PROBABILITY MEASURES ON INFINITE-DIMENSIONAL STIEFEL MANIFOLDS

ELEONORA BARDELLI†
Microsoft Deutschland GmbH, Germany
ANDREA CARLO GIUSEPPE MENNucci∗
Scuola Normale Superiore
Piazza dei Cavalieri 7, 56126, Pisa, Italia

Abstract. An interest in infinite-dimensional manifolds has recently appeared in Shape Theory. An 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.

Suppose that $H$ is an infinite-dimensional separable Hilbert space. Let $S \subset H$ be the sphere, $p \in S$. Let $\mu$ be the push forward of a Gaussian measure $\gamma$ from $T_p S$ onto $S$ using the exponential map. Let $v \in T_p S$ be a Cameron–Martin vector for $\gamma$; let $R$ be a rotation of $S$ in the direction $v$, and $\nu = R_\# \mu$ be the rotated measure. Then $\mu, \nu$ are mutually singular. This is counterintuitive, since the translation of a Gaussian measure in a Cameron–Martin direction produces equivalent measures.

Let $\gamma$ be 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 $\gamma$ measure.

Instead it is possible to project a Gaussian measure to a Stiefel manifold to define a probability.

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 [15] provided the first general definition of a Gaussian measure on an infinite-dimensional space. 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 [13, 14, 17]. 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 $n \in \mathbb{N}, n \geq 1$ and $H$ be a Hilbert space. The Stiefel manifold $St(n,H)$ is the subset of $H^n$ consisting of orthonormal $n$-tuples of vectors. In symbols,

$$St(n,H) = \{(v_1, \ldots, v_n) \in H^n \mid \forall i, j \text{ with } i \neq j, \langle v_i, v_j \rangle_H = 0 \text{ and } \forall i, \|v_i\|_H = 1 \}.$$ 

The Stiefel manifold is a smooth embedded submanifold of $H^n$, 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 $\dim(H) \geq n$ (otherwise $St(n,H)$ is empty).

In [21] Younes studied the space $\mathcal{M}$ of smooth immersed closed planar curves. Those ideas were then revisited in Younes et al [22], 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 $St(2,L^2)$, where $L^2 = L^2([0,1])$ is the usual Hilbert space 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 [20] 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 $St(n,L^2)$ (adapting a method described in [9]); they proposed a novel method for tracking shapes bounded by curves, that is based on a simple first-order dynamical system on $St(2,L^2)$.

Moreover Harms and Mennucci [12] proved that any two points in the Stiefel manifold (respectively in the Grassmann manifold) are connected by a minimal length geodesic.

Since the manifold $St(2,L^2)$ 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 $\varphi : A \rightarrow V$ be a local chart; then the push-forward of $\mathcal{H}^n$ using $\varphi^{-1}$ 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 a finite-dimensional differentiable manifold $M$ is by using volume densities. All volume densities are equivalent. If $M$ is

\(^2\) “Equivalent” means “mutually absolutely continuous”. See Definition 1.2.
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 \( H^n \). 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.

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 \( T_p M \) is isomorphic to a subspace of \( H \), so we can define a probability measure \( \mu \) on \( T_p M \), e.g. a Gaussian measure. At the same time, the exponential map \( \exp_p : T_p M \to M \) is smooth, so we can push forward the probability \( \mu \) to define the probability \( \nu = \exp_p \# \mu \) on \( M \). In some texts this procedure is known as “wrapping”. This method undoubtedly 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 \( \mathcal{L}^n \). 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 \( T_p M \). Again, if we push forward \( \mathcal{L}^n \) using this local diffeomorphism, we obtain a measure that is equivalent to \( H^n \) near \( p \). (The \( n \)-dimensional Hausdorff measure \( H^n \) 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 \( T_p M \) the tangent space to \( M \) in \( p \). Let \( \gamma \) be a measure on \( T_p M \), equivalent to the Lebesgue measure \( \mathcal{L}^n \). 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 \( T_p M = M \) for all \( p \in M \); we view \( M \) as a (trivial) Riemannian manifold by associating the norm \( \| \cdot \|_M \) to each tangent space \( T_p M \). In this case the wrapping is trivial. Let \( p_1 = 0, p_2 \neq 0 \). Fix a non-degenerate Gaussian measure \( \mu \) on \( M \). Since \( \exp_{p_1} \) is the identity map, then the wrapping
\[ \nu_1 = \exp_{p_1} \# \mu \text{ of } \mu \text{ itself. At the same time the wrapping } \nu_2 = \exp_{p_2} \# \mu \text{ of } \mu \text{ is the translation of } \mu, \text{ translated by the vector } p_2. \text{ It is well known that } \nu_1 \text{ and } \nu_2 \text{ are equivalent if and only if } p_2 \text{ lies in the Cameron-Martin space of } \mu_1, \text{ otherwise they are mutually singular. See next section 2 for detailed definitions and further results.} \]

The matter becomes even more intricate in the case of an infinite-dimensional manifold. Let \( S \) be the unit sphere in an 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 non-trivial 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 \( \ell^2 \) 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 4.3.4 that the projection method works fine in the case of the Stiefel manifold \( \text{St}(n,H) \). Indeed, for any non-degenerate Gaussian measure \( \eta \) on \( H^n \), the projection from \( H^n \) to the nearest point in \( \text{St}(n,H) \) is defined for \( \eta \)-almost all points. So we can project \( \eta \) onto \( \text{St}(n,H) \) 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 \( v^*(x) = \langle v, x \rangle_H \).

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” \( \mu \) 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” $\mu$ on $M$ (or more simply “probability”) we will mean a
measure $\mu$ such that $\mu(M) = 1$.

When $\mu$ is a probability the push-forward $\psi_\# \mu$ is a probability on $N$, and is
usually called the “distribution” or the “law” of $\psi$ on $N$.

**Definition 1.2.** Let $\mu$ and $\nu$ be measures on $M$.

- The measure $\nu$ is called “absolutely continuous” with respect to $\mu$ if $\nu(A) = 0$
  for every set $A \in \mathcal{B}(M)$ with $\mu(A) = 0$. We will write $\nu \ll \mu$ in this case.
- The measures $\mu, \nu$ are “equivalent” if they are mutually absolutely continuous.
  We will write $\mu \sim \nu$ in this case.
- The measures $\nu, \mu$ 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 \cite{4} and \cite{7}.

