Probability Measures
on Infinite Dimensional Embedded Manifolds
with applications to
Computer Vision

Tesi di laurea

Candidata Eleonora Bardelli
Relatore Andrea Mennucci

Corso di Laurea Magistrale in Matematica
Anno Accademico 2011/2012
## Contents

1 Introduction
   1.1 Notation ............................................. 6

2 Motivation ............................................ 9
   2.1 Shape spaces ....................................... 10
   2.2 The Stiefel manifold ................................. 14
   2.3 The filtering problem ................................. 16
   2.4 Filtering and tracking ............................... 20

3 Gaussian measures .................................... 25
   3.1 Finite dimensional Gaussian measures ............... 25
   3.2 Gaussian measures in Hilbert spaces ............... 28
   3.3 The Cameron-Martin space ............................ 36
   3.4 Miscellaneous facts .................................. 43

4 Probability measures on manifolds .................. 47
   4.1 Hausdorff measures ................................... 47
   4.2 Projections .......................................... 51
   4.3 The exponential map .................................. 63
Chapter 1

Introduction

This thesis aims to provide examples of probability measures on infinite dimensional embedded manifolds. We try to generalize some approaches commonly used to define measures on finite dimensional manifolds. However, when studying probability on infinite dimensional spaces many difficulties arise, and hence some finite dimensional methods could fail.

The text is divided into three chapters and only in the last one probability on manifolds is addressed. In the first chapter, we discuss what motivated us to the study of probability measures on infinite dimensional embedded manifolds and explain the connection with computer vision. The second chapter provides an introduction to Gaussian measures in Hilbert spaces, which are widely used in the third chapter.

Computer vision is a branch of computer science that studies methods for processing and analyzing digital images in order to extract interesting features. For example the tracking problem consists of recognizing and following a moving object in a movie.

Objects in images are usually identified by their shape and different shape spaces have been proposed in the literature. Some of them are infinite dimensional, for example the space of closed plane curves. In the thesis we concentrate on a particular space of closed plane curves, previously proposed in [35, 31]. Our space is endowed with a differential structure and a Riemannian metric, which is closely related to a Stiefel manifold, and can be embedded in a infinite dimensional Hilbert space.

The problem of tracking is often formulated as a Bayesian filtering problem. The motion of the object is modeled by a stochastic equation and the probability distribution of its shape is estimated, conditioning on the current frame. The problem of Bayesian filtering has many applications in engineering, it is well studied in $\mathbb{R}^n$, and several methods to solve it exactly or approximately are known. However it is difficult to rigorously generalize these methods (and the problem itself) to infinite dimensional spaces.

It would be desirable to know something more on Bayesian filtering on some infinite dimensional shape spaces. Before that, it is at least necessary to study
some examples of probability measures on them.

In the second chapter we provide an introduction to Gaussian measures in Hilbert spaces, following [3]. Such measures have many properties in common with normal distributions in $\mathbb{R}^n$ and are characterized by their Fourier transform.

Unlike the finite dimensional case, absolute continuity of different Gaussian measures with respect to each other is not guaranteed in general and fails in some common situations, for example when translating or scaling a measure. To see this fact, a deeper analysis is needed, involving the Cameron-Martin space.

The third chapter deals with probability measures on manifolds. A few different approaches to endow a manifold with a measure are presented, all borrowed from the study of probability on finite dimensional manifolds, and their generalization to the infinite dimensional case is analysed.

A natural way to put a measure on an embedded manifold is provided by the restriction of the Hausdorff measure. A “Gaussian” generalization of it to finite codimension objects in a Hilbert space can be found in [13], but it is far less natural.

A simple way to sidestep the problem of defining a measure on a manifold is to use the push forward of a measure defined on the ambient space under some function.

A reasonable function to choose in finite dimension is the projection onto the nearest point of the manifold, which is unique for almost every point with respect to the Lebesgue measure. Unfortunately this kind of projection may be ill-defined in infinite dimensional spaces. In particular we show that, for any given Gaussian measure, there is a manifold such that the set of those points for which the projection does not exist is non negligible.

On the other hand, this approach can be suitable for Stiefel manifolds. Indeed, we prove that the projection is well-defined for almost every point, with respect to any Gaussian measure.

The push-forward of measures under the exponential map is examined last. In finite dimension, provided measures on tangent spaces are equivalent (mutually absolutely continuous) to the Lebesgue measure, the push-forward measures are equivalent, even when using exponential maps from different points. This fails in infinite dimension and we show that the same Gaussian measure, projected from the north or the south pole of a sphere, gives two measures singular with respect to each other.

### 1.1 Notation

In this Section we define some common notation, that are used throughout the text.

We denote by $\mathbb{N}$ the set of natural numbers and, by convenience for sequence numbering, we let it start from 1. As usual the sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, and the Euclidean norm in $\mathbb{R}^n$ by $|.|$. 
1.1. NOTATION

The symbol $S^1$ denotes the unit circle in $\mathbb{R}^2$

$$S^1 = \{ x \in \mathbb{R}^2 \mid \| x \| = 1 \}.$$ 

It is a submanifold of $\mathbb{R}^2$ and has a Riemannian structure induced by the Euclidean product in the ambient space $\mathbb{R}^2$. When integrating on $S^1$, we consider on it the Hausdorff measure, which coincide with the arc length measure.

Given a manifold $M$, we denote by

$$C^k(M, \mathbb{R}^n)$$

the set of functions from $M$ to $\mathbb{R}^n$ which are continuously differentiable $k$ times and functions in $C^0(M, \mathbb{R}^n)$ are just continuous. For example $C^1(S^1, \mathbb{R}^2)$ is the set of continuously differentiable functions from the unit sphere to $\mathbb{R}^2$.

The remaining of this section is devoted to fix the terminology and notation for probability.

Let $\Omega$ be a set and $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ a $\sigma$-algebra. We call the pair $(\Omega, \mathcal{F})$ a measurable space. By the term measure we mean a countably additive nonnegative finite function $\mu : \mathcal{F} \to [0, +\infty)$. When we need to talk about other kind of measures, we explicitly say signed measure or not finite measure. We call such a triple $(\Omega, \mathcal{F}, \mu)$ a measure space. When $\mu(\Omega) = 1$ the triple is called a probability space.

Given a set $\Omega_1$, a measurable space $(\Omega_2, \mathcal{F}_2)$ and a family $\mathcal{G}$ of functions $\Omega_1 \to \Omega_2$, we call $\sigma$-algebra generated by $\mathcal{G}$ the smallest $\sigma$-algebra with respect to whom all functions in $\mathcal{G}$ are measurable.

Given two measures $\mu$ and $\nu$ on $\mathcal{F}$, we denote by

$$\mu \ll \nu$$

the absolute continuity of $\mu$ with respect to $\nu$. If the two measures are such that $\mu \ll \nu$ and $\nu \ll \mu$ we say that they are equivalent and denote this fact by

$$\mu \sim \nu.$$ 

On the opposite side, we say that two measures are orthogonal or singular to each other if their are concentrated on disjoint sets.

If $\Omega$ is a topological space, the Borel $\sigma$-algebra $\mathcal{B}(\Omega)$ is the smallest $\sigma$-algebra that contains all open sets. In the following, when not otherwise specified, measures on a topological space $\Omega$ are assumed to defined on $\mathcal{B}(\Omega)$.

A measure $\mu$ on a topological space $\Omega$ is called a Radon measure if for all $B \in \mathcal{B}(X)$ and for all $\varepsilon > 0$ there exists a compact set $C$ such that $\mu(B \setminus C) < \varepsilon$.

Let $(\Omega_1, \mathcal{F}_1, \mu)$ be a measure space, $(\Omega_2, \mathcal{F}_2)$ a measurable space. A random variable $f$ is a measurable function $\Omega_1 \to \Omega_2$. If $\Omega_2 = \mathbb{R}$, $f$ is called a real random variable.

The measure $f_\sharp \mu$ on $\mathcal{F}_2$ defined by

$$f_\sharp \mu(A) = \mu(f^{-1}(A))$$
is called the image of $\mu$ under the function $f$. Note that the measure $f_*\mu$ is unchanged if $f$ is modified on a null set. A change of variables formula relates the integrals with respect to a measure and an image of it, see Theorem 4.1.11 of [11] for more details.

**Proposition 1.1.1.** Let $\varphi : \Omega_2 \to \mathbb{R}$ a measurable function. Then $\varphi$ is integrable with respect to $f_*\mu$ if and only if $\varphi \circ f$ is integrable with respect to $\mu$ and in that case one has

$$
\int_{\Omega_2} \varphi \, d(f_*\mu) = \int_{\Omega_1} \varphi \circ f \, d\mu.
$$
Chapter 2

Motivation

One topic in computer vision is the analysis of a digital image to recognize some objects in it. The image can be static or be a dynamic video, in which a object should be detected and then followed as it moves.

The problem of recognizing the contours of objects in a fixed image is often referred to as image segmentation, while the problem of following a moving object is called tracking. Of course the first should be addressed also when dealing with the second.

In the last decades of the XX century, the most used technique to segment an image was to first identify edges, and then try to connect them to form a whole contour.

The current approach works in the opposite direction. The algorithm is initialized with a closed curve and then moves it to match a contour in the image. This is done by defining an energy function on all curves, which is supposed to be small on objects contour, and looking for the minimum of it.

This approach was introduced in [18] and is now widely used. It is referred to as active contours.

The energy function usually contains an edge-based term, which attracts the minimum towards the edges, and also a smoothness term, which prevents the minimum from being too irregular.

The active contours method raises some interesting mathematical problems. The main one is how to minimize a function whose domain is a space of curves. In turn, we have to define more properly what a curve is and the geometric structure of the space of curves.

Defining properly a space of curves, or a shape space as it is called, is an important issue because a good choice can make the minimization easier. This problem has been studied also in recent years, with the proposal of some new shape spaces, for example in [25, 35, 31, 26, 7].

The most common minimization algorithms are based on gradient flow techniques. To implement such a technique, it is necessary to have a differential structure and a metric on the shape space. This is the reason for which some recent developments focus on shape spaces with a linear or differential structure.
CHAPTER 2. MOTIVATION

Figure 2.1: An example of tracking. Two overlapped objects move in opposite directions.

and study metrics on them.

2.1 Shape spaces

There are many ways to define a shape space. Many have been used in literature and each has his own advantages. As explained above, it is useful to define on the shape space also a differential structure and a metric. In doing that, we shall also keep in mind computational issues, since in the end we aim to numerically implement the theory.

The intuitive idea of a “contour” is that of a subset of $\mathbb{R}^2$. This a structureless dataset, not apt to calculus. We will need to give to the space of contours a form of differential structure.

One simple way is to see contours as the image of a closed curve, i.e. a function $c: S^1 \rightarrow \mathbb{R}^2$ with some regularity. A contour could obviously be represented by a curve in many ways; even if we consider only non intersecting curves, a curve can always be reparametrized to get a different curve with the same image. This requires a quotient: a proper definition of shapes as a quotient of curves under reparametrization is given in [24] and [25].

The main limitation when dealing with curves, is that they are always connected. For example this can be a difficulty when tracking two overlapped figures that move in opposite direction. In this situation, the shape evolves creating two connected components. Figure 2.1 provides an example of this phenomenon.

A different approach, that addresses this limitation, is the so called level set method, introduced in [27]. In this case, one represent a contour as the zero set of a regular function. Again there are many ways to represent the same contours, and a quotient should be performed.

In both cases, the shape space is infinite dimensional.

We now provide proper definitions about curves and show some examples of Riemannian-like metrics on spaces of curves.

By the word curve we mean a function $c: S^1 \rightarrow \mathbb{R}^2$. The set of all continuously differentiable curves is $C^1(S^1, \mathbb{R}^2)$. Given a differentiable curve $c(t)$, we denote by $\dot{c}$ or $\frac{d}{dt}c$ its derivative. The curve $c$ is said to be immersed if its
2.1. SHAPE SPACES

derivative never vanish
\[ \dot{c}(t) \neq 0 \quad \text{for all } t \in S^1 \]
and we denote by \( M \) the set of \( C^1 \) immersed curves.

