is the minimum also on the Stiefel manifold.

We define this notation. For every $y \in H^p$ and $A \in \mathbb{R}^{n \times n}$ the product $Ay \in H^p$ can be defined taking linear combinations of the components $y_1 \ldots y_p$, i.e.

$$(Ay)_i = \sum_{j=1}^{p} A_{ij} y_j \ . \tag{15}$$

Let xx^T be the $p \times p$ symmetric positive definite matrix whose entries are the scalar product between the components of x,

$$\left(xx^T\right)_{ij} = \langle x_i, x_j \rangle \quad . \tag{16}$$

We claim that the minimum point z is obtained as

$$z = B^{-1} x \quad , \quad B = \sqrt{xx^T} \quad , \tag{17}$$

where B is the unique symmetric positive definite matrix such that $B^2 = xx^T$.

We proceed to prove this claim.

Let A be an orthonormal matrix that diagonalizes xx^T,

$$A\, xx^T A^T = D = \mathrm{diag}(d_1, \ldots, d_n) \, . \tag{18}$$

Consider the mapping $f_A \colon H^p \to H^p$, $f_A(y) = Ay$. The key fact about f_A are:

- it is an isometry of H^p;
- it maps the Stiefel manifold into itself;
- the components of Ax are orthogonal.

It is an isometry because $A^T A = Id_{p\times p}$. Indeed, using matrix notation,

$$\|Ay\|^2_{H^p} = (Ay)^T Ay = y^T (A^T A) y = y^T y = \|y\|^2_{H^p} \ .$$

The fact $v \in \mathrm{St}\,(p, H)$ can be written as $vv^T = Id_{p\times p}$ and then

$$vv^T = Id \iff A(vv^T)A^T = Id \iff Av(Av)^T = Id \,.$$

The components of $y = Ax$ are orthogonal because A diagonalizes xx^T

$$Ax(Ax)^T = A\,xx^T A^T = \mathrm{diag}(d_1, \ldots, d_n)\,. \tag{19}$$

Note that $d_j > 0$ and that $\|y_j\| = \sqrt{d_j}$.

By the observation above, there is a unique point $\tilde{y}$ in $\mathrm{St}\,(p, H)$ that minimizes the distance from Ax, and is given by

$$(\tilde{y}_1, \ldots, \tilde{y}_p) \ \text{ with } \ \tilde{y}_j = y_j / \|y_j\| = y_j / \sqrt{d_j} \ , \tag{20}$$

or, in short, $\tilde{y} = D^{-1/2} y$. Because of the proprieties of f_A, the point $z = A^T \tilde{y}$ is a point on $\mathrm{St}\,(p, H)$ at minimal distance from x.

Combining the Equations (16), (18), (20) and (19) we obtain $z = A^T D^{-1/2} Ay$, but $B = A^T \sqrt{D} A$ so we have derived Equation (17).

Regarding point *2*, let $v \in \mathrm{St}\,(p, V)$ be a minimum of the distance from x. Let $X = \mathrm{Span}(x_1 \ldots x_p)$, and decompose H as $X + X^\perp$ and every component of v as $v_i = v_i^x + v_i^\perp$. Consider $\widetilde{v} = (v_1^x - v_1^\perp, \ldots, v_p^x - v_p^\perp)$. The vector $\widetilde{v}$ still lays on the Stiefel, indeed

$$\begin{aligned}\langle \widetilde{v}_i, \widetilde{v}_j \rangle = \left\langle v_i^x - v_i^\perp, v_j^x - v_j^\perp \right\rangle = \left\langle v_i^x, v_j^x \right\rangle + \left\langle v_i^\perp, v_j^\perp \right\rangle = \\ = \left\langle v_i^x + v_i^\perp, v_j^x + v_j^\perp \right\rangle = \langle v_i, v_j \rangle \ .\end{aligned}$$

Since the x_i are linearly dependent, the v_i could not all lay in X, but there is some $v_i^\perp \neq 0$ and then $v \neq \widetilde{v}$. Moreover,

$$|v_i - x_i|^2 = |\widetilde{v}_i - x_i|^2 \quad \text{for all } i = 1 \ldots p$$

and then there are at least two minima. □

The above proof shows that the minimum can actually be easily computed.