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

**Definition 2.1.** A probability measure $\gamma$ on $\mathbb{R}$ is said to be Gaussian if it is either
a Dirac measure, or has density
\[
  x \mapsto \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x - m)^2}{2\sigma^2} \right)
\]
with respect to the Lebesgue measure for some parameters $\sigma, m \in \mathbb{R}$. In the first
case the measure is called degenerate.

**Definition 2.2.** Let $H$ be a Hilbert space. A measure $\gamma$ on $H$ is said to be Gaussian
if for all $x \in H$ the push-forward measure $x^*\gamma$ 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 $\mu$ on $H$
is
\[
  \widehat{\mu} : H^* \to \mathbb{C} , \quad \widehat{\mu}(f) := \int_H \exp(\overline{f(y)}) \, d\mu(y) .
\]

**Theorem 2.3** (Theorem 2.3.1 in \cite{4}, or Theorem 1.12 in \cite{7}). Let $\gamma$ be a Gaussian
measure on a Hilbert space $H$. Then there exist a vector $m \in H$ and a symmetric
non-negative nuclear (i.e., trace class) operator $K$ such that
\[
  \widehat{\gamma}(f) = \exp \left( i \langle m, x \rangle_H - \langle Kx, x \rangle_H \right) .
\]
Vice versa, given any $m$ and $K$ as above, there exists a unique Gaussian measure
$\gamma$ satisfying (2).

We recognize that $m$ is the mean and $K$ is the covariance operator of $\gamma$, 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
\[
  \mathbb{E}[v^*] := \int_H v^*(x) \, d\gamma(x) = \langle m, v \rangle_H ,
\]
\[
  \text{Cov}[v^*, w^*] := \int_H v^*(x - m) \, w^*(x - m) \, d\gamma(x) = \langle Kv, w \rangle_H .
\]
For this reason we will indicate \( \gamma \) with the usual notation \( N(m, K) \). When \( m \) is zero, we say that \( \gamma \) is centered. When the kernel of \( K \) is \( \{0\} \), we say that \( \gamma \) 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

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

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 \( m_n = \langle m, e_n \rangle_H \in \mathbb{R} \) and \( \sigma_i \geq 0 \) is the eigenvalue such that \( Ke_n = \sigma_ne_n \), then \( e_n^* \sim N(m_n, \sigma_n) \), that is, \( m_n \) is the mean and \( \sigma_n \) the variance of the real Gaussian random variable \( e_n^* \). Moreover \( \sum_{n=0}^{\infty} \sigma_n = \text{Tr}(K) < \infty \). Obviously \( \gamma \) is non-degenerate if \( \sigma_n > 0 \) 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 \( \gamma \) be a centered Gaussian measure on a Hilbert space \( H \), let \( K \) be its covariance. The Cameron–Martin space \( \text{CM}(\gamma) \) of \( \gamma \) is the range (i.e., the image) of \( K^{1/2} \). In symbols,

\[
\text{CM}(\gamma) = K^{1/2}(H) .
\]

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

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

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

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

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

\[
W : \text{CM}(\gamma) \to L^2(H, \gamma) , \ z \mapsto W_z
\]

where \( W_z \) is defined by \( W_z(x) = \langle x, K^{-1/2}z \rangle_H \). This mapping is an isometry from \( \text{CM}(\gamma) \) (with the norm of \( H \)) to \( L^2(H, \gamma) \), so it extends to a 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 \( \gamma \) 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 \( \gamma \).
• If \( h \in \text{CM}(\gamma) \) then \( \mu \) and \( \gamma \) are equivalent, and the Radon–Nicodym derivative is given by the Cameron–Martin formula

\[
\frac{d\mu}{d\gamma}(x) = \exp \left( -\frac{1}{2} \|a\|^2_H + W_a(x) \right)
\]

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

• Moreover the total variation distance is bounded by

\[
\|\gamma - \mu\|_{TV} \leq 2\sqrt{1 - \exp \left( -\frac{1}{4} \|a\|^2_H \right)}.
\]

• If \( h \notin \text{CM}(\gamma) \) then \( \mu \) and \( \gamma \) 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 \( \gamma \) on the tangent space \( T_pM \) in \( p \) and then push forward \( \gamma \) 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 [16]. The exponential map \( \exp_p : T_pM \to M \) is defined as

\[
\exp_p(v) = \sigma_v(1)
\]

where \( \sigma_v \) is the geodesic starting from \( \sigma_v(0) = p \) with tangent vector \( \dot{\sigma}_v(0) = v \).

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 \( \exp_p \# \gamma \) 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 \( \gamma \) be a measure on \( T_pM \), absolutely continuous wrt the Lebesgue measure \( \mathcal{L}^n \). Then its push-forward \( \mu = \exp_p \# \gamma \) under the exponential map is absolutely continuous wrt the Hausdorff measure. In symbols

\[
\gamma \ll \mathcal{L}^n \Rightarrow \exp_p \# \gamma \ll \mathcal{H}^n.
\]

If moreover \( \gamma \) is equivalent to the Lebesgue measure, then \( \mu \) is equivalent to the Hausdorff measure. In symbols

\[
\gamma \sim \mathcal{L}^n \Rightarrow \exp_p \# \gamma \sim \mathcal{H}^n.
\]

**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 $f_\# \mathcal{L}^n$ is absolutely continuous wrt $\mathcal{H}^n$ if and only if $\mathcal{L}^n(C_f) = 0$. Let now $f = \exp_p$ so that $\mu = f_\# \mathcal{L}^n$. Let us divide $C_f = \bigcup_{i=1}^{n-1} \Gamma_i$ with

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

The Theorem 4.1 of [19] proves that each of the above sets $\Gamma_i$ is locally contained in a $(n-i)$-dimensional submanifold of $T_p M$. So $\mathcal{L}^n(C_f) = 0$.