The set \( C^1(S^1, \mathbb{R}^2) \) has a linear structure and it is a Banach space with the norm
\[ \| c \|_{C^1} = \sup_{t \in S^1} |c(t)| + \sup_{t \in S^1} |\dot{c}(t)|. \]
The immersed curves are an open subset of this Banach space, and so have a differential structure as well. A tangent vector in a point \( c \) can be canonically identified with an element of the Banach space \( C^1(S^1, \mathbb{R}^2) \) and so we indicate tangent vectors as \( C^1 \) functions \( h: S^1 \rightarrow \mathbb{R}^2 \).

Let \( \text{Diff}(S^1) \) be the group of diffeomorphism of the circle, i.e. the set of \( C^1 \) functions \( \varphi: S^1 \rightarrow S^1 \) such that \( \varphi^{-1} \) is continuously differentiable as well. This group acts by reparametrization on immersed curves,
\[ \text{Diff}(S^1) \times M \rightarrow M, \quad (\varphi, c) \mapsto c \circ \varphi \]

A geometric curve is an element of the quotient
\[ M/\text{Diff}(S^1). \]
This quotient turns out to be almost a manifold modeled on \( C^1(S^1, \mathbb{R}^2) \), but has some singular points, see [6] for more details.

Quantities that do not depend on parametrization and can be defined on the quotient are often referred to as geometric quantities. Geometric energy functions and metrics are preferred in computer vision, since they better represent the actual contour and are less influenced by the representation.

A common way to define geometric quantities is to evaluate non geometric quantities on the parametrization by arc length. If \( c \) is an immersed curve, the parametrization by arc length of \( c \) is the curve \( c \circ \varphi \), such that \( \varphi \in \text{Diff}(S^1) \) preserves orientation and the norm of the derivative \( \frac{d}{dt} c \circ \varphi \) is constant on \( S^1 \).

Some common geometric quantities of this kind are the following.
Let \( c \) be an immersed curve and \( h \in C^1(S^1, \mathbb{R}^2) \). The derivative by arc length in \( c \) is
\[ \partial_{s,c} h(t) = \frac{\dot{h}(t)}{\| \dot{c}(t) \|}. \]

If \( f: \mathbb{R}^2 \rightarrow \mathbb{R}^n \) is a measurable function with values in \( \mathbb{R}^n \) for some \( n \in \mathbb{N} \), the integral by arc length of \( f \) on \( c \) is
\[ \int_c f \, ds = \int_{S^1} f \circ c(t) \| \dot{c}(t) \| \, dt, \]

clearly this definition makes sense also when \( f \) is defined only on the image of the curve \( c \).
The centroid of an immersed curve $c$ is

$$\text{avg}(c) = \int_c c \, ds$$

and its length is

$$\text{len}(c) = \int_c 1 \, ds.$$ 

In applications it is also useful to perform a quotient by translation and scaling. Curves up to translations and scaling are often identified with curves of centroid the origin and length 1. We denote the set of immersed curves with centroid the origin and length 1 as

$$M_d = \{ c \in M_i \mid \text{avg}(c) = 0 \text{ and } \text{len}(c) = 1 \}.$$ 

By the inverse function theorem, $M_d$ is a manifold modeled on $C^1(S^1, \mathbb{R}^2)$, see [19, Theorem 5.9] for reference about this theorem in Banach spaces.

Before talking about metrics, note that the differential structure we have defined on $M_i$ is not modeled on a Hilbert space, which the natural place where a metric can be defined. What is usually done to overcome this problem, is to define a pointwise metric and then ad hoc prove that some energies have a gradient and it is sufficiently regular to admit a gradient flow.

The simpler pointwise metric that we can consider on $M_i$ is the $L^2$ metric. Given a curve $c$ and two tangent vectors $h, k \in C^1(S^1, \mathbb{R}^2)$, the $L^2$ metric is defined as

$$\langle h, k \rangle_{L^2,c} = \int_{S^1} \langle h(t), k(t) \rangle \, dt$$

where $\langle \cdot, \cdot \rangle$ is the standard scalar product on $\mathbb{R}^2$. With this metric, the shape space $M_i$ is a subspace of $L^2(S^1, \mathbb{R}^2)$.

The first improvement that can be done to the above metric is to make it geometric. This leads to the definition of the $H^0$ metric,

$$\langle h, k \rangle_{H^0,c} = \int_c \langle h, k \rangle \, ds.$$ 

This metric is widely used in computer vision, and often implicitly assumed when doing gradient flows without talking about a metric.

The metric $H^0$ is sometimes defined as

$$\langle h, k \rangle_c = \frac{1}{\text{len}(c)} \int_c \langle h, k \rangle \, ds,$$

adding a conformal factor $\text{len}(c)$ to make it scale invariant, i.e.

$$\langle h, k \rangle_c = \langle h, k \rangle_{\lambda c} \quad \text{for all } \lambda > 0.$$ 

The conformal factor only changes the velocity of geodesics and gradient flows but not their trajectories.
The $H^0$ metric also induces a distance on $M_i$, that we call $H^0$ distance. As usual, the distance of two curves is the infimum of the length of the paths in $M_i$ between them.

However, the $H^0$ metric has some undesirable features. It was shown in [23] that the $H^0$ distance induces a pathological distance on the quotient of $M_i$ by parametrization, i.e. any two curves can be made arbitrarily close by reparametrizing them.

Moreover gradient flows of some common energies are very irregular and quickly evolve towards non-smooth curves, while others are even ill defined. In general, $H^0$ gradient flows are very sensitive to noise and numerically unstable. Regularization terms can be added to the energies, but this changes the minimization problem to be solved. More details about these phenomena can be found in [32].

Other metrics, presented for example in [32, 8, 25], use first or higher derivatives of tangent vectors. These are usually referred to as Sobolev-type metrics. An example of such a metric is the $H^1$ metric, defined as

$$
\langle h,k \rangle_{H^1,c} = \frac{1}{\text{len}(c)} \int_c \langle h,k \rangle \, ds + \text{len}(c) \int_c \langle \partial_s c h, \partial_s c k \rangle \, ds ,
$$

where the factors $\text{len}(c)$ make it scale invariant.

This kind of metrics address some of the problems of the $H^0$ metric. The distance induced on the quotient of $M_i$ by reparametrization is not identically null, as proven in [21], and gradient flows are more regular compared to the $H^0$ gradient flows (see [32]). Moreover, as noted in [33], some ill defined gradient flows with respect to the $H^0$ metric are well defined with respect to Sobolev-type metrics, and then more energies can be minimized with this kind of metrics.

Some metrics could also be designed to outline some geometric features of the curves motion, or to induce easily computable geodesics and gradients.

Easiness of computation is the main feature of a metric introduced in [34, 35] and recently proposed in [31]. This metric is closely related to a Stiefel manifold and it is for us the main motivation for studying that kind of manifolds.

The original metric of [35] was defined only on the submanifold $M_d$ of curves with centroid the origin and length 1. Given a curve $c \in M_d$ and two tangent vectors $h, k \in T_c M_d$ the metric is defined as

$$
\langle h,k \rangle_{St,c} = \int_c \langle \partial_s c h, \partial_s c k \rangle \, ds .
$$

In next section we explain the connection with Stiefel manifolds, which is not at all evident from this definition.

In the remaining of this section, we see some geometric consideration that allow to extend a metric defined on $M_d$ to the whole $M_i$, as proposed in [31].

Roughly, to get immersed curves from $M_d$, we should specify the centroid and the scale. This is done by the function $\Phi$

$$
\Phi: \quad \mathbb{R}^2 \times \mathbb{R} \times M_d \rightarrow M_i \\
(x, \lambda, c) \quad \mapsto \quad x + e^\lambda c .
$$
It is easy to show that $\Phi$ is a diffeomorphism with inverse

$$
\Phi^{-1}: \quad M_i \rightarrow \mathbb{R}^2 \times \mathbb{R} \times M_d
$$

$$
c \mapsto \left( \text{avg}(c), \log \text{len}(c), \frac{c - \text{avg}(c)}{\text{len}(c)} \right).
$$

On $\mathbb{R}^2$ and $\mathbb{R}$ we can consider the standard scalar product as a metric. Given a metric $\langle \cdot , \cdot \rangle_{M_d}$ on $M_d$, we can consider the product metric on $\mathbb{R}^2 \times \mathbb{R} \times M_d$ and define a metric on $M_i$ as the pull back under $\Phi^{-1}$ of this product metric.

For example, given a curve $c \in M_i$ and tangent vectors $h, k \in T_c M_i$, the extension of the metric $\langle \cdot , \cdot \rangle_{St}$ defined above can be computed as follows. First decompose the curve and the tangent vectors using $\Phi$,

$$
c^d = \frac{c - \text{avg}(c)}{\text{len}(c)}
$$

$$
(h^a, h^l, h^d) = D\Phi^{-1}h \quad (k^a, k^l, k^d) = D\Phi^{-1}k
$$

and then compute the product metric

$$
\langle h, k \rangle_{St,c} = \langle h^a, k^a \rangle + h^l k^l + \int_{c^d} \partial_{s,c^d} h^d \partial_{s,c^d} k^d \, ds.
$$

The differential of $\Phi^{-1}$ can be written in a closed form, see [31] for details, and this allows the gradient of some commonly used energies to be explicitly written in a nice form.

With respect to the metric $\langle \cdot , \cdot \rangle_{St}$ on $M_i$, centroid translations, scale changes and deformations of the curve are orthogonal. Moreover, the relative weights of these components can be tuned adding coefficients as follows

$$
\langle h, k \rangle_c = \lambda_a \langle h^a, k^a \rangle + \lambda_l h^l k^l + \lambda_d \int_{c^d} \partial_{s,c^d} h^d \partial_{s,c^d} k^d \, ds.
$$

The ability to separate these components is important in computer vision applications, in fact an object is usually identified by the $M_d$ component of a curve, while position and scale can depend on the location of the camera and other minor factors.

### 2.2 The Stiefel manifold

**Definition 2.2.1.** Let $p \in \mathbb{R}$ and $H$ a Hilbert space. The Stiefel manifold $\text{St}(p, H)$ is the subset of $H^p$ consisting of orthonormal $p$-uples of vectors.

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

It is easy to check that the Stiefel manifold is actually a manifold, modeled on $H^p$, even when $H$ is infinite dimensional. This can be done using the inverse function theorem, see [19, Theorem 5.9] for reference.
2.2. THE STIEFEL MANIFOLD

The space $H^p$ is naturally a Hilbert space, with the scalar product

$$\langle (v_1, \ldots, v_p), (w_1, \ldots, w_p) \rangle_{H^p} = \sum_{i=1}^{p} \langle v_i, w_i \rangle_{H}$$

and this scalar product induces also a Riemannian metric on the Stiefel manifold.

With a little abuse of notation, we write $\text{St}(2, C^0)$ to indicate the set

$$\text{St}(2, C^0) = \{(e, f) \in \text{St}(2, L^2(S^1, \mathbb{R}^2)) \mid e \text{ and } f \in C^0(S^1, \mathbb{R}^2) \}.$$  

The set $\text{St}(2, C^0)$ can be seen also as a submanifold of $C^0 \times C^0$ and it is then a manifold modeled on $C^0(S^1, \mathbb{R}^2)$. Let also $\text{St}_0$ be the open subset in $\text{St}(2, C^0)$ of pairs $(e, f)$ that never vanish simultaneously,

$$\text{St}_0 = \{(e, f) \in \text{St}(2, C^0) \mid e(t)^2 + f(t)^2 \neq 0 \text{ for all } t \in S^1 \}.$$  

A two fold covering of $M_d$ can be defined on $\text{St}_0$ and the pull-back of the metric $\langle \cdot, \cdot \rangle_{\text{St}}$ defined at page 13 is the metric induced by the inclusion of $\text{St}_0$ in $\text{St}(2, C^0)$.