4.3.2 Properties of the projection

If V is a Hilbert space and $p \in \mathbb{N}, p \geq 1$ a natural number, we define

$$\operatorname{Ind}(p, V) = \{(x_1, \dots, x_p) \in V^p \mid x_1, \dots, x_p \text{ are linearly independent}\} .$$

Proposition 4.9. *This set is open, and the projection $\pi : \operatorname{Ind}(p, V) \to \operatorname{St}(p, V)$ is smooth.*

Proof. Let xx^T be defined as in Equation (16). Then $x \in \operatorname{Ind}(p, V)$ if only if $\det(xx^T) \neq 0$. The previous theorem proved that inside this set $\pi(x) = (xx^T)^{-1/2} x$, and this is a smooth function. $\square$

Proposition 4.10. *Consider $x = (x_1, \dots, x_p) \in H^p$, suppose that $x_1, \dots, x_p$ are linearly independent. Let $z = (z_1, \dots, z_p) \in \operatorname{St}(p, H)$ be the unique point at minimum distance, as in Prop. 4.8. Let $V \subset H$ be a vector subspace, suppose that $x_j \in V$ for $j = 1, \dots p$: then $z_j \in V$ for $j = 1, \dots p$.*

The projection on the Stiefel Manifold shares a property with the projection on the sphere.

Proposition 4.11. *Consider $x = (x_1, \dots, x_p) \in H^p$, suppose that $x_1, \dots, x_p$ are linearly independent. Let $t > 0$ and $y = tx$. Then both x and y project to the same point $z \in \operatorname{St}(p, H)$.*

Both proofs follows from the relation (17).

4.3.3 Projection of a Gaussian measure

If V is a Hilbert space and $p \in \mathbb{N}, p \geq 1$ a natural number, we define for convenience the complement of $\operatorname{Ind}(p, V)$ as

$$\operatorname{Dep}(p, V) = \{(x_1, \dots, x_p) \in V^p \mid x_1, \dots, x_p \text{ are linearly dependent}\} .$$

By Proposition 4.9 this is a closed subset of V^p. To prove that the projection is defined almost everywhere, we need the following lemma.

Lemma 4.12. *Let $p \leq n$ be positive natural numbers and consider the linear space $(\mathbb{R}^n)^p$ with the Lebesgue measure $(\mathscr{L}^n)^p = \mathscr{L}^n \times \cdots \times \mathscr{L}^n$. Then the set $\operatorname{Dep}(p, \mathbb{R}^n) \subseteq (\mathbb{R}^n)^p$ is negligible.*

Proof. We prove this lemma by induction on p. The case $p = 1$ is trivial, because $\operatorname{Dep}(1, \mathbb{R}^n)$ contains only the origin.

Suppose the lemma true for $p - 1$ and decompose $\operatorname{Dep}(p, \mathbb{R}^n)$ as

$$\operatorname{Dep}(p-1, \mathbb{R}^n) \times \mathbb{R}^n \cup \left\{ (x_1, \dots, x_p) \in (\mathbb{R}^n)^p \;\middle|\; \substack{(x_1, \dots, x_{p-1}) \notin \operatorname{Dep}(p-1, \mathbb{R}^n) \\ x_p \in \operatorname{Span}(x_1, \dots, x_{p-1})} \right\} .$$

The first set is negligible thanks to inductive hypothesis. Moreover for each $x_1, \dots, x_{p-1} \in \mathbb{R}^n$, the set of $x_p \in \mathbb{R}^n$ linearly dependent from them is a subspace of dimension at most $p-1 < n$ and so it is $\mathscr{L}^n$-negligible. By Fubini's theorem, the second set is negligible too, and so also $\operatorname{Dep}(p, \mathbb{R}^n)$ is negligible. $\square$

Lemma 4.13. *Let H be a Hilbert space, $p \in \mathbb{N}, p \geq 1$ and $\mathrm{St}\,(p, H)$ a Stiefel manifold. Let also γ be a non degenerate Gaussian measure on H^p. Then the set $\mathrm{Dep}(p, H)$ is γ-negligible.*

Proof. If H is finite dimensional, we assume that $\dim(H) \geq p$ (otherwise $\mathrm{St}\,(p, H)$ is empty); then the result follows from Lemma 4.12. Let us consider then the case when H is infinite dimensional.