Suppose now moreover that $\gamma$ is equivalent to the Lebesgue measure, we want to prove that $\mu$ 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 $\mu$, i.e., $\mathcal{H}^n \ll \mu$. We will use some facts that are explained in [18]. 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 [18] that $\mathcal{H}^n(K_p) = 0$, so we will ignore $K_p$ in the following. Let $E \subseteq \Omega_p$ be a Borel set such that $\mu(E) = 0$. Let $B = f^{-1}(E)$, then by definition of push-forward $\mathcal{L}^n(B) = 0$. There exists an open star-shaped set $O \subseteq T_p M$ 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 $\mathcal{L}^n(D) = 0$. The second point in the above Lemma shows then that $\mathcal{H}^n(E) = 0$.

**Corollary 3.2.** Let us consider, for $i = 1, 2$, a point $p_i \in M$, a probability measure $\mu_i$ defined on $T_{p_i} M$ that is equivalent to the Lebesgue measure on $T_{p_i} M$. Let $\nu_i = \exp_p, \# \mu_i$ be the wrapping of $\mu_i$ on $M$. Then the two measures $\nu_1, \nu_2$ 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 [20] when devising a discrete stochastic process on the Stiefel manifold $\text{St}(2, L^2)$, 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 rotations, 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 also when Gaussian measures are not concentrated.
on finite-dimensional spaces. Indeed if a Gaussian measure is supported on an infinite-dimensional closed subspace, then we may restrict the following analysis to that subspace, and the restriction of the measure would be a non-degenerate Gaussian measure 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 $\gamma$ be a non-degenerate Gaussian measure on it. Let $v_n = \sigma_n, \text{ that is, } f_n = v_n/\sigma_n$, so that the random variable $f_n$ has standard Gaussian distribution $N(0, 1)$. Since the joint distribution of $(f_1, \ldots, f_n)$ 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]), $\gamma$ is concentrated on the Borel set

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

(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 $\gamma$, and for any $v \in C$, then $v + r \in C$. In symbols,

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

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 $\bar{x} = x + r$, so for all $i$

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

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

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

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

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{r_i^2x_i}{\sigma_i^2} = 0$$
for $\gamma$-almost any $x$. Similarly, again by the law of large numbers, for $\gamma$-almost every $x$,

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

Summing up we obtain the desired result.

**Lemma 3.4.** Let $\gamma$ 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

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

where $H' = \text{Span}(e_2, e_3, \ldots)$ and let $\pi$ be the orthogonal projection on $H'$. By the independence of the coordinate functions, $\gamma$ can be decomposed as

$$\gamma = e_1^\star \gamma \otimes \pi^\star \gamma.$$ 

We compute the measure of the sphere using Fubini’s theorem for the product measure $e_1^\star \gamma \otimes \pi^\star \gamma$. For every $x \in H'$, there are at most two $x_1$ such that $(x_1, x') \in S_r$. Since $e_1^\star \gamma$ is a Gaussian measure on $\mathbb{R}$, finite sets are negligible with respect to it. It follows that $S_r$ is negligible for $\gamma$, since every slice at $x' \in H'$ fixed is negligible with respect to $e_1^\star \gamma$. 

It is worth noting 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 $T_p S$. 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

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

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

**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 functionals on $H$.

Reasoning as in Lemma 3.3, we obtain that $\gamma$ is concentrated on the set

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

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 $\mu_1$, $\mu_2$ the push-forward of $\gamma$ under $\exp_p$ and $\exp_{-p}$

$$\mu_1 = \exp_p \gamma \quad \mu_2 = \exp_{-p} \gamma.$$
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 lie on the equator.

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, $\mu_1$ is concentrated on $C_1$ and $\mu_2$ is concentrated on $C_2$.

To prove that $\mu_1$ and $\mu_2$ 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

$$
\exp_p(x) = \cos(||x||_H)p + \sin(||x||_H)\frac{x}{||x||_H}
$$

$$
\exp_{-p}(x) = -\cos(||x||_H)p + \sin(||x||_H)\frac{x}{||x||_H}
$$

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 $\mu_1$ (and also for $\mu_2$). Let $S_r$ be the sphere of radius $r$ in $T$, then

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

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 $q \in S$. Let $\gamma$ be a centered non-degenerate Gaussian measure on $T_qS$. Let $\mu = \exp_{q\#} \gamma$ the wrapping of $\gamma$ on $S$. Let $r \in T_qS, r \neq 0$ be a vector that is in the Cameron–Martin space of $\gamma$. Let $\hat{q}r$ be the plane spanned by $q, r$. We define a rotation $R : H \to H$ in this way: $R$ rotates any vector in the plane $\hat{q}r$ by a fixed angle, whereas $R$ keeps fixed any vector orthogonal to $\hat{q}r$. Suppose that $R$ is not the identity map. Let $\nu = R_\# \mu$. Then $\mu, \nu$ are mutually singular.
The rotation $R$ rotates $q$ in the direction $r$. So we may think of $R$ as a “Cameron–Martin rotation”. This would mislead us into thinking that $\nu$ and $\mu$ are equivalent. Instead they are mutually singular.

We remark this fact.

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

We will need the following Lemma.

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

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

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 $\gamma$, then the conditional measures $\nu_z$ 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.

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

$$p = e_1 \cos \theta + r \sin \theta,$$
$$q = e_1 \cos \theta - r \sin \theta.$$

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

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

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 $V^\bot$, whereas it rotates vectors in the plane $V$ by the angle $\theta$ (so that $Re_1 = p$ and $Rq = e_1$).

Note that the rotation defined in the statement of this theorem (and used in the following Remark 3.8) is the square of the rotation here defined in the proof: indeed $R^2 q = p$. 

The hypotheses of the theorem as follows. Let $\tilde{R} = D_q R : T_q S \to T_p S$, then $\tilde{R}$ coincides with the parallel transport along $\xi$. Let $\tilde{\gamma} = \tilde{R}_\# \gamma$, let then $\tilde{\mu} = \exp_{p\#} \tilde{\gamma}$ the wrapping of $\tilde{\gamma}$ on $S$. Then $\tilde{\mu} = \nu$. So the probability $\nu$ is also obtained by identifying $T_q S \to T_p S$ using parallel transport, and then wrapping.
We have excluded 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 $R^2 x = -x$, 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 $\tilde{R}_p : T_{e_1}S \rightarrow T_p S$ and $\tilde{R}_q : T_q S \rightarrow T_{e_1}S$ be the tangent maps.