The two fold covering $\Psi : \text{St}_0 \to M_d$ is defined by the conditions that

$$\frac{d}{dt} \Psi(e, f)(t) = \frac{1}{2}(e^2 - f^2, 2ef)(t) \quad \text{for all } t \in S^1$$

$$\text{avg}(\Psi(e, f)) = 0$$

for every $(e, f) \in \text{St}_0$.

We now check that $\Psi$ is well defined and give an overview of how defining a local inverse, referring the reader to [35] for a complete proof of the fact that $\Psi$ is a two fold covering and it is an isometry.

All the following (and the definition) is much clearer if we identify the plane $\mathbb{R}^2$ with the complex plane $\mathbb{C}$.

Let $(e, f) = e + if$ be couple of functions in $\text{St}_0$ and consider its square $(e + if)^2$. The fact that $e$ and $f$ are orthonormal, implies that

$$\int_{S^1} e^2 - f^2 = 1 - 1 = 0$$

$$2 \int_{S^1} ef = 0.$$  

This means that $(e + if)^2$ is the derivative of some closed curve $c \in C^1(S^1, \mathbb{R}^2)$. The fact that $e$ and $f$ never vanish simultaneously, implies also that $c \in M_i$. Moreover the length of $c$ is

$$\text{len}(c) = \int_{S^1} |(e + if)^2| = \int_{S^1} e^2 + f^2 = 2$$

The integral curve $c$ is determined up to translation, so we can choose it in a unique way to get $\text{avg}(c) = 0$. Now it is sufficient to scale the curve $c$ by a factor $\frac{1}{2}$ to get that $\Psi$ is well defined.
Conversely, if \( c \in M_d \), its derivative never vanish, and so it is possible to extract a continuous square root of \( 2 \dot{c} \). Let \( e, f \) be such that \( 2 \dot{c} = 2(e + if)^2 \). The property that \( c \) is closed implies
\[
\oint_{S^1} e^2 = \oint_{S^1} f^2 = 0.
\]
The length of \( c \) can be computed as
\[
1 = \text{len}(c) = \oint_{S^1} |\frac{1}{2}(e + if)|^2 = \frac{1}{2} \oint_{S^1} e^2 + \frac{1}{2} \oint_{S^1} f^2
\]
and so we get that \( e \) and \( f \) are orthonormal.

### 2.3 The filtering problem

In this section we introduce the so called filtering problem. This problem often arises in engineering, when an estimate of some parameters, based on noisy measurements, is needed.

After that, we explain how this problem fit to the tracking problem of computer vision.

As a naming convention, in this section we use capital Latin letters to indicate random variables and the corresponding lower-case letter to indicate elements in their range. For example, we denote by \( X \) a \( \mathbb{R}^n \)-valued random variable and by \( x \) an element of \( \mathbb{R}^n \). Greek letters are also used to indicate some random variables.

In engineering a common problem is to give an estimate of the physical state of a system, based on some measurements. The system evolves in time accordingly to known, but possibly probabilistic, laws and the measurement does not necessarily give complete information and be affected by errors.

Suppose given a probability space \((\Omega, \mathcal{F}, \mu)\) on which all the subsequent random variables and stochastic processes are defined.

The state of the system is represented by a discrete time random variable \( X_t \) with values in \( \mathbb{R}^n \), called state vector. The evolution in time is modeled as
\[
X_{t+1} = f(X_t, \xi_t)
\]
where \( f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \) is a given transition function and the system noise \( \xi_t \) is a sequence of independent \( \mathbb{R}^m \)-valued random variables of known distribution. The \( \xi_t \) are also independent of the past state vectors \( X_0 \ldots X_{t-1} \). The starting state \( X_0 \) is supposed to be known.

The measurement is another random variable \( Y_t \) defined as
\[
Y_t = g(X_t, \eta_t)
\]
2.3. THE FILTERING PROBLEM

where the measurement function \( g \) is given and the measurement noise \( \eta_t \) is a sequence of independent random variables of known distribution. The \( \eta_t \) are also independent of past state vectors and system noise.

The measurement \( Y_t \) became available at some moment of time. The filtering problem consist of giving the “better” estimate of \( X_t \) knowing all the measurement until time \( t \), that is the values \( y_1 \ldots y_t \) of the variables \( Y_1, \ldots, Y_t \). Following a common notation, we denote a tuple \( y_1 \ldots y_t \) by \( y_{1:t} \).

The estimate can be done computing the conditional probability distribution of \( X_t \) given the measurements \( Y_{1:t} = y_{1:t} \),

\[
p(X_t | Y_{1:t} = y_{1:t}).
\]

This probability is often called the posterior.

Before going on, we spend a few words about the posterior probability, saying what we mean by that word and symbol.

Given two events \( A \) and \( B \subseteq \Omega \), such that \( P(B) \neq 0 \), there is no doubt on what the conditional probability is. The probability of \( A \) given \( B \) is

\[
\frac{\mu(A \cap B)}{\mu(B)}
\]

and similarly can be defined the conditional probability distribution of a random variable \( X \) given a non negligible event \( B \). When conditioning on negligible events, as it can easily happens in the above case, this should be refined.

Consider two random variables \( X \) and \( Y \) with values in \( \mathbb{R}^n \) and \( \mathbb{R}^m \) respectively. Suppose the distribution of the couple \( (X, Y) \) has density with respect to the Lebesgue measure on \( \mathbb{R}^n \times \mathbb{R}^m \) and call \( f(x, y) \) the density. Let also \( f_Y \) be the density of \( Y \) with respect to \( \mathcal{L}^m \) (the \( \sharp \) notation is defined in Section 1.1),

\[
f_Y(y) = \int_{\mathbb{R}^n} f(x, y) \, d\mathcal{L}^n(x).
\]

Then the probability of \( X \) conditioned to \( Y = y \) is an absolutely continuous measure with respect to \( \mathcal{L}^n \) defined by the density

\[
p(X|Y = y)(\cdot) = \frac{f(\cdot, y)}{f_Y(y)}
\]

if \( f_Y(y) \neq 0 \) and identically 0 otherwise.

We wish to remark that the symbol \( p(X|Y = y) \) denotes a function, which is a density with respect to the Lebesgue measure. We rarely refer the measure itself, and when needed we indicate it by \( p(X|Y = y) \mathcal{L}^n \), although this notation make sense only in case \( X \) and \( Y \) are absolutely continuous with respect to the Lebesgue measure.

It holds an integration by part formula

\[
(X, Y)_t \mu = p(X|Y = y) \mathcal{L}^n \, dY_t \mu(y),
\]
which means

$$\int_{\mathbb{R}^n \times \mathbb{R}^m} \varphi(x, y) \, d(X, Y)_{t} \mu(x, y) =$$

$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \varphi(x, y) p(X|Y = y)(x) \, d\mathcal{L}^n(x) \, dY_{t} \mu(y) =$$

$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \varphi(x, y) \frac{f(x, y)}{f_Y(y)} \, d\mathcal{L}^n(x) f_Y(y) \, d\mathcal{L}^m(y)$$

for all integrable \( \varphi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \).

Conditional probability can be defined in a very general setting, asking that the integration by parts formula still holds. However not much of what we say in the following is meaningful without additional hypothesis and in engineering literature all the measures are often implicitly assumed to be absolutely continuous with respect to the Lebesgue measure. For this reasons we restrict our presentation of the filtering problem to random variables with density with respect to the Lebesgue measure.

Back to the filtering problem, we have a model for the evolution of a stochastic process \( X_t \). At some time we get a measurement \( y_t \), which we would like to process real-time to get the posterior probability \( p(X_t|Y_{1:t} = y_{1:t}) \). This information is of course contained in the model and an inductive formula for the posterior can be written.

The starting point to write the inductive formula is the relation

$$p(X_t|Y_{1:t} = y_{1:t})(x) =$$

$$= \frac{p(Y_t|X_t = x)(y_t)}{p(Y_t|Y_{1:t-1} = y_{1:t-1})(y_t)} p(X_t|Y_{1:t-1} = y_{1:t-1})(x)$$

which make sense when \( p(Y_t|Y_{1:t-1})(y_t) \neq 0 \). Note that, by the integration by parts formula, this is true for \( (Y_t, Y_{1:t-1}) \mu \) almost every \( y_t, y_{1:t-1} \).

A closed form formula is supposed to be available for \( p(Y_t|X_t = x) \) and \( p(X_t|X_{t-1} = x_{t-1}) \). In real world models, these can usually be easily deduced from the model. From these and the posterior at time \( t-1 \), the term \( p(X_t|Y_{1:t-1} = y_{1:t-1}) \), usually referred to as prior, can be written in integral form

$$p(X_t|Y_{1:t-1} = y_{1:t-1})(x) =$$

$$= \int_{\mathbb{R}^n} p(X_t|X_{t-1} = x')(x) p(X_{t-1}|Y_{1:t-1} = y_{1:t-1})(x') \, d\mathcal{L}^n(x').$$

The term \( p(Y_t|Y_{1:t-1} = y_{1:t-1}) \) needs not to be computed, because it just normalizes the density to have integral equal to 1.

Past data is supposed to have been processed yet, leading to the posterior at time \( t-1 \) (or to an approximation of it), so what is needed to compute the posterior at time \( t \) is to do an integral.

In some cases, that integral can be explicitly solved. For example, this is the case when the transition and measurement functions are linear, and noise and
starting state random variables have Gaussian laws. The algorithm to compute
the posterior in this case is called Kalman filter, see [2, 17] for a more detailed
description.

Otherwise, the posterior should be approximated. This is done by algorithms
like the extended Kalman filter, see [2] for details, and sequential importance
sampling or particle filtering methods, see [2, 15].

Algorithms that compute the exact or approximate posterior using the in-
ductive formula above are in general called Bayesian filters.

A limitation of the above description of the filtering problem is the fact that
the random variables $X_t$ and $Y_t$ are asked to take values in $\mathbb{R}^n$ and $\mathbb{R}^m$ for some
$n$ and $m$. Actually, this is not really needed.

Suppose $X_t$ takes values in $\Omega_X$ and $Y_t$ in $\Omega_Y$ for each $t$ and to have some
“reference” measures $\mu_t$ on $\Omega_X$ and $\nu_t$ on $\Omega_Y$. Replace the hypothesis that all
random variables have density with respect to the Lebesgue measure with the
one that, for every $n, m \in \mathbb{N}$, for every $t_1 \ldots t_n, s_1 \ldots s_m \in \mathbb{N}$, the law of

$$(X_{t_1}, \ldots , X_{t_n}, Y_{s_1}, \ldots , Y_{s_m})$$

has density with respect to the product measure

$$\mu_{t_1} \otimes \cdots \otimes \mu_{t_n} \otimes \nu_{s_1} \otimes \cdots \otimes \nu_{s_m} .$$

Under this hypothesis, the conditional probability can be defined as in the
case of random variables with density with respect to the Lebesgue measure
and it easy to see that what we have said about the calculation of the posterior
make sense.

The problem of this generalization is how to choose the reference measures
such that the hypothesis of absolute continuity is satisfied by the dynamic and
measurement models, which often are given and should model real world.

In the case the reference measures are Lebesgue measures, the hypothesis of
absolute continuity is very reasonable and satisfied but nearly all models used
in engineering.

Another situation in which the absolute continuity hypothesis is easily met is
when some variables take values in an embedded manifold $M$ and we choose as
reference measure an Hausdorff measure (for the definition of Hausdorff measure
and further reference see 4.1). Again a lot of commonly used models satisfy the
hypothesis.

Things get more complicated when trying to formulate the filtering problem
in a infinite dimensional Hilbert space or manifold. We talk about probability
measures on Hilbert spaces, with special attention to Gaussian measures, from
Section 3.2 and there we see that absolute continuity surprisingly fails in some
common situations, for example when considering translated or scaled measures.
For this reason it could be hard, if not impossible, to choose some good reference
measures for a given model. We know no rigorous formulation of the filtering
problem that works well in infinite dimensional spaces, and in literature the
approach is more heuristic than rigorous.
2.4 Filtering and tracking

Recall that the tracking problem consists of tracking the motion and deformation of an object in a sequence of digital images. This problem can be given a formulation very similar to the one of the filtering problem.