Fix an orthonormal frame $\{e_i\}_{i\leq p}$ in H, consider the projection f

$$\begin{array}{llll} f\colon & H & \to & \mathbb{R}^p \\ & x & \mapsto & (\langle x, e_1\rangle, \dots, \langle x, e_p\rangle) \end{array}$$

and define a continuous linear projection f^p from H^p to $(\mathbb{R}^p)^p$ in this way

$$f^p\colon (x_1, \dots, x_p) \mapsto (f(x_1), \dots, f(x_p))\,.$$

Since f is linear, the image of $\mathrm{Dep}(p, H)$ is contained in $\mathrm{Dep}(p, \mathbb{R}^p)$, so it sufficient to prove that the inverse image of this set is negligible, or equivalently that $\mathrm{Dep}(p, \mathbb{R}^p)$ is $f_\sharp\gamma$-negligible.

By Lemma 4.12, $\mathrm{Dep}(p, \mathbb{R}^p)$ is negligible for the Lebesgue measure $(\mathscr{L}^p)^p$. Since γ is non degenerate, $f_\sharp\gamma$ is a non degenerate Gaussian measure on $(\mathbb{R}^p)^p$ and then it is absolutely continuous with respect to $(\mathscr{L}^p)^p$. It follows that

$$f_\sharp\gamma\left(\mathrm{Dep}(p, \mathbb{R}^p)\right) = 0$$

and so $\mathrm{Dep}(p, H)$ is negligible. ☐

Corollary 4.14. *Assume H, p and γ as in above lemma. Then for almost every $x \in H^p$ there exists a unique point $\pi(x) \in \mathrm{St}\,(p, H)$ that realizes the minimum of the distance from x, i.e.*

$$d(\pi(x), x) = d_{\mathrm{St}(p,H)}(x)\,.$$

Proof. By Proposition 4.8 for all points $x \notin \mathrm{Dep}(p, H)$ there exists a unique point at minimal distance on $\mathrm{St}\,(p, H)$. By the above lemma $\mathrm{Dep}(p, H)$ is negligible for the measure γ. ☐

Note that the above results hold also when H is not separable.

We so summarize.

Theorem 4.15. *Let H be a Hilbert space, $p \in \mathbb{N}, p \geq 1$ and $\mathrm{St}\,(p, H)$ a Stiefel manifold. Let also γ be a non degenerate Gaussian measure on H^p. Then the projection $\pi_\#\gamma$ is a well defined Radon probability measure on $\mathrm{St}\,(p, H)$.*

Proof. To define $\pi_\#\gamma$ we restrict γ to $\mathrm{Ind}(p, H)$, and we push forward this restriction using the map π. Theorem 7.1.7 in [3] ensures that $\pi_\#\gamma$ is Radon. ☐

4.3.4 Properties of projected Gaussian measures

Let H be a Hilbert space, $p \in \mathbb{N}, p \geq 1$ and $\mathrm{St}\,(p, H)$ a Stiefel manifold. Let also γ be a non degenerate Gaussian measure on H^p. Then by the above Theorem, the projection $\pi_\# \gamma$ is a well defined Radon probability on $\mathrm{St}\,(p, H)$.

We show two simple properties of these kind of probabilities.

Lemma 4.16. *Suppose that θ and γ are equivalent probability measures on a space Ω and that $\pi : \Omega \to \Omega'$ is a measurable map. Then $\pi_\# \theta$ and $\pi_\# \gamma$ are equivalent.*

The proof is simple and is omitted. This result follows.