We assume that $\gamma$ 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 $\sigma_k^2$ be the eigenvalue of $K$ in direction $e_k$.

We assume that $r$ is in the Cameron–Martin space of $\gamma$. We push forward $\gamma$ to $\gamma_p$ using $\tilde{R}_p$, and pull it back to $\gamma_q$ 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 $\gamma_p$ and $\tilde{q}$ is in the Cameron–Martin space of $\gamma_q$.

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 $\gamma_q$ and $\nu = (\exp_p)_\# \gamma_p$ is the wrapping of $\gamma_p$, 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 a unique way) as

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

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 $\gamma_p$ is the same as the joint distribution of $(b, \tilde{w})$ according to $\gamma_q$. 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

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

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

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

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

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

$$\exp_p(v) = \left(p \cos(t) + a\tilde{p} \text{sinc}(t)\right) + \tilde{v} \text{sinc}(t),$$

$$\exp_q(w) = \left(q \cos(s) + b\tilde{q} \text{sinc}(s)\right) + \tilde{w} \text{sinc}(s).$$

If these maps reach the same point, then

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

$$\tilde{v} \text{sinc}(t) = \tilde{w} \text{sinc}(s). \quad (7)$$
We know that \( \hat{v} = v - a\hat{p} \) and that \( \hat{p} \) is in the Cameron–Martin space of \( \gamma_p \). By applying Lemma 3.3 we assert that

\[
\lim_{j \to \infty} \frac{1}{j} \sum_{i=1}^{j} \frac{|\hat{v}_i|^2}{\sigma_i^2} = 1.
\]

for \( \gamma_p \)-almost any \( v \). The same holds for \( w \) as well, \textit{mutatis mutandis}. By Lemma 3.4 we can assume that \( \text{sinc}(s) \neq 0 \) and \( \text{sinc}(t) \neq 0 \). So for \( \gamma_p \)-almost all \( v \) and \( \gamma_q \)-almost all \( w \),

\[
\lim_{j \to \infty} \frac{1}{j} \sum_{i=1}^{j} \frac{(\hat{v}_i)^2}{\sigma_i^2} = 1 = \lim_{j \to \infty} \frac{1}{j} \sum_{i=1}^{j} \frac{(|\hat{w}_i|)^2}{\sigma_i^2} = \frac{\text{sinc}^2 t}{\text{sinc}^2 s} \lim_{j \to \infty} \frac{1}{j} \sum_{i=1}^{j} \frac{(\tilde{v}_i)^2}{\sigma_i^2},
\]

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

We will now use this fact to the best. Foremost, we elaborate on the equation (6).

We know that the frame \( p, \hat{p} \) is obtained by \( q, \tilde{q} \) by rotating by an angle \( 2\theta \). So

\[
\begin{pmatrix}
\cos(2\theta) & \sin(2\theta) \\
-\sin(2\theta) & \cos(2\theta)
\end{pmatrix}
\begin{pmatrix}
\cos t \\
\sin t
\end{pmatrix}
= \begin{pmatrix}
\cos s \\
\sin s
\end{pmatrix}.
\]

(9)

Let

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

The points in \( E_0 \) are all the positive zeros of \( \text{sinc} \), while the points in \( E_1 \) are all the positive zeros of its derivative \( \text{sinc}' \). 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 \( \text{sinc}(s) \) 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: \( \varphi \) is a homeomorphism between maximal sub-intervals of \( I_n, I_k \); each one of these sub-intervals has a zero of \( \text{sinc} \) as one of its endpoints; \( \varphi \) and its inverse are analytic; when \( t \in I_n \) and \( s \in I_k \) the relation \( \text{sinc}(t) = \pm \text{sinc}(s) \) holds if and only if \( s = \varphi(t) \).

Recall that \( v \) is distributed according to \( \gamma_p \). By Lemma 3.9, for almost any \( \hat{v} \), the conditional distribution of \( u \) is Gaussian and non-degenerate. Let us fix such a \( \hat{v} \).

From (9) we extract the identity

\[
\cos(2\theta) \cos(t) + a \sin(2\theta) \text{sinc}(t) - \cos(\varphi(t)) = 0,
\]

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

\[
\pm \cos(2\theta) \pm 1
\]

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. We remarked, after Lemma 3.9, that in this case the conditional measures
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
\((X, \mathcal{F}_X, P)\), a measurable map \( f : X \to M \) and endow \( M \) with the push-forward
measure \( f_*P \).

Example 4.1. Let \( S^n \subseteq \mathbb{R}^{n+1} \) be the \( n \)-dimensional unit sphere and \( \gamma \) a Gaussian
measure on \( \mathbb{R}^{n+1} \) with mean 0 and covariance operator the identity. Consider the
projection
\[
\pi : \mathbb{R}^{n+1} \setminus \{0\} \to S^n, \quad x \mapsto \frac{x}{\|x\|_{\mathbb{R}^{n+1}}},
\]
which is defined \( \gamma \) almost everywhere. Then the measure \( \pi_*\gamma \) on \( S^n \) coincides with
the Hausdorff measure \( H^n \) restricted to the sphere and normalized.

4.1. Finite-dimensional manifolds. The above example can be properly gener-
alized, 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 con-
sider 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 points \( x \) 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 \( \gamma \) which is locally absolutely continuous with respect to the
Lebesgue measure, the measure \( \pi_*\gamma \) is well defined on \( M \).

Proof. Here is a sketch of the proof, the detailed arguments may be found in [18]
and references therein. The distance function \( d_M \) is Lipschitz. At all points where
\( d_M \) is differentiable, the projection point is unique. Let \( \Sigma \) be the set where \( d_M \)
is not differentiable. By Rademacher Theorem \( \Sigma \) is negligible.