The state vector $X_t$ is the contour of a real object. Real world objects are not usually thought in three dimensions, but identified with their projection on a plane, and $X_t$ is a random variable with values in a shape space $S$, for example (a finite dimensional subspace of) regular closed curves with values in $\mathbb{R}^2$.

The $t$-th image $I_t$ is instead thought of as a measurement. To get a filtering problem, one should also specify a dynamic model for the state vector and a model for the measurement, i.e. which process leads from the projection of the real world on a plane to the image. Before making some examples of these models, we spend a few words on the spaces in which the random variables take values.

The image is usually regarded as matrix of pixels, i.e. an element of $\mathbb{R}^{n_1 \times n_2}$ with $n_1$ and $n_2 \in \mathbb{N}$ the sides’ size, and so $I_t$ takes values in a finite dimensional vector space.

In the converse, there are various shape spaces. Some of them are finite dimensional vector spaces or manifolds and for these the Bayesian filtering techniques make sense and have been successfully used in literature, see [16, 28] for example. Other useful shape spaces are not finite dimensional, for example the Stiefel manifold presented in Section 2.2 which is an infinite dimensional manifold embedded in a Hilbert space. Heuristic algorithms, which mimic Bayesian filtering algorithms, have been proposed also for these spaces, see for example [29]. To our knowledge there is no rigorous formulation of this techniques and few examples of probability measures on these shape spaces have been studied.

We present examples of dynamic and measurement models both for $\mathbb{R}^n$ and infinite dimensional manifolds shape spaces.

The dynamic model obviously depends on what kind of objects one needs to follow. Accurate models can be done when one knows in advance the kind of objects being tracked. In the case of a general purpose tracker, there is no prior knowledge of the object’s motion. The simpler model is a kind a “Brownian motion”, the object evolves accordingly to

$$X_{t+1} = X_t + \xi_t$$

where $\xi_t$ is a noise random variable and $X_t$ takes values in $\mathbb{R}^n$. Usually $\xi_t$ has a distribution clustered around 0 and the equation just models the fact that at time $t$ the object is likely to be close to where it was at time $t-1$.

If the shape space is a manifold $M$, we can consider a noise random variable $\xi_t$ with values in the tangent space $T_{X_t} M$ and write the model as

$$X_{t+1} = \exp_{X_t} \xi_t$$

where $\exp_{X_t}$ is the exponential map based in $X_t$, see Section 4.3 for the definition and further reference.
Another general purpose model ask also some coherence to the velocity of the object. The state vector is a couple \((X_t, v_t)\), where \(v_t\) represents the velocity of the object at time \(t\), and when \(X_t\) takes values in \(\mathbb{R}^n\) the model looks like

\[
X_{t+1} = X_t + v_t + \xi_t \\
v_{t+1} = X_{t+1} - X_t
\]

where \(\xi_t\) is a noise random vector, usually clustered around 0.

If the shape space is a manifold, \(v_t\) and \(\xi_t\) belong to \(TX_t M\) and the model is defined by the equations

\[
X_{t+1} = \exp_{X_t} (v_t + \eta_t) \\
v_{t+1} = -\exp_{X_t}^{-1} (X_{t+1} - X_t).
\]

Regarding the measurement model, is often defined imposing the value of the conditional probability \(p(I_t|X_t = x_t)\). Choose an energy \(E(i_t, x_t)\), defined on \(\mathbb{R}^{n_1 \times n_2} \times S\), which takes a local minimum when \(x_t\) is on the contour of an object, and define

\[
p(I_t|X_t = x_t) = \frac{1}{z} e^{-E(i_t, x_t)} \mathcal{L}^{n_1 \times n_2}
\]

where \(z \in \mathbb{R}\) is a normalization factor. In this way we are not really modelling the measurement, but just saying that the chosen energy is likely to be minimized on the contours of objects.

An example of energy that can be used in the definition above is the Chan-Vese energy. Before defining it, we should introduce some notation. Let \(x\) be a shape in the shape space \(S\) and \(i \in \mathbb{R}^{n_1 \times n_2}\) an image. Denote by \(D\) the set

\[
D = \{(a, b) \mid 1 \leq a \leq n_1, 1 \leq b \leq n_2\}
\]

and by \(i_{ab}\), with \((a, b) \in D\), the values of the pixels of \(i\).

Given a planar curve \(x\), we denote by \(\hat{x}\) a suitable discrete approximation of the topological interior of \(x\). The topological interior is considered here because it is the region occupied by the object whose contour is \(x\) (see Figure 2.2).

We use integral notation to indicate summations on the image pixels, i.e. if \(A \subseteq D\) and \(f \in \mathbb{R}^{n_1 \times n_2}\),

\[
\int_A f = \sum_{(i,j) \in A} f_{ij}
\]

and the mean of \(f\) in \(A\) is

\[
\bar{f}_A = \frac{1}{|A|} \int_A f
\]

where \(|A|\) is the cardinality of \(A\).

The Chan-Vese energy can now be defined as

\[
E(i, x) = \int_{\hat{x}} (i - \text{avg}(\hat{x}))^2 + \int_{D \setminus \hat{x}} (i - \text{avg}(D \setminus \hat{x}))^2
\]
Figure 2.2: The interior of a curve. Interiors are striped.
where
\[ \text{avg}(\vec{x}) = \frac{1}{\vec{x}} \int i \quad \text{and} \quad \text{avg}(D \setminus \vec{x}) = \frac{1}{D \setminus \vec{x}} \int i. \]

Roughly, this energy gives a measure of how much uniform are the region inside and outside the shape.
Chapter 3

Gaussian measures

In this section we make an introduction to probability measures in Hilbert spaces. We are mostly interested in Gaussian measures. The main reference for this kind of results is the book [3], which treats the subject with great generality, considering the case of locally convex spaces.

We restrict our presentation to Hilbert spaces. Gaussian measures in Hilbert spaces can be treated writing everything in coordinates, and some books follow this approach, for example [9]. We think the abstract setting of [3] is more clear, and in this introduction we follow the book [3].

3.1 Finite dimensional Gaussian measures

Definition 3.1.1. A measure on $\mathbb{R}$ is called Gaussian if it is the Dirac measure $\delta_m$ at a point $m$ or if it 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 $m \in \mathbb{R}$ and $\sigma > 0$.

If we put $\sigma = 0$ for any Dirac measure, the parameters $m$ and $\sigma^2$ are the mean and variance of $\mu$, namely

$$m = \int_{\mathbb{R}} x \, d\mu(x) \quad \sigma^2 = \int_{\mathbb{R}} (x-m)^2 \, d\mu(x).$$

A Gaussian measure on $\mathbb{R}$ is called non degenerate if it is not a Dirac measure.

A real random variable $X$ on a probability space $(\Omega, \mathcal{F}, \mu)$ is called Gaussian if $X_{\sharp \mu}$ is a Gaussian measure on $\mathbb{R}$.

Definition 3.1.2. A measure $\mu$ on $\mathbb{R}^n$ is called Gaussian if for all linear functional $f$, the induced measure $f_{\sharp \mu}$ is Gaussian.
The measure $\mu$ is called non degenerate if for all linear functional $f$, the measure $f \mu$ is non degenerate.

We recall now some properties of Gaussian random variables and finite dimensional Gaussian measures. We do not provide many proofs, referring the reader to Sections 9.4 and 9.5 of [11] for a more detailed discussion.

**Proposition 3.1.3.** The Fourier transform of a Gaussian measure $\gamma$ with mean $m$ and variance $\sigma$ on $\mathbb{R}$ is

$$\hat{\gamma}(\xi) = \exp \left( i m \xi - \frac{1}{2} \sigma^2 \xi^2 \right).$$

**Corollary 3.1.4.** Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of centered Gaussian random variables on a measure space $(\Omega, F, \mu)$. Suppose they converge almost surely to a random variable $X$. Then $X$ is Gaussian and centered.

*Proof.* Let $\sigma_n^2$ be the variance of $X_n$. By the dominate convergence theorem

$$\hat{X_n \mu}(\xi) = \int_{\mathbb{R}} e^{i \xi x} \ d\mu = \lim_{n \to \infty} \int_{\mathbb{R}} e^{i \xi X_n} \ d\mu = \lim_{n \to \infty} e^{-\frac{1}{2} \sigma_n^2 \xi^2},$$

in particular the rightmost limit exists. This implies that the limit

$$\sigma = \lim_{n \to \infty} \sigma_n^2$$

exists. The Fourier transform of $X$ is

$$\hat{X \mu}(\xi) = \lim_{n \to \infty} e^{-\frac{1}{2} \sigma_n^2 \xi^2} = e^{-\frac{1}{2} \sigma^2 \xi^2}$$

and then by Proposition 3.1.3 and injectivity of the Fourier transform, $X$ is Gaussian and centered. \qed

An other result about the convergence of Gaussian random variables is the following, see [3, Theorem 1.1.4].

**Proposition 3.1.5.** Let $\{X_n\}_{n \in \mathbb{N}}$ a sequence of independent centered Gaussian random variables of variances $\sigma_n^2$ on a probability space $(\Omega, F, \mu)$. Then the following conditions are equivalent:

1. the series $\sum_{n=1}^{\infty} X_n$ converges almost everywhere;
2. there exists a subsequence of partial sums $\sum_{i=1}^{n_k} X_i$ that converges almost everywhere as $k \to \infty$;
3. the series $\sum_{n=1}^{\infty} X_n$ converges in probability;
4. the series $\sum_{n=1}^{\infty} X_n$ converges in $L^2(\mu)$;
5. the series $\sum_{n=1}^{\infty} \sigma_n^2$ is finite.

Gaussian measures in $\mathbb{R}^n$ are characterized by their Fourier transform.
Proposition 3.1.6. A measure $\mu$ on $\mathbb{R}^n$ is Gaussian if and only if its Fourier transform is equal to

$$\hat{\mu}(\xi) = \exp \left( i \langle m, \xi \rangle - \frac{1}{2} \langle K \xi, \xi \rangle \right)$$

for some $m \in \mathbb{R}^n$ and some symmetric nonnegative matrix $K \in \mathbb{R}^{n \times n}$.

If $\mu$ is a Gaussian measure, the vector $m$ and matrix $K$ given by the Proposition above are called the mean and covariance matrix of $\mu$. They are related to the mean and variance of image measures of $\mu$ under linear maps. Indeed if $x \in \mathbb{R}^n$ and $x^* : \mathbb{R}^n \to \mathbb{R}$ is the linear random variable

$$x^*(y) = \langle x, y \rangle$$

given by the standard scalar product in $\mathbb{R}^n$, the measure $x^* \mu$ has mean $\langle m, x \rangle$ and variance $\langle K x, x \rangle$. Moreover, the covariance of two linear random variables $x_1^*$ and $x_2^*$ is $\langle K x_1, x_2 \rangle$.

By choosing an orthonormal base in $\mathbb{R}^n$ with respect to which $K$ is diagonal, $\mu$ can be decomposed as a product of one-dimensional Gaussian measures. This proves the following Corollary.

**Corollary 3.1.7.** Let $\mu$ be a Gaussian measure on $\mathbb{R}^n$ and $m$, $K$ as in Proposition 3.1.6. Then

1. the support of $\mu$ is the orthogonal to $\text{Ker}(K)$, in particular it is a subspace and it coincides with $\mathbb{R}^n$ if and only if $K$ is invertible;
2. $\mu$ has density with respect to the Lebesgue measure if and only if $K$ is invertible;
3. $\mu$ is non degenerate if and only if $K$ is invertible.

**Proposition 3.1.8.** Let $X_1$ and $X_2$ be real Gaussian random variable. Then any linear combination $\alpha X_1 + \beta X_2$, with $\alpha$, $\beta$ in $\mathbb{R}$, is Gaussian as well.

Given a probability space $(\Omega, \mathcal{F}, \mu)$, we often regard a real Gaussian random variable $X$ as an element of $L^2(\mu)$. This make sense, indeed

$$\int_{\Omega} f^2 \, d\mu = \int_{\mathbb{R}} x^2 \, df \mu(x) < +\infty$$

because Gaussian measures on $\mathbb{R}$ has second moment. In the case where $\Omega = \mathbb{R}^n$ and $\mu$ is a Gaussian measure, all linear functional could be regarded as elements of $L^2(\mu)$.