Proposition 4.17. *Suppose that θ and γ are equivalent non degenerate Gaussian measures on H^p. Then $\pi_\# \theta$ and $\pi_\# \gamma$ are equivalent.*

A vice versa does not hold.

Remark 4.18. Let H and γ be as above. Let $t > 0$ and θ be a rescaling of γ, that is, $\theta(B) = \gamma(tB)$ for any $B \in H^p$ Borel set. Then the projections $\pi_\# \gamma$ and $\pi_\# \gamma$ are identical. This derives from Corollary 4.11. Note that θ and γ are mutually singular, unless $t = 1$. So two mutually singular Gaussian measures can project to identical measures.

We already remarked that, if $O : \mathbb{R}^p \to \mathbb{R}^p$ is a rotation, then the map $v \mapsto Ov$ defined in Equation (15) is an isometry and sends $\mathrm{St}\,(p, H)$ into itself. It is quite easy to provide a Gaussian probability that is invariant *w.r.t.* this action.

Proposition 4.19. *Let μ be a non degenerate centered Gaussian measure on H, and let $\gamma = \mu^p = \mu \otimes \cdots \otimes \mu$ be the product probability on H^p. Then γ is invariant with respect to the action of μ and hence the projected measure $\pi_\# \gamma$ is a probability on $\mathrm{St}\,(p, H)$ that is invariant* w.r.t. *the action $v \mapsto Ov$.*

Acknowledgments

We thanks Prof. Luigi Ambrosio for many corrections and suggestions (and in particular for Proposition 4.3).

A preliminary version of this work was presented at a conference in the programme *Infinite-dimensional Riemannian geometry with applications to image matching and shape analysis*, under support by the Erwin Schrödinger Institute in Vienna.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] C. J. Atkin. The Hopf-Rinow theorem is false in infinite dimensions. *Bull. London Math. Soc.*, 7(3):261–266, 1975.

[3] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[4] Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[5] Leo Breiman. *Probability*. Addison-Wesley, 1968.

[6] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[7] Giuseppe Da Prato. *An introduction to infinite-dimensional analysis*. Springer, 2006.

[8] Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[9] Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1999.

[10] Herbert Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[11] Philipp Harms and Andrea Mennucci. Geodesics in infinite dimensional Stiefel and Grassmann manifolds. *Comptes rendus - Mathématique*, 350:773–776, 2012. http://cvgmt.sns.it/papers/harmen10/.

[12] David G Kendall. The diffusion of shape. *Advances in applied probability*, pages 428–430, 1977.

[13] David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[14] Andrei Nikolaevich Kolmogorov. La transformation de laplace dans les espaces linéaires. *CR Acad. Sci. Paris*, 200:1717–1718, 1935.

[15] Serge Lang. *Fundamentals of differential geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999.

[16] Huiling Le and David G Kendall. The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. *The Annals of Statistics*, pages 1225–1271, 1993.

[17] Carlo Mantegazza and Andrea Carlo Mennucci. Hamilton–Jacobi equations and distance functions on Riemannian manifolds. *Applied Math. and Optim.*, 47(1):1–25, 2003.

[18] Andrea C. G. Mennucci. Regularity and variationality of solutions to Hamilton-Jacobi equations. I. Regularity. *ESAIM Control Optim. Calc. Var.*, 10(3):426–451 (electronic), 2004.

[19] Ganesh Sundaramoorthi, Andrea Mennucci, Stefano Soatto, and Anthony Yezzi. A new geometric metric in the space of curves, and applications to tracking deforming objects by prediction and filtering. *SIAM J. Imaging Sci.*, 4(1):109–145, 2011.

[20] Laurent Younes. Computable elastic distances between shapes. *SIAM J. Appl. Math.*, 58(2):565–586 (electronic), 1998.

[21] Laurent Younes, Peter W. Michor, Jayant Shah, and David Mumford. A metric on shape space with explicit geodesics. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 19(1):25–57, 2008.