In the case when \( M \) is a smooth submanifold, moreover, \( \Sigma \) and its closure both
have Hausdorff dimension at most \( m - 1 \); see [18]. 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 $\gamma$ be a Gaussian measure on $H$. Then for $\gamma$-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 $\Sigma$ where $d_M$ is not Gâteaux differentiable has measure $\gamma(S) = 0$. The rest of the proof works as in the finite-dimensional case. \hfill $\square$

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

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

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 $f(x) = \|T(x)_H\|^2$. 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 [16] 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 $x \in S$,

$$\|x\|_H > c.$$
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 $n \in \mathbb{N}$,
\[ d_S(0) \leq \inf_{n \in \mathbb{N}} \|ca_n^{-1}e_n\|_H = \inf_{n \in \mathbb{N}} ca_n^{-1} = c \]
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
\[ E_S = \left\{ x \in H \left| \sum_{i=1}^{+\infty} \left( \frac{1}{1-a_i^2} \right)^{2} x_i^2 < c^2 \right. \right\} \]
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 $\lambda$ is negative). If $\lambda$ 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.,
\[ \sum_{i=1}^{+\infty} a_i^2 x_i^2 < c^2 \quad \text{for all } x \in E_S \]
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 $f: H \to \mathbb{R}$ 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 $g: H \to \mathbb{R}$ be the square of the function we want to minimize on $S$, i.e., $g(y) = \|y - x\|^2_H$. 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

\[
x_i = (1 - \lambda a_i^2) z_i. \tag{11}
\]

This equation gives us some information about \( \lambda \). 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 \( z_i \):

\[
f(z) = \sum a_i^2 \left( \frac{1}{1 - \lambda a_i^2} \right)^2 x_i^2 = \sum \left( \frac{1}{1 - a_i^2} \right)^2 x_i^2 < c^2
\]
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 \( y = \hat{y} + \hat{y} \), where \( \hat{y} \in \text{Span}(e_1, \ldots, e_n) \) and \( \hat{y} \in \text{Span}(e_{n+1}, \ldots, e_{n+1})^\perp \). 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 \( \{a_{n+1}, a_{n+2}, \ldots\} \) 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

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

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 Proposition 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 \( \gamma \) a Gaussian measure on it. Then there exists a manifold \( S \) embedded and closed in \( H \) and a set \( E_S \) of positive \( \gamma \)-measure such that for every \( x \in E_S \) the distance from \( x \) has no minimum on \( S \).

**Proof.** We assume that \( \gamma \) is centered and non-degenerate. Choose an orthonormal basis \( \{e_i\}_{i\in\mathbb{N}} \) of \( H \) that diagonalizes the covariance \( K \) of \( \gamma \). The coordinate functions \( x \mapsto x_i = \langle x, e_i \rangle_H \) are independent Gaussian random variables; we denote by \( \sigma^2_i \) their variances; we know that \( \sum_{i=1}^{+\infty} \sigma^2_i < +\infty \) 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 \( f: H \to \mathbb{R} \cup \{+\infty\} \) 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

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

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 \( \gamma \) almost everywhere and, up to negligible sets,

\[
H = \bigcup_{n\in\mathbb{N}} \{ f(x) < n \} \cup \{ +\infty \}.
\]

we choose \( c \) large enough so that \( E_S \) is not negligible for \( \gamma \).

**Remark 4.7.** One may wonder if the provided example is “complete”. There are many meanings attached to the word “complete”. When a Riemannian manifold is finite dimensional, they are equivalent, due to the Hopf-Rinow theorem. In the infinite dimensional case, they are not. The example ellipsoid \( S \) presented in Lemma 4.4. is complete in some senses.

1. It is **metrically complete** (i.e. any Cauchy sequence converges) since it is a closed subset of a Hilbert space.
2. It is **geodesically complete**, that is, any geodesic segment can be infinitely prolonged; this is due to the fact that the geodesic spray is bounded in \( TS \).

Since it is quite similar to the Grossman example [11], then there are points in \( S \) that cannot be connected by a minimal length geodesic.
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 4.8.** Let $H$ be a separable Hilbert space and $S \subseteq H$ the unit sphere. Then the function

$$\pi: x \mapsto \frac{x}{\|x\|_H}$$

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

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

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

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

with $x_i \in H$.

**Proposition 4.9.** Let $H$ be a Hilbert space, $h \in \mathbb{N}, h \geq 1$, $\text{St}(h, H) \subseteq H^h$ the Stiefel manifold and $x \in H^h$. Then

1. if the components $x_1, \ldots, x_h$ are linearly independent, there exists a unique point $\pi(x) \in \text{St}(h, H)$ that realizes the minimum of the distance from $x$;
2. if $x_1, \ldots, x_h$ 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 finite dimensional and $\dim(H) \geq h$, or when $H$ is not separable.

**Proof.** The proof is divided into three steps. In the first step we prove that the minimum exists. In the second step we prove uniqueness in case 1. In the third step we prove that the minimum is not unique in case 2.

**Step 1 - The minimum exists.** Let $v = (v_1, \ldots, v_h)$ be a generic vector in $\text{St}(h, H)$. Since $x$ is fixed and $\|v_i\|_H = 1$, proving that the distance has a minimum is the same as proving that the function $g: \text{St}(h, H) \to \mathbb{R}$,

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

has a maximum.

If $H$ is finite-dimensional then the Stiefel manifold is compact and $g$ is continuous so it readily follows that $g$ has a maximum.

In the case where $H$ is infinite-dimensional, let $X = \text{Span}(x_1, \ldots, x_h) \subset H$ 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 $h + q$ dimensional subspaces of $H$ containing $X$ and call them “nice” subspaces. Let $Y$ be a “nice” subspace and $y_1, \ldots, y_h$ an
orthonormal basis of the orthogonal to $X$ in $Y$. The vectors $x_1, \ldots, x_q, y_1, \ldots, y_h$ are a basis of $Y$.

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

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

where

$$S = \left\{ a \in \mathbb{R}^{h \times (h+q)} \mid \forall j \sum_{i=1}^{q} a_{j,i} (x_i, x_i)_H + \sum_{i=1}^{h} a_{j,i+q}^2 = 1, \forall j, k, j \neq k \sum_{i=1}^{q} a_{j,i} a_{k,i} (x_i, x_i)_H + \sum_{i=1}^{h} a_{j,i+q} a_{k,i+q} = 0 \right\}.$$

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 $Y^h \cap \text{St}(h, H)$ is

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

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 $Y^h$ for some “nice” subspace $Y$.

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

**Step 2 - The minimum is unique in case 1.** To show that the minimum is unique when $x_1, \ldots, x_h$ are linearly independent we explicitly compute the minimum, choosing a suitable basis of $H^h$. The explicit computation shows also that a point at minimal distance exists, but since step 1 is anyway necessary to prove case 2, we will not stress this fact.