**Proposition 3.1.9.** Let $X_1, \ldots, X_n$ be centered Gaussian real random variables. Suppose $X_1$ is orthogonal to $X_2, \ldots, X_n$ in $L^2(\gamma)$. Then $X_1$ is independent of the $\sigma$-algebra generated by $X_2, \ldots, X_n$. 

CHAPTER 3. GAUSSIAN MEASURES

3.2 Gaussian measures in Hilbert spaces

In this section we define Gaussian measures in Hilbert spaces and outline some of their properties, giving also some proof. The reference for other proofs remains [3].

First of all we should say why we generalize Gaussian measures to Hilbert spaces and not, for example, the Lebesgue measure. The fact is that there is no analogue of Lebesgue measure on infinite dimensional spaces, since by a known lemma says that translation invariant measures are not so interesting.

Lemma 3.2.1. Let $H$ be a separable infinite dimensional Hilbert space and $\mu$ a translation invariant, possibly not finite, measure on $H$. Then either $\mu$ is identically 0 or it is $+\infty$ on all open sets.

Proof. Let $\mu$ be a measure as in the statement and suppose that there exists an open set of finite measure. Then there exists also an open ball $B_0$ of finite measure. Call $3r$ its radius. Being $H$ infinite dimensional, there exists a sequence of balls $\{B_n\}_{n\in\mathbb{N}}$ of radius $r$, contained in $B_0$ and disjoint. By $\sigma$-additivity of $\mu$

$$\sum_{n=1}^{\infty} \mu(B_n) \leq \mu(B_0) < +\infty,$$

but by translation invariance all the $B_n$ have the same measure and so it should hold

$$\mu(B) = 0 \quad \text{for all balls } B \text{ of radius } r.$$

By separability, $H$ is covered by a countable union of balls of radius $r$ and so $\mu$ is identically 0. \qed

Gaussian measures could instead be generalized to Hilbert spaces and retain some nice properties.

In the following let $H$ be a separable Hilbert space. The definition of Gaussian measure and some properties are also true in more general spaces, not necessarily separable, but we restrict ourselves to these.

When talking about Gaussian measures, it could be confusing to identify $H$ and its dual. For this reason, we keep them distinct, denoting by $H^*$ the dual space and by $x^*$ the linear functional associated to $x \in H$, i.e. for every $x \in H$, $x^*$ is a function defined by

$$x^*(y) = \langle x, y \rangle \quad \text{for all } y \in H.$$

Definition 3.2.2. Let $H$ be a separable Hilbert space. A measure $\gamma$ on $H$ is said to be Gaussian if for all $x^* \in H^*$ the image measure $x^* \gamma$ on $\mathbb{R}$ is Gaussian.

Note that all Gaussian measures are probability measures. Indeed, given $x^* \in H^*$, it holds $\gamma(H) = x^*_\mathbb{R} \gamma(\mathbb{R})$ and all Gaussian measures on $\mathbb{R}$ are probability measures.

As in the finite dimensional case, all continuous linear functional in $H^*$ can be regarded as real random variables and they belong to $L^2(\gamma)$. A measure
3.2. GAUSSIAN MEASURES IN HILBERT SPACES

on $H$ is Gaussian if and only if all continuous linear functional are Gaussian random variables.

To do an example of Gaussian measure on a Hilbert space, we need a lemma about $\sigma$-algebras. This Lemma can be stated in a more general form, see for example Theorem A.3.7 in [3].

Lemma 3.2.3. Let $H$ be a separable Hilbert space, $\{e_n\}$ an orthonormal basis of $H$ and $e^*_n$ the coordinate functions relative to that base.

Then the Borel $\sigma$-algebra $\mathcal{B}(H)$ is generated by the family $\{e^*_n\}_{n \in \mathbb{N}}$ of functions $H \to \mathbb{R}$.

Proof. Let $\mathcal{E}$ be the $\sigma$-algebra generated by $\{e^*_n\}_{n \in \mathbb{N}}$. Finite linear combinations of the $e^*_n$ are measurable with respect to $\mathcal{E}$. It could be verified that translations by elements in $\text{Span}(e_n)$ are measurable too.

The inclusion $\mathcal{E} \subseteq \mathcal{B}(H)$ is true.

Let $\{x_m\}_{m \in \mathbb{N}} \subseteq \text{Span}(e_n)$ a countable set dense in $H$. Since $x_m$ are dense, $\mathcal{B}(H)$ is generated by balls centered in $\{x_m\}$ and also by closed balls with those centers. Translations by $x_m$ are $\mathcal{E}$ measurable and so to prove the inclusion $\mathcal{B}(H) \subseteq \mathcal{E}$ it sufficient to see that all closed balls centered in the origin belong to $\mathcal{E}$.

Let $B$ be a closed ball of radius $r$ with center the origin. For every element $x_m \notin B$, consider the half-space $L_m = \{x \in H \mid \langle x_m, x \rangle \leq r |x_m| \}$. Then

$$L_m \supseteq B \quad \text{and} \quad x_m \notin L_m.$$

The intersection

$$L = \bigcap_{m \text{ t.c. } x_m \notin B} L_m$$

is equal to $B$. Indeed, if $x \notin B$, by density there exists $x_m$ such that

$$|x - x_m| < \frac{|x| - r}{2}.$$

With elementary calculations, it can be seen that $x \notin L_m \supseteq B$.

Then $B$ is a countable intersection of $L_m$, which are $\mathcal{E}$ measurable, and then it is $\mathcal{E}$ measurable too.

Example 3.2.4. Let $\mathbb{R}^\infty$ be a product of real lines and $\mathcal{F}$ the product $\sigma$-algebra of infinitely many copies of $\mathcal{B}(\mathbb{R})$. We denote by $(x_n)_{n \in \mathbb{N}}$ the elements of $\mathbb{R}^\infty$. Let also $\{\sigma_n\}_{n \in \mathbb{N}}$ be a sequence such that

$$\sum_{n \in \mathbb{N}} \sigma_n^2 < +\infty$$

and $\gamma_n$ centered Gaussian measures on $\mathbb{R}$ with variance $\sigma_n^2$.

Define the probability measure $\tilde{\gamma}$ on $(\mathbb{R}^\infty, \mathcal{F})$ as the product of $\gamma_n$. Since

$$\int_{\mathbb{R}^\infty} \sum_{n \in \mathbb{N}} x_n^2 \, d\tilde{\gamma} = \sum_{n \in \mathbb{N}} \sigma_n^2 < +\infty,$$
the function $\sum x_n^2$ is finite almost everywhere and then the measure $\tilde{\gamma}$ is concentrated on the set $\ell^2 = \{(x_n)_{n\in\mathbb{N}} \mid \sum x_n^2 < +\infty\}$, which is measurable.

The set $\ell^2$ is a Hilbert space with the usual scalar product

$$\langle (x_n), (y_n) \rangle_{\ell^2} = \sum_{n\in\mathbb{N}} x_n y_n.$$  

Subsets of $\ell^2$ can be seen as subsets of $\mathbb{R}^\infty$ and so the measure $\tilde{\gamma}$ could be evaluated on the restriction of $\mathcal{F}$ to $\ell^2$, namely subsets of $\ell^2$ that belongs to $\mathcal{F}$.

By Lemma 3.2.3, the Borel $\sigma$-algebra $\mathcal{B}(\ell^2)$ is generated by the coordinates functions $e^*_n: (x_m)_{m\in\mathbb{N}} \mapsto x_n$, which are measurable with respect to $\mathcal{F}$ and so $\mathcal{B}(\ell^2)$ is contained in the restriction of $\mathcal{F}$ to $\ell^2$.

We can then consider $\tilde{\gamma}$ as a probability measure on $\ell^2$. This is a Gaussian measure. Indeed it follows from the definition that $e^*_n$ are Gaussian random variables and by Proposition 3.1.8, linear combinations of $e^*_n$ are Gaussian as well. Every linear functional on $\ell^2$ is pointwise limit of such linear combinations and so by Corollary 3.1.4 it is Gaussian too.

Gaussian measures are characterized by their Fourier transform. We recall that the Fourier transform of a measure $\mu$ is the function $\hat{\mu}: H^* \to \mathbb{R}$

$$\hat{\mu}(x^*) = \int e^{ix^t(y)} \, d\mu(y).$$

**Proposition 3.2.5.** A measure $\mu$ on $H$ is Gaussian if and only if its Fourier transform is

$$\hat{\mu}(x^*) = \exp \left( iL(x^*) - \frac{1}{2} B(x^*, x^*) \right)$$

for some linear function $L$ on $H^*$ and some symmetric nonnegative bilinear function $B$ on $H^*$.

Moreover

$$L(x^*) = \int x^* \, d\mu$$  \hspace{1cm} (3.1)$$

and

$$B(x^*_1, x^*_2) = \int (x^*_1 - L(x^*_1))(x^*_2 - L(x^*_2)) \, d\mu.$$  \hspace{1cm} (3.2)$$

**Proof.** Let $\mu$ be a measure with such a Fourier transform. For every element $x^* \in H^*$ the Fourier transform of the measure $x^* \mu$ is

$$\hat{x^* \mu}(\xi) = \int_{\mathbb{R}} e^{i\xi t} \, dx^* t \mu(t) = \int_{H} e^{i\xi x^t(y)} \, d\mu(y) = \hat{\mu}(\xi x^*) =$$

$$= \exp \left( iL(\xi x^*) - \frac{1}{2} B(\xi x^*, \xi x^*) \right) = \exp \left( i\xi L(x^*) - \frac{1}{2} \xi^2 B(x^*, x^*) \right).$$

By Proposition 3.1.3, this is also the Fourier transform of the Gaussian measure with mean $L(x^*)$ and variance $B(x^*, x^*)$. Since the Fourier transform is injective, $x^* \mu$ is Gaussian. By the definition, $\mu$ is Gaussian as well.
3.2. GAUSSIAN MEASURES IN HILBERT SPACES

Conversely, if $\mu$ is a Gaussian measure, its Fourier transform can be computed using Proposition 3.1.3,

$$\hat{\mu}(x^*) = \int_\mathbb{H} e^{ix^*(y)} \, \text{d}\mu(y) = \int_\mathbb{R} e^{it} \, \text{d}x^* \mu(t) = \hat{x^*} \mu(1) = \exp \left( iL(x^*) - \frac{1}{2} B(x^*, x^*) \right),$$

where $L$ and $B$ are defined by Equation 3.1 and 3.2. Clearly $L$ is linear and $B$ is symmetric, nonnegative and bilinear, so we have proven the first part of the proposition, Equation 3.1, and Equation 3.2 in the special case where $x_1^* = x_2^*$.

To prove Equation 3.2 it is sufficient to note that left and right hand side are symmetric bilinear forms that induce the same norm and this is sufficient to say that they are equal.

The operators $L$ and $B$ have some important continuity property, as stated in next theorem.

We recall that if $K$ is a symmetric nonnegative compact operator on $H$, by a well know theorem on the diagonalization of symmetric compact operators (see e.g. [5, Teorema VI.11]), there exists an orthonormal basis $\{e_n\}_{n \in \mathbb{N}}$ of $H$ made of eigenvectors for $K$. Let $\sigma_n^2$ be the correspondent eigenvalues. We say that $K$ is “trace-class” if the sum of the eigenvalues converges,

$$\sum_{n=1}^{\infty} \sigma_n^2 < \infty.$$  

The term trace-class operator is usually referred to a much more general class of operators, see e.g. [14], but here we just need to say in short that the sum of the eigenvalues of a symmetric nonnegative compact operator converges.

**Theorem 3.2.6.** Let $\gamma$ be a Gaussian measure on a separable Hilbert space $H$, $L$ and $B$ defined as in Proposition 3.2.5. Then there exist a vector $m_\gamma \in H$ and a symmetric nonnegative compact “trace-class” operator $K$ such that

$$L(x^*) = x^*(m_\gamma) = \langle m_\gamma, x \rangle \quad (3.3)$$
$$B(x_1^*, x_2^*) = \langle Kx_1, x_2 \rangle. \quad (3.4)$$

Conversely, for any such $m$ and $K$ there exists a Gaussian measure $\gamma$ on $H$ with Fourier transform

$$\hat{\gamma}(x^*) = \exp \left( i \langle m, x \rangle - \frac{1}{2} \langle Kx, x \rangle \right)$$

**Proof.** We first show that $L$ and $B$ are continuous. By dominate convergence theorem, the Fourier transform of $\gamma$ is continuous $H^* \to \mathbb{R}$ with respect to the weak topology on $H^*$ and then also with respect to the strong topology. The Fourier transform is

$$\hat{\gamma}(x^*) = \exp \left( iL(x^*) - \frac{1}{2} B(x^*, x^*) \right).$$
and so \( x^* \mapsto B(x^*, x^*) \) and \( x^* \mapsto L(x^*) \) are continuous as well.

The function \( L \) is a linear continuous functional on \( H^* \), and then there exists \( m_\gamma \) satisfying Equation (3.3).

Since \( B \) is a bilinear operator, from the continuity of \( x^* \mapsto B(x^*, x^*) \) follows also the continuity of \( B: H \times H \to \mathbb{R} \), and so there exists a linear continuous operator \( K \) that satisfies Equation (3.4). Symmetry and nonnegativity of \( K \) follow from symmetry and nonnegativity of \( B \).

To see that \( K \) is compact, consider a bounded sequence \( \{x_n\}_{n \in \mathbb{N}} \subseteq H \). By weak compactness, there exists a weakly convergent subsequence \( \{x_{n_k}\} \) and translating we can suppose \( x_{n_k} \rightharpoonup 0 \). By the continuity of the Fourier transform outlined before,

\[
\langle Kx_{n_k}, x_{n_k} \rangle \to 0.
\]

Since \( K \) is continuous and nonnegative, there exists its square root \( \sqrt{K} \) by a well known functional analysis theorem (see Theorem 12.33 of [30]) and \( \sqrt{K} \) is a continuous symmetric nonnegative operator. The equation above can then be written as

\[
\left\langle \sqrt{K}x_{n_k}, \sqrt{K}x_{n_k} \right\rangle \to 0,
\]

which means \( \sqrt{K}x_{n_k} \to 0 \) in the strong topology of \( H \), and as a consequence \( Kx_{n_k} \to 0 \) as well.

It remains to show that \( K \) is trace class. Translating the measure by \(-m_\gamma\), it is possible to reduce to the case when \( \gamma \) is centered. We prove that centered Gaussian measures have second moment, i.e.

\[
\int_H |x|^2 \, d\gamma < \infty.
\]

Let \( \{e_n\}_{n \in \mathbb{N}} \) an orthonormal basis of \( H \) made of eigenvectors for \( K \). The functionals \( e_n^* \) are orthogonal Gaussian random variables and so, by Proposition 3.1.9, they are independent.

Since the norm can be written as

\[
|x|^2 = \sum_{n=1}^{\infty} e_n^*(x)^2
\]

and since \( e_n^* \) are independent, the series \( \sum_{n=1}^{\infty} e_n^* \) restricted to a bounded set converges in \( L^2(\gamma) \).

With a diagonal argument, it is possible to extract a subsequence \( n_k \) such that the partial sums \( \sum_{n=1}^{n_k} e_n^* \) converge almost everywhere as \( k \to \infty \). By Proposition 3.1.5 this implies

\[
\sum_{n=1}^{\infty} \int_H e_n^* \, d\gamma < \infty
\]

and then

\[
\int_H |x|^2 \, d\gamma(x) = \int_H \sum_{n=1}^{\infty} e_n^*(x)^2 \, d\gamma(x) < \infty.
\]
Now the fact that $K$ is trace class is straight forward, indeed the eigenvalue $\sigma_n^2$ relative to $e_n$ is

$$\sigma_n^2 = \langle Ke_n, e_n\rangle = B(e_n^*, e_n^*) = \int_H e_n^2 \, d\gamma$$

by definition of $B$ and we have just proven that the sum of these terms converges.

To see the converse, we have to show that that there exists a Gaussian measure whose $L$ and $B$ functions satisfy Equations (3.3) and (3.4). Since $K$ is symmetric and compact, there exists an orthonormal basis $\{e_n\}_{n \in \mathbb{N}}$ of $H$ made of eigenvectors for $K$. Let $\sigma_n^2$ be the correspondent eigenvalues. The construction of Example 3.2.4 leads to a centered Gaussian measure $\gamma$ such that the coordinates functions are independent and have covariance $\sigma_n^2$. For such measure, the function $B$, which is continuous as we have proven above, is

$$B(x^*, y^*) = B \left( \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n^*, \sum_{m=1}^{\infty} \langle y, e_m \rangle e_m^* \right) = \sum_{n=1}^{\infty} \langle x, e_n \rangle \langle y, e_n \rangle \sigma_n^2 = \sum_{i=1}^{\infty} \langle x, e_n \rangle \langle y, e_n \rangle \langle Ke_n, e_n \rangle = \langle Kx, y \rangle$$

and so Equation (3.4) is satisfied. To get a measure that satisfies also Equation (3.3) is it sufficient to translate $\gamma$ by $m_\gamma$. \qed

The vector $m_\gamma$ is called the mean of $\gamma$ and $\gamma$ is said to be centered if $m_\gamma$ is the origin. The translate of $\gamma$ by $-m_\gamma$ is still a Gaussian measure and it is centered, so we can always suppose a Gaussian measure to be centered up to a translation.

Theorem 3.2.6 has many interesting corollaries. The first corollary below is an intermediate step of the proof.

Corollary 3.2.7. Every Gaussian measure $\gamma$ on a Hilbert space $H$ has second moment, namely

$$\int_H |x - m_\gamma|^2 \, d\gamma(x) < +\infty.$$ 

In $\mathbb{R}^n$, chosen a system of coordinates, there is a “standard” Gaussian measure, the one with center in the origin and covariance matrix the identity, which is rotationally symmetric. Theorem 3.2.6 says that in infinite dimensional spaces, Gaussian measures are not symmetric, and the variance of the coordinates functions should go to zero quite fast.

Corollary 3.2.8. The function $\exp \left( -\frac{1}{2} |x|^2 \right)$ is the Fourier transform of no measure on an infinite dimensional separable Hilbert space.
Proof. Suppose by contradiction that there exists a measure $\mu$ on the Hilbert space $H$. By Proposition 3.2.5 $\mu$ is Gaussian measure, but then by Theorem 3.2.6 the identity should be a trace class operator and this could be true only if $\dim H < \infty$. \qed

Corollary 3.2.9. Let $\gamma$ be a Gaussian measure on a Hilbert space $H$. Then there exists an orthonormal basis of $H$ such that the coordinate functions are independent.

In some sense, $\gamma$ can be seen a product of one dimensional Gaussian measures.

Proof. The operator $K$ is compact and symmetric, there is an orthonormal basis $\{e_n\}_{n \in \mathbb{N}}$ of eigenvectors for $K$. The coordinate functions $e^*_n$ are orthogonal in $L^2(\gamma)$, indeed if $n \neq m$

$$\langle e^*_n, e^*_m \rangle_{L^2(\gamma)} = \langle Ke_n, e_m \rangle = 0,$$

but, since $e^*_n$ are Gaussian random variables, by Proposition 3.1.9 this implies that they are also independent. \qed

In general, finite measures on Hilbert spaces have another couple of useful properties, which of course are true also for Gaussian measures.

Lemma 3.2.10. Let $H$ a separable Hilbert space, $\{e_n\}_{n \in \mathbb{N}}$ an orthonormal basis and $P_n : H \to H$ the projection on the subspace generate by $e_1, \ldots, e_n$. Let also $\mu$ be a finite measure and consider the image measures $P_n\# \mu$. Then $P_n\# \mu \rightharpoonup C_b \mu$, i.e. for every $\varphi : H \to \mathbb{R}$ continuous and bounded

$$\int \varphi \ dP_n\# \mu \to \int \varphi \ d\mu.$$

Proof. Let $\varphi \in C_b$ be a continuous bounded function. Since $\{e_n\}_{n \in \mathbb{N}}$ is a basis of $H$, for every $x \in H$,

$$P_n(x) \to x$$

and then also $\varphi \circ P_n \to \varphi$ pointwise. It follows that

$$\int \varphi \ dP_n\# \mu = \int \varphi \circ P_n \ d\mu \to \int \varphi \ d\mu$$

by dominate convergence, because $\varphi$ is bounded and $\mu$ is finite. \qed

The second property says that also in separable Hilbert spaces measures are characterized by their Fourier transform.

Lemma 3.2.11. Let $\mu$ and $\nu$ two measures on a separable Hilbert space and suppose their Fourier transforms coincide $\hat{\mu} = \hat{\nu}$.

Then $\mu = \nu$. 
3.2. GAUSSIAN MEASURES IN HILBERT SPACES

Proof. By Lemma 3.2.3, $H^*$ generates the Borel $\sigma$-algebra $B(H)$. This means that it is sufficient to prove

$$\mu(B) = \nu(B)$$

for all $B$ in the family

$$\mathcal{F} = \{(x_1^*, \ldots, x_n^*)^{-1}(B') \mid n \in \mathbb{N}, x_1^* \ldots x_n^* \in H^*, B' \in B(\mathbb{R})\}.$$

Indeed, $\mathcal{F}$ is closed under finite intersection, the family of sets on which $\mu$ and $\nu$ coincide is a Dynkin system and the smaller $\sigma$-algebra that contains $\mathcal{F}$ coincide with the one generated by $H^*$.

Equivalently we can prove

$$(x_1^*, \ldots, x_n^*)_{\sharp} \mu = (x_1^*, \ldots, x_n^*)_{\sharp} \nu$$

for all $x_1^* \ldots x_n^* \in H^*$ and $n \in \mathbb{N}$. This follows from the injectivity of the Fourier transform for measures on $\mathbb{R}^n$. Indeed, letting $f = (x_1^*, \ldots, x_n^*)$,

$$\hat{f}_{\mu}(\xi) = \hat{\mu}(\xi_1 x_1^* + \cdots + \xi_n x_n^*) = \hat{\nu}(\xi_1 x_1^* + \cdots + \xi_n x_n^*) = \hat{f}_{\nu}(\xi)$$

for all $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$. $\square$

The last remark is that all measures on a separable Hilbert space are Radon. This follows from the Ulam’s lemma, which is true in general for measures in complete metric spaces, see also Theorem 7.1.4 in [11].

**Proposition 3.2.12** (Ulam’s lemma). Let $(\Omega, d)$ be a complete separable metric space and $\mu$ a finite measure on $\Omega$. Then $\mu$ is Radon.

Proof. We first prove that for every $\varepsilon$ there exists a compact set $C_{\varepsilon}$ such that $\mu(\Omega \setminus C_{\varepsilon}) < \varepsilon$. Let $\{x_n\}_{n \in \mathbb{N}}$ be a dense set in $\Omega$ and denote by $B_m(x_n)$ the ball of radius $\frac{1}{m}$ and center $x_n$. The union of that balls at $m$ fixed covers $\Omega$ by density

$$\bigcup_{n=1}^{+\infty} B_m(x_n) = \Omega$$

and so for every $m$ there exists $N_m$ such that, called $U_m$ the set

$$U_m = \bigcup_{n=1}^{N_m} B_m(x_n),$$

it holds

$$\mu(\Omega \setminus U_m) < 2^{-m} \varepsilon.$$ 

Now let

$$U = \bigcap_{m=1}^{+\infty} U_m.$$
and \( C_\varepsilon \) be the closure of \( U \). The set \( U \) is totally bounded by construction, and so, since \( \Omega \) is complete, its closure \( C_\varepsilon \) is compact. We can also estimate the measure of its complement as

\[
\mu(\Omega \setminus C_\varepsilon) \leq \mu(\Omega \setminus U) = \mu \left( \bigcup_{m=1}^{\infty} \Omega \setminus U_m \right) < \sum_{m=1}^{\infty} 2^{-m} \varepsilon = \varepsilon
\]

and so \( C_\varepsilon \) is the set we were looking for.