First of all, note that if the components of $x$ are orthogonal, then the minimum is unique and is given by the formula:

$$v_{\text{min}} = \left( \frac{x_1}{\|x_1\|_H}, \ldots, \frac{x_h}{\|x_h\|_H} \right).$$

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

Now we would like to find an isometry of $H^h$ that preserves the Stiefel manifold and maps $x$ to a vector whose components are orthogonal.

We recall the following notations. Let $A \in \mathbb{R}^{h \times h}$ and $y \in H^h$, the product $Ay \in H^h$ is defined as follows

$$(Ay)_i = \sum_{j=1}^{h} A_{i,j} y_j.$$

As usual, we denote by $A$ also the function $A : H^h \rightarrow H^h, y \mapsto Ay$.

We denote by $yy^T \in \mathbb{R}^{h \times h}$ the symmetric positive definite matrix whose entries are the scalar product between the components of $y$

$$(yy^T)_{i,j} = (y_i, y_j)_H.$$
Consider the matrix $xx^T$ and let $A$ be an orthonormal matrix that diagonalizes it,
\[ Axx^TA^T = D = \text{diag}(d_1, \ldots, d_n). \] (18)

Then the function $A: H^h \to H^h$ is an isometry of $H$, since $A$ is orthonormal. Indeed, using matrix notation,
\[ \|Ay\|_{H^h}^2 = (Ay)^TAy = y^T(A^TA)y = y^Ty = \|y\|_{H^h}^2. \]
It preserves the Stiefel manifold, since $v \in \text{St}(h, H)$ is equivalent to $vv^T = \text{Id}_{h \times h}$ and then
\[ vv^T = \text{Id} \iff A(vv^T)A^T = \text{Id} \iff Av(Av)^T = \text{Id}. \]
Moreover, the components of $Ax$ are orthogonal because $A$ diagonalizes $xx^T$
\[ Ax(Ax)^T = Axx^TA^T = D = \text{diag}(d_1, \ldots, d_n). \] (19)

By the observation above, there is a unique point in $\text{St}(h, H)$ that minimizes the distance from $Ax$.
Since $A$ is an isometry of $H$, and it preserves the Stiefel manifold, the minimum of the distance from $x$ is also unique.

Note that there is an explicit formula to compute this point. Combining the equations (17), (18), (19) and the formula (15), we obtain
\[ v_{\text{min}} = A^T \left( \frac{(Ax)_1}{\sqrt{d_1}}, \ldots, \frac{(Ax)_h}{\sqrt{d_h}} \right) = A^TD^{-1/2}Ax = B^{-1}x \] (20)
where $B = \sqrt{xx^T}$ is the unique symmetric positive definite matrix such that $B^2 = xx^T$.

**Step 3 - The minimum is not unique in case 2.** Let $v \in \text{St}(h, H)$ be a minimum of the distance from $x$. Let $X = \text{Span}(x_1, \ldots, x_h)$, and decompose $H$ as $X + X^\perp$ and every component of $v$ as $v_i = v_i^x + v_i^\perp$. Consider
\[ \bar{v} = (v_1^x - v_1^\perp, \ldots, v_h^x - v_h^\perp). \]
The vector $\bar{v}$ still lies on the Stiefel manifold.
Since the $x_i$ are linearly dependent, the $v_i$ could not all lie in $X$, but there is some $v_i^\perp \neq 0$ and then $v \neq \bar{v}$. Moreover,
\[ \|v_i - x_i\|_{H^h}^2 = \|\bar{v}_i - x_i\|_{H^h}^2 \quad \text{for all } i = 1, \ldots, h \]
and then there are at least two minima.

The above proof shows that the minimum can actually be easily computed.
The last step of the above proof is exemplified in the following example.

**Example 4.10.** Let $H = \mathbb{R}^2$ and $h = 2$. Hence $\text{St}(2, \mathbb{R}^2) = O(2)$. Consider $x \in H^2$ given by $x_1 = x_2 = (1, 1)$. The there are two points $v, \bar{v} \in \text{St}(2, \mathbb{R}^2)$ at minimal distance from $x$, namely
\[ v_1 = (1, 0), \ v_2 = (0, 1), \ \bar{v}_1 = (0, 1), \ \bar{v}_2 = (1, 0). \]
4.3.2. Properties of the projection. If $H$ is a Hilbert space and $h \in \mathbb{N}, h \geq 1$ a natural number, we define

$$\text{Ind}(h, H) = \{ (x_1, \ldots, x_h) \in H^h \mid x_1, \ldots, x_h \text{ are linearly independent} \}.$$ 

**Proposition 4.11.** This set is open, and the projection $\pi : \text{Ind}(h, V) \to \text{St}(h, H)$ is smooth.

**Proof.** Let $xx^T$ be defined as in Equation (17). Then $x \in \text{Ind}(h, H)$ if only if $\det(xx^T) \neq 0$. We proved in Equation (20) that inside this set $\pi(x) = (xx^T)^{-1/2}x$, and this is a smooth function.

**Proposition 4.12.** Consider $x = (x_1, \ldots, x_h) \in \text{Ind}(h, H)$. Let $z = (z_1, \ldots, z_h) \in \text{St}(h, H)$ be the unique point at minimum distance, as in Prop. 4.9. Let $V \subset H$ be the vector subspace spanned by $x_j$ for $j = 1, \ldots, h$: then $V$ is also the vector space spanned by $z_j$ for $j = 1, \ldots, h$.

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

**Proposition 4.13.** Consider $x \in \text{Ind}(h, H)$. Let $t > 0$ and $y = tx$. Then both $x$ and $y$ project to the same point $z \in \text{St}(h, H)$.

Both proofs follow from the relation (20).

4.3.3. Properties of the projection of linearly dependent frames. We define for convenience the complement of $\text{Ind}(h, V)$ as

$$\text{Dep}(h, V) = \{ (x_1, \ldots, x_h) \in V^h \mid x_1, \ldots, x_h \text{ are linearly dependent} \}.$$  \hfill (21)

By Proposition 4.11 this is a closed subset of $H^h$. Moreover $\text{Dep}(h, H)$ has empty interior, so it is a first category set.