It is a well known result, see for example Theorem 7.1.3 in [11], that if \( \mu \) is a finite measure on a metric space, then for every Borel set \( B \) and every \( \varepsilon \) there exists a closed set \( C \subseteq B \) such that \( \mu(B \setminus C) < \varepsilon \). Intersecting \( C \) and \( C_\varepsilon \) we get a compact set contained in \( B \) such that

\[
\mu(B \setminus (C \cap C_\varepsilon)) \leq \mu(B \setminus C) + \mu(\Omega \setminus C_\varepsilon) < 2\varepsilon
\]

and we are done.

The fact that Gaussian measures are Radon implies that they are concentrated on a countable union of compact sets. This could be a bit surprising, because compact sets are quite “small” in Hilbert spaces. As we will see in the following, it is possible to show other “small” sets on which a Gaussian measure is concentrated and this phenomenon is a big difference between the finite and infinite dimensional case.

### 3.3 The Cameron-Martin space

Let \( H \) be a separable Hilbert space and denote its scalar product by \( \langle \cdot, \cdot \rangle_H \) and its norm by \( |\cdot|_H \). Let also \( \gamma \) be a Gaussian measure on \( H \). As in the previous section, we denote by \( H^* \) the dual of \( H \) and for each \( x \in H \), \( x^*: H \to \mathbb{R} \) is the continuous linear functional defined by

\[
x^*(y) = \langle x, y \rangle.
\]

We denote the dual norm on \( H^* \) by \( |\cdot|_H \) as well.

By Theorem 3.2.6 there exists the mean of \( \gamma \) and it is denoted by \( m_\gamma \), namely for each \( x^* \in H^* \)

\[
\int_H x^* \, d\gamma = x^*(m_\gamma).
\]

We now define also the covariance of \( \gamma \), which is a scalar product on \( H^* \) defined by

\[
\langle x_1^*, x_2^* \rangle_\gamma = \int_H (x_1^* - x_1^*(m_\gamma))(x_2^* - x_2^*(m_\gamma)) \, d\gamma.
\]

Note that if \( \gamma \) is centered the covariance \( \langle \cdot, \cdot \rangle_\gamma \) is the scalar product in \( L^2(\gamma) \).

The covariance of \( \gamma \) induces on \( H^* \) a norm

\[
|x^*|_\gamma^2 = \int_H (x^* - x^*(m_\gamma))^2 \, d\gamma,
\]

which in general is different from the dual norm induced by \( H \).
3.3. THE CAMERON-MARTIN SPACE

Definition 3.3.1. The covariance operator $R'_\gamma : H^* \to H$ is defined by the equation

$$x^*_1 (R'_\gamma x^*_2) = \langle x^*_1, x^*_2 \rangle_\gamma .$$

This definition is well posed thanks to Theorem 3.2.6, which states that the covariance is continuous with respect to the dual norm.

Definition 3.3.2. We denote by $H^*_\gamma$ the closure in $L^2(\gamma)$ of the set

$$\{ x^* - x^*(m_\gamma) \mid x^* \in H^* \} .$$

Note that, unless $\gamma$ is centered, $H^* \nsubseteq H^*_\gamma$, but $H^*_\gamma$ contains translations of the elements in $H^*$.

Lemma 3.3.3. Every $f \in H^*_\gamma$ is a centered Gaussian random variable.

Proof. Let $f \in H^*_\gamma$ and $x^*_n$ a sequence in $H^*$ such that $f_n = x^*_n - x^*_n(m_\gamma)$ converges to $f$ in $L^2(\gamma)$. Possibly extracting a subsequence, we can suppose that the convergence is almost sure. By Corollary 3.1.4, the limit random variable $f$ is Gaussian and centered.

Lemma 3.3.4. For every $f \in H^*_\gamma$ the linear functional on $H^*$

$$x^* \mapsto \int_H [x^* - x^*(m_\gamma)] f \, d\gamma = \langle f, x^* - x^*(m_\gamma) \rangle_{L^2(\gamma)}$$

is continuous with respect to the $|\cdot|_H$ norm and so, the covariance operator admits an “extension” to $H^*_\gamma$ defined by

$$x^*(R_\gamma f) = \int_H [x^* - x^*(m_\gamma)] f \, d\gamma .$$

Beware that, unless $\gamma$ is centered, this not at all an extension, because in general $H^* \nsubseteq H^*_\gamma$. It happens instead that

$$R'_\gamma(x^*) = R_\gamma(x^* - x^*(m_\gamma)) ,$$

$R'_\gamma$ is defined on $H^*$ and $R_\gamma$ is defined on $H^*_\gamma$. Sometimes, see e.g. [3], these two operators are both denoted by $R_\gamma$.

Proof. The functional defined in Equation 3.5 is clearly continuous with respect to the $|\cdot|_\gamma$ norm on $H^*$. But this norm is bounded by the $|\cdot|_H$ norm, indeed

$$|x^*|_\gamma^2 = \int_H [x^*(y - m_\gamma)]^2 d\gamma(y) \leq |x^*|_H^2 \int_H |y - m_\gamma|_H^2 d\gamma(y)$$

and

$$\int_H |y - m_\gamma|_H^2 d\gamma < \infty$$

is a finite constant by Corollary 3.2.7.

\qed
By duality $H$ could be seen as a set of linear functional on $H^*$. If we consider on $H^*$ the norm $|·|_γ$, not all the elements of $H$ are continuous with respect to this norm. The continuous ones are the Cameron-Martin space.

**Definition 3.3.5.** The Cameron-Martin space, denoted by $H_γ$, is a subspace of $H$ defined as

$$H_γ = \left\{ x \in H \mid \sup_{y^* \in H^*, |y^*|_γ \leq 1} y^*(x) < +\infty \right\}.$$ 

On the Cameron-Martin space is defined the norm

$$|x|_{H_γ} = \sup_{y^* \in H^*, |y^*|_γ \leq 1} y^*(x).$$

With a little abuse of notation, we define $|·|_{H_γ}$ on the whole $H$, letting

$$|x|_{H_γ} = +\infty \text{ for every } x \notin H_γ.$$

The Cameron-Martin space is closely related to the structure of the Gaussian measure $γ$ and we now outline some of its properties.

**Proposition 3.3.6.** The Cameron-Martin space is the image of $H_γ^*$ through the operator $R_γ$,

$$H_γ = R_γ(H_γ^*).$$

Moreover, for every $f \in H_γ^*$,

$$|R_γ f|_{H_γ} = |f|_{L^2(γ)}.$$

**Proof.** Let $f$ be an element of $H_γ^*$. Then, for every $y^* \in H^*$ such that $|y^*|_γ \leq 1$, by definition of $R_γ$

$$y^*(R_γ f) = \langle y^* - y^*(m_γ), f \rangle_{L^2(γ)} \leq |y^*|_γ |f|_{L^2(γ)} \leq |f|_{L^2(γ)} < +\infty$$

and so $R_γ f \in H_γ$.

Conversely, if $x \in H_γ$ the linear functional $x^{**}: H^* \to \mathbb{R}$

$$x^{**}: y^* \mapsto y^*(x)$$

is continuous with respect to the $|·|_γ$ norm on $H^*$. This means that a correspondent functional $x_γ^{**}$ can be defined $H_γ^* \to \mathbb{R}$, letting

$$x_γ^{**}(y^* - y^*(m_γ)) = y^*(x)$$

for every $y^* \in H^*$ and extending this function to $H_γ^*$ by continuity with respect to the $L^2(γ)$ norm. Since $H_γ^*$ is an Hilbert space with the $L^2(γ)$ norm, by Riesz theorem, the functional $x_γ^{**}$ is represented by some $f \in H_γ^*$. This $f$ is such that for every $y \in H^*$

$$y^*(x) = x_γ^{**}(y^* - y^*(m_γ)) = \langle f, y^* - y^*(m_γ) \rangle_{L^2(γ)}.$$
3.3. THE CAMERON-MARTIN SPACE

and this means that \( x = R_\gamma(f) \) by definition of \( R_\gamma \).

Riesz theorem gives also the equality of the norms, indeed

\[
|f|_{L^2(\gamma)} = \sup_{g \in H_\gamma^*} \frac{x_{\gamma}^*(g)}{|g|_{L^2(\gamma)}} = \sup_{y^* \in H^*} \frac{x_{\gamma}^*(y^* - y^*(m_\gamma))}{|y^* - y^*(m_\gamma)|_{L^2(\gamma)}} = \sup_{y^* \in H^*} \frac{y^*(x)}{|y^*|_\gamma} = |x|_{H_\gamma}
\]

and we are done.

The following Corollary is a direct consequence of the above Proposition, we state it here for future reference.

**Corollary 3.3.7.** The Cameron-Martin space \( H_\gamma \) with the scalar product

\[
\langle x, y \rangle_{H_\gamma} = \langle R_\gamma^{-1}(x), R_\gamma^{-1}(y) \rangle_{L^2(\gamma)}
\]

is a Hilbert space, where \( R_\gamma \) is the operator defined in Lemma 3.3.4.

The norm induced by the scalar product \( \langle \cdot, \cdot \rangle_{H_\gamma} \) is the \( |\cdot|_{H_\gamma} \) norm defined in Definition 3.3.5 and the function \( R_\gamma \) is an isometry between \((H_\gamma, \langle \cdot, \cdot \rangle_{H_\gamma})\) and \((H_\gamma^*, \langle \cdot, \cdot \rangle_{L^2(\gamma)})\).

In \( \mathbb{R}^n \), we are used to the fact that all Gaussian measures are absolutely continuous with respect to each other. If the space is infinite dimensional, this is not true. A measure and a translation of it can be singular with respect to each other. The next lemmas and propositions aims to show this fact.

We denote by \( \tau_x : H \to H \) the translation of \( x \) in \( H \), \( \tau_x(y) = y + x \), so that the translated of a measure \( \gamma \) by \( x \) is \( \tau_x^\sharp \gamma \).

**Lemma 3.3.8.** Let \( \gamma \) and \( \nu \) two finite measures on a measure space \((\Omega, F)\). Then they are orthogonal if and only if the total variation of their difference is

\[
\|\gamma - \nu\| = \gamma(\Omega) + \nu(\Omega)
\]

Proof. It is a well known fact that, when considering the difference of two positive measures, the total variation is given by the formula

\[
\|\gamma - \nu\| = \sup \{ \gamma(E) + \nu(F) \mid E, F \text{ measurable, } E \cap F = \emptyset \}.
\]

From the above formula it is clear that \( \|\gamma - \nu\| \leq \gamma(\Omega) + \nu(\Omega) \). If \( \gamma \) and \( \nu \) are singular, then there exist two disjoint sets \( E \) and \( F \) such that \( \gamma \) is concentrated on \( E \) and \( \nu \) is concentrated on \( F \). It follows that

\[
\|\gamma - \nu\| \geq \gamma(E) + \nu(F) = \gamma(\Omega) + \nu(\Omega)
\]

and one implication is proven.
To see the converse, suppose \( \| \gamma - \nu \| = \gamma(\Omega) + \nu(\Omega) \). Then for every \( n \in \mathbb{N} \) there exists a set \( E_n \) such that
\[
\gamma(E_n) + \nu(\Omega \setminus E_n) \geq \gamma(\Omega) + \nu(\Omega) - 2^{-n},
\]
in particular \( \gamma(E_n) \geq \gamma(\Omega) - 2^{-n} \) and \( \nu(E_n) \geq \nu(\Omega) - 2^{-n} \). Let
\[
E = \bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} E_n \quad \text{and} \quad F = \bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} \Omega \setminus E_n.
\]
The sets \( E \) and \( F \) are disjoint. Indeed, if \( x \in E \cap F \), then \( x \) eventually belongs to both \( E_n \) and \( \Omega \setminus E_n \), and this can not be true.