The result 4.12 has a natural analogy in the case of linearly dependent vectors, as follows.

**Proposition 4.14.** Consider $x \in \text{Dep}(h, H)$. Let $V \subseteq H$ be a vector subspace of dimension $h$ such that $x \in V^h$: then there is a point at minimum distance $z = (z_1, \ldots, z_h) \in \text{St}(h, H)$ such that $V$ is spanned by $z_j$ for $j = 1, \ldots, h$. Vice versa for any $z \in \text{St}(h, H)$ at minimum distance from $x$, calling $V$ the vector space spanned by $z_j$ for $j = 1, \ldots, h$, we have that $x \in V^h$.

**Proof.** We fix $V$ as in the statement. Let $(x^n)_n \subseteq V^h$ be a sequence such that $x^n \to_n x$ and, for any fixed $n$, $x^n \in \text{Ind}(h, H)$. Let $z^n \in \text{St}(h, H)$ be the unique point at minimum distance from $x^n$. By Prop. 4.12 $z^n \in V^h$; up to a subsequence we can assume that $z^n \to_n z$. It is easily proved that $z \in \text{St}(h, H)$ and $z$ is a point at minimum distance from $x$; indeed for any $v \in \text{St}(h, H)$ we know that $\|v - x^n\|_H \geq \|z^n - x^n\|_H$ hence passing to the limit $\|v - x\|_H \geq \|z - x\|_H$.

The second claim follows from the theory of Lagrange multipliers: if $z$ is a point at minimum distance then $x = \lambda z$ where $\lambda \in \mathbb{R}^{h \times h}$ is a symmetric matrix. 

When the components of $x$ are are linearly dependent, the proof of Prop. 4.9 explains why the point at minimum distance is not unique. In particular, for any vector subspace $V \subseteq H$ of dimension $h$ and such that $x \in V^h$, there are at least two points at minimum distance, both contained in $V^h$. See also the related example 4.10.
When \( \dim(H) \geq h + 1 \) the above proposition provides another reason: indeed for any different \( V \) we obtain a different minimum.

Similarly the result 4.13 has a natural analogy in the case of linearly dependent vectors, as follows.

**Proposition 4.15.** Let \( x \in \text{Dep}(h,H) \), \( t > 0 \) and \( y = tx \): in this case the set of all points in \( \text{St}(h,H) \) at minimum distance from either \( x \) or \( y \) is the same.

**Proof.** We noted in the proof of 4.9 that those minima are also the maxima of the function \( g(v) = \sum_{i=1}^{h} \langle x_i, v_i \rangle_H \) (that was already defined in eqn. (14)).

4.3.4. **Projection of a Gaussian measure.** If \( H \) is a Hilbert space and \( h \in \mathbb{N}, h \geq 1 \) a natural number, we again define for convenience the complement of \( \text{Ind}(h,H) \) as \( \text{Dep}(h,H) \) (as in Eq. (21)).

To prove that the projection is defined almost everywhere, we need the following lemma.

**Lemma 4.16.** Let \( h \leq n \) be positive natural numbers and consider the linear space \( (\mathbb{R}^n)^h \) with the Lebesgue measure \( (\mathcal{L}^n)^h = \mathcal{L}^n \times \cdots \times \mathcal{L}^n \). Then the set \( \text{Dep}(h,\mathbb{R}^n) \subseteq (\mathbb{R}^n)^h \) is negligible.

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

Suppose that the lemma is true for \( h - 1 \) and decompose \( \text{Dep}(h,\mathbb{R}^n) \) as

\[
\text{Dep}(h - 1,\mathbb{R}^n) \times \mathbb{R}^n \cup \left\{ (x_1, \ldots, x_h) \in (\mathbb{R}^n)^h \mid (x_1, \ldots, x_{h-1}) \notin \text{Dep}(h-1,\mathbb{R}^n), x_h \in \text{Span}(x_1, \ldots, x_{h-1}) \right\}.
\]

The first set is negligible thanks to the inductive hypothesis. Moreover for each \( x_1, \ldots, x_{h-1} \in \mathbb{R}^n \), the set of \( x_h \in \mathbb{R}^n \) linearly dependent from them is a subspace of dimension at most \( h - 1 < n \) and so it is \( \mathcal{L}^n \)-negligible. By Fubini’s theorem, the second set is negligible too, and so also \( \text{Dep}(h,\mathbb{R}^n) \) is negligible.

**Lemma 4.17.** Let \( H \) be a Hilbert space, \( h \in \mathbb{N}, h \geq 1 \) and \( \text{St}(h,H) \) a Stiefel manifold. Let also \( \gamma \) be a non-degenerate Gaussian measure on \( H^h \). Then the set \( \text{Dep}(h,H) \) is \( \gamma \)-negligible.

**Proof.** If \( H \) is finite-dimensional, we assume that \( \dim(H) \geq h \) (otherwise \( \text{St}(h,H) \) is empty); then the result follows from Lemma 4.16. Let us consider then the case when \( H \) is infinite dimensional.

Fix an orthonormal frame \( \{e_i\}_{i \leq h} \) in \( H \), consider the projection \( f \)

\[
f : H \rightarrow \mathbb{R}^h, \quad x \mapsto (\langle x, e_1 \rangle_H, \ldots, \langle x, e_h \rangle_H)
\]

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

\[
f^h : (x_1, \ldots, x_h) \mapsto (f(x_1), \ldots, f(x_h)).
\]

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

By Lemma 4.16, \( \text{Dep}(h,\mathbb{R}^h) \) is negligible for the Lebesgue measure \( (\mathcal{L}^h)^h \). Since \( \gamma \) is non-degenerate, \( f^h \gamma \) is a non-degenerate Gaussian measure on \( (\mathbb{R}^h)^h \) and then it is absolutely continuous with respect to \( (\mathcal{L}^h)^h \). It follows that

\[
f^h \gamma (\text{Dep}(h,\mathbb{R}^h)) = 0
\]
and so \( \text{Dep}(h, H) \) is negligible.