Moreover
\[
\gamma(E) \geq \gamma\left( \bigcap_{n \geq k} E_n \right) \geq \gamma(\Omega) - \sum_{n=k}^{\infty} 2^{-n}
\]
for every \( k \in \mathbb{N} \). Taking the supremum on \( k \),
\[
\gamma(E) = \gamma(\Omega)
\]
and similarly can be proven that \( \nu(F) = \nu(\Omega) \). Then \( \gamma \) and \( \nu \) are singular, because concentrated on disjoint sets.

\[\square\]

**Lemma 3.3.9.** Let \( H \) be a separable Hilbert space and \( \gamma \) a Gaussian measure on \( H \), \( g \) a function in \( H_\gamma^* \). Then the measure \( \nu \) given by density
\[
x \mapsto \exp\left( g(x) - \frac{1}{2} |g|_{L^2(\gamma)}^2 \right)
\]
with respect to \( \gamma \) is a Gaussian measure with Fourier transform
\[
\hat{\nu}(x^*) = e^{ix^*(R, g)} \hat{\gamma}(x^*) .
\]

The above lemma can be proven computing the Fourier transform of the measure \( \nu \). For details see Proposition 2.4.2 of [3].

**Theorem 3.3.10.** Let \( \gamma \) be a Gaussian measure on a separable Hilbert space \( H \) and \( x \in H \). Then,

1. if \( x \in H_\gamma \), the measures \( \tau_x \gamma \) and \( \gamma \) are equivalent;
2. if \( x \notin H_\gamma \), the measures \( \tau_x \gamma \) and \( \gamma \) are orthogonal.

In particular, the Cameron-Martin space can be characterized as
\[
H_\gamma = \{ x \in H \mid \tau_x \gamma \sim \gamma \} .
\]
Proof. Point 1 follows from Lemma 3.3.9. Indeed the Fourier transform of $\tau_{x^\sharp}\gamma$ is

$$\hat{\tau}_{x^\sharp}\gamma(y^*) = \int_H e^{iy^*(t+x)} d\gamma(t) = e^{iy^*(x)}\tilde{\gamma}(y^*)$$

and since $x \in H_\gamma$, by Proposition 3.3.6, there exists $g \in H^*_\gamma$ such that $x = R_\gamma(g)$.

By the injectivity of the Fourier transform (see Lemma 3.2.11) and Lemma 3.3.9, the measure $\tau_{x^\sharp}\gamma$ has density

$$y \mapsto \exp\left(g(y) - \frac{1}{2} |g|_{L^2(\gamma)}^2 \right)$$

with respect to $\gamma$.

To prove point 2 we show that the total variation of the difference is

$$\|\tau_{x^\sharp}\gamma - \gamma\| = 2$$

and this is sufficient to prove singularity because of Lemma 3.3.8 and the fact that both are probability measures. Since the Cameron-Martin space is invariant under translations of $\gamma$, we can assume $\gamma$ centered without loss of generality.

Keeping in mind that the total variation can be written as

$$\|\tau_{x^\sharp}\gamma - \gamma\| = \sup_{A \in B(H)} |\tau_{x^\sharp}\gamma(A) - \gamma(A)|,$$

it is easy to see the inequality

$$\|\tau_{x^\sharp}\gamma - \gamma\| \leq 2$$

and that for each measurable $f : H \to \mathbb{R}$,

$$\|\tau_{x^\sharp}\gamma - \gamma\| \geq \|f_\sharp\tau_{x^\sharp}\gamma - f_\sharp\gamma\|.$$

Since $x \notin H_\gamma$, for each $n \in \mathbb{N}$ there exists $y^*_n$ in $H^*$ such that $|y^*_n\gamma| = 1$ and $y^*_n(x) > n$. Observe that

$$y^*_n \tau_{x^\sharp}\gamma = (y^*_n \circ \tau_x)\gamma = \tau_{y^*_n(x)}\gamma y^*_n \gamma.$$

By assumption, $y^*_n(x) \to \infty$ and $\nu = y^*_n\gamma$ does not depend on $n$, since it is a standard Gaussian measure on $\mathbb{R}$, with mean 0 and variance 1.

Putting all together we get the other inequality

$$\|\tau_{x^\sharp}\gamma - \gamma\| \geq \|y^*_n \tau_{x^\sharp}\gamma - y^*_n \gamma\| = \|\tau_{y^*_n(x)\nu - \nu}\| \to_{n \to \infty} 2$$

because $y^*_n(x) \to \infty$.

Another characterization of the Cameron-Martin space could be given in terms of full measure subspaces.

**Proposition 3.3.11.** Let $\gamma$ be a centered Gaussian measure on a separable Hilbert space $H$. Then the Cameron-Martin space is the intersection of all measurable subspace of full $\gamma$ measure.
Proof. Let $L$ be a measurable subspace of full $\gamma$ measure. For all $h \in H_\gamma$, $\gamma(L - h) = \tau_h \gamma(L) = 1$ since $\tau_h \gamma \ll \gamma$ by Theorem 3.3.10. Then $\gamma$ is concentrated on $(L - h) \cap L$. But this intersection is empty unless $h \in L$ and so $H_\gamma \subseteq L$.

Conversely, if $h \not\in H_\gamma$ there exists a sequence $\{x_n^*\}_{n \in \mathbb{N}} \subseteq H^*$ such that $|x_n^*|_\gamma = 1$ and $x_n^*(h) > n$. Let $L$ be the measurable subspace of $H$ defined as

$$L = \left\{ x \in H \mid \sum_{i=1}^{+\infty} \frac{x_n^*(x)}{n^2} < +\infty \right\}.$$ 

The element $h$ does not belong to $L$. Denoted by $f$ the function

$$f(x) = \sum_{i=1}^{+\infty} \frac{x_n^*(x)}{n^2},$$

it holds

$$\int_H |f| \, d\gamma \leq \sum_{i=1}^{+\infty} \frac{1}{n^2} \int_H |x_n^*| \, d\gamma \leq \sum_{i=1}^{+\infty} \frac{1}{n^2} \left( \int_H (x_n^*)^2 \, d\gamma \right)^{\frac{1}{2}} = \sum_{i=1}^{+\infty} \frac{1}{n^2} |x_n^*|_\gamma < +\infty$$

by the Cauchy-Schwartz inequality, the fact that $\gamma$ is centered and the fact that $|x_n^*|_\gamma = 1$. Then $f$ should be finite almost everywhere and $L$ is a subspace of full $\gamma$ measure that does not contain $h$.

We have then proven that the intersection of all subspaces on which $\gamma$ is concentrated is exactly the Cameron-Martin space. \hfill \□

The next theorems, are related to the question of “how many” directions there are in the Cameron-Martin space. They show that the Cameron-Martin space is small in some sense, but big in another.

We present this two results here to keep together all facts about the Cameron-Martin space, but in the proofs we need some results from the next section.

**Proposition 3.3.12.** Let $\gamma$ a Gaussian measure on a separable Hilbert space $H$ and suppose that $\gamma$ is not concentrated on any finite dimensional subspace. Then the Cameron-Martin space is negligible.

Of course, to get the hypothesis of this Proposition satisfied, $H$ should be infinite dimensional.

**Proof.** By Lemma 3.4.1, the Hilbert space $H^*_\gamma$ is infinite dimensional. We can then take a countable orthonormal sequence $\{f_n\}_{n \in \mathbb{N}} \subseteq H^*_\gamma$. The function $f_n$ are Gaussian random variables by Lemma 3.3.3, they are orthogonal and so by Lemma 3.1.9 they are independent. Moreover, since they are centered and have variance equal to 1, they have the same law, that we can call $\nu$. 
The Cameron-Martin space is such that
\[ H_\gamma = \bigcup_{n \in \mathbb{N}} \{ x \in H \mid |x|_{H_\gamma} < n \} \subseteq \bigcup_{n \in \mathbb{N}} \bigcap_{i \in \mathbb{N}} \{ x \in H \mid f_i(x) < n \} \]
since the \( f_n \) have variance equal to 1. Using that the \( f_n \) are independent, the measure of those intersections can be estimated as
\[
\gamma \left( \bigcap_{i \in \mathbb{N}} \{ f_i(x) < n \} \right) \leq \gamma \left( \bigcap_{i=1}^{m} \{ f_i < n \} \right) = \prod_{i=1}^{m} \gamma(\{ f_i < n \}) = \gamma(\{ f_1 < n \})^m \to 0
\]
as \( m \to \infty \) and so we have covered the Cameron-Martin space with a countable union of null sets.

The next proposition characterize the support of a Gaussian measure, and its corollary gives a nice characterization of non degenerate Gaussian measures. We omit the proof, referring the reader to Theorem 3.6.1 of [3].

**Proposition 3.3.13.** Let \( \gamma \) be a Gaussian measure on a separable Hilbert space \( H \), \( m_\gamma \) its mean and \( H_\gamma \) the Cameron-Martin space.

Then the support of \( \gamma \) is an affine subspace obtained translating by \( m_\gamma \) the closure of the Cameron-Martin space.

\[ \text{Supp}(\gamma) = m_\gamma + \overline{H_\gamma} \]

**Corollary 3.3.14.** A Gaussian measure \( \gamma \) on a Hilbert space \( H \) is non degenerate if and only if its support is the whole \( H \).

**Proof.** Let \( f \in H^* \). The Gaussian measure \( f_\sharp \gamma \) is degenerate if and only if \( \gamma \) is concentrated on \( f(m_\gamma) + \ker(f) \), which is a closed subset of \( H \).

If \( \gamma \) is degenerate, by definition, there exists \( x^* \in H^* \) such that \( x_\sharp^* \gamma \) is degenerate and so the support is smaller than \( H \).

Conversely, the support of \( \gamma \) is the affine subspace \( m_\gamma + \overline{H_\gamma} \) by Proposition 3.3.13. If \( \overline{H_\gamma} \) is smaller than \( H \), there exists \( x \in H \) orthogonal to \( \overline{H_\gamma} \) and then \( x_\sharp^* \gamma \) is degenerate. \( \square \)

### 3.4 Miscellaneous facts

In this section we present some other interesting facts about Gaussian measures. They are not really related to each other, and most of them are cited in previous or following chapters. For this reason we try to make every statement as self-contained as possible.

We still denote by \( \gamma \) a Gaussian measure on a separable Hilbert space \( H \) and use the notation of Definition 3.3.5 and 3.3.2 for the Cameron-Martin space and the space \( H^*_\gamma \).
Lemma 3.4.1. Let $\gamma$ be a Gaussian measure on a Hilbert space $H$. Then $H^*_\gamma$ is infinite dimensional if and only if $\gamma$ is not concentrated on any finite dimensional subspace of $H$.

Proof. Suppose $\gamma$ concentrated on a finite dimensional subspace $L$. Then for all $x^* \in H^*$ such that $x$ is orthogonal to $L$, the functional $x^*$ is 0 almost everywhere and so $x^* = 0$ in $L^2(\gamma)$. This means that $H^*_\gamma$ is the closure of $L^*$, which is finite dimensional, in $L^2(\gamma)$ and so it is finite dimensional as well.

Conversely, suppose that there is no finite dimensional subspace on which $\gamma$ is concentrated and consider the smallest closed subspace $L$ that contains the topological support of $\gamma$. Since $H$ is separable, $\gamma$ is concentrated on its support and then also on $L$. By hypothesis, $L$ should be infinite dimensional.

Consider an element $x \in L$ such that $x^* = 0$ in $L^2(\gamma)$. This means that $\gamma$ is concentrated on $L' = L \cap \{ y \in H \mid \langle x, y \rangle_H = 0 \}$.

The subspace $L'$ is a closed set, and so it contains the support of $\gamma$, but then should be equal to $L$, which is the smallest closed subspace with this property. This is possible only if $x = 0$.

This says that the inclusion $L^* \to L^2(\gamma)$ is injective. Since $L^*$ is an infinite dimensional vector space and the inclusion is linear, its image should be infinite dimensional as well. It is contained in $H^*_\gamma$ and so we are done. $\square$

In the previous section we saw that the behaviour of a Gaussian measure under translation is different from the finite dimensional