**Corollary 4.18.** Assume \( H, h \) and \( \gamma \) as in the above lemma. Then for almost every \( x \in H^h \) there exists a unique point \( \pi(x) \in \text{St}(h, H) \) that realizes the minimum of the distance from \( x \), i.e.,

\[
d(\pi(x), x) = d_{\text{St}(h, H)}(x).
\]

**Proof.** By Proposition 4.9 for all points \( x \notin \text{Dep}(h, H) \) there exists a unique point at minimal distance on \( \text{St}(h, H) \). By the above lemma \( \text{Dep}(h, H) \) is negligible for the measure \( \gamma \).

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

We so summarize.

**Theorem 4.19.** Let \( H \) be a Hilbert space, \( h \in \mathbb{N}, h \geq 1 \) and \( \text{St}(h, H) \) a Stiefel manifold. Let also \( \gamma \) be a non-degenerate Gaussian measure on \( H^h \). Then the projection \( \pi \# \gamma \) is a well defined Radon probability measure on \( \text{St}(h, H) \).

**Proof.** To define \( \pi \# \gamma \) we restrict \( \gamma \) to \( \text{Ind}(h, H) \), and we push forward this restriction using the map \( \pi \). Theorem 7.1.7 in [3] ensures that \( \pi \# \gamma \) is Radon.

### 4.3.5. Properties of projected Gaussian measures.

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

We show two simple properties of this kind of probabilities.

**Lemma 4.20.** Suppose that \( \theta \) and \( \gamma \) are equivalent probability measures on a space \( \Omega \) 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 4.21.** Suppose that \( \theta \) and \( \gamma \) are equivalent non-degenerate Gaussian measures on \( H^h \). Then \( \pi \# \theta \) and \( \pi \# \gamma \) are equivalent.

A vice versa does not hold.

**Remark 4.22.** Let \( H \) and \( \gamma \) be as above. Let \( t > 0 \) and \( \theta \) be a rescaling of \( \gamma \), that is, \( \theta(B) = \gamma(tB) \) for any \( B \in H^h \) Borel set. Then the projections \( \pi \# \gamma \) and \( \pi \# \gamma \) are identical. This derives from Corollary 4.13. Note that \( \theta \) and \( \gamma \) are mutually singular, unless \( t = 1 \). So two mutually singular Gaussian measures can project to identical measures.

We already remarked that, if \( O : \mathbb{R}^h \to \mathbb{R}^h \) is a rotation, then the map \( v \mapsto Ov \) defined in Equation (16) is an isometry and sends \( \text{St}(h, H) \) into itself. It is quite easy to provide a Gaussian probability that is invariant wrt this action.

**Proposition 4.23.** Let \( \mu \) be a non-degenerate centered Gaussian measure on \( H \), and let \( \gamma = \mu^h = \mu \otimes \cdots \otimes \mu \) be the product probability on \( H^h \). Then \( \gamma \) is invariant with respect to the action of \( O \) and hence the projected measure \( \pi \# \gamma \) is a probability on \( \text{St}(h, H) \) that is invariant wrt the action \( v \mapsto Ov \).

**Acknowledgments.** We thank 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] L. Ambrosio, N. Gigli and G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, 2nd edition, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

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

[3] V. I. Bogachev, Measure Theory. Vol. I, II, Springer-Verlag, Berlin, 2007, URL http://dx.doi.org/10.1007/978-3-540-34514-5.

[4] V. I. Bogachev, Gaussian Measures, vol. 62 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1998.

[5] L. Breiman, Probability, Addison-Wesley, 1968.

[6] D. Burago, Y. Burago and S. Ivanov, A Course in Metric Geometry, vol. 33 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2001.

[7] G. Da Prato, An Introduction to Infinite-Dimensional Analysis, Springer, 2006.

[8] M. P. do Carmo, Riemannian Geometry, Mathematics: Theory & Applications, Birkhäuser Boston Inc., Boston, MA, 1992, Translated from the second Portuguese edition by Francis Flaherty.

[9] A. Edelman, T. A. Arias and S. T. Smith, The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl., 20 (1999), 303–353, URL http://dx.doi.org/10.1137/S0895479895290954.

[10] H. Federer, Geometric Measure Theory, Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969.

[11] N. Grossman, Hilbert manifolds without epiconjugate points, Proc. Amer. Math. Soc., 16 (1965), 1365–1371.

[12] P. Harms and A. Mennucci, Geodesics in infinite dimensional Stiefel and Grassmann manifolds, Comptes rendus - Mathématique, 350 (2012), 773–776, http://cvgmt.sns.it/paper/336/.

[13] D. G. Kendall, The diffusion of shape, Advances in applied probability, 428–430.

[14] D. G. Kendall, Shape manifolds, procrustean metrics, and complex projective spaces, Bulletin of the London Mathematical Society, 16 (1984), 81–121.

[15] A. N. Kolmogorov, La transformation de laplace dans les espaces linéaires, CR Acad. Sci. Paris, 200 (1935), 1717–1718.

[16] S. Lang, Fundamentals of Differential Geometry, vol. 191 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1999.

[17] H. Le and D. G. Kendall, The Riemannian structure of Euclidean shape spaces: A novel environment for statistics, The Annals of Statistics, 21 (1993), 1225–1271.

[18] C. Mantegazza and A. C. Mennucci, Hamilton–Jacobi equations and distance functions on Riemannian manifolds, Applied Math. and Optim., 47 (2003), 1–25.

[19] A. C. G. Mennucci, Regularity and variationality of solutions to Hamilton-Jacobi equations. I. Regularity, ESAIM Control Optim. Calc. Var., 10 (2004), 426–451 (electronic), URL http://dx.doi.org/10.1051/cocv:2004014.

[20] G. Sundaramoorthi, A. Mennucci, S. Soatto and A. 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 (2011), 109–145, URL http://dx.doi.org/10.1137/090781139.

[21] L. Younes, Computable elastic distances between shapes, SIAM J. Appl. Math., 58 (1998), 565–586 (electronic), URL http://dx.doi.org/10.1137/S0036139995287685.

[22] L. Younes, P. W. Michor, J. Shah and D. Mumford, A metric on shape space with explicit geodesics, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl., 19 (2008), 25–57, URL http://dx.doi.org/10.4171/RLM/506.

Received December 2015; revised June 2016.

E-mail address: eleonora.bardelli@sns.it
E-mail address: andrea.mennucci@sns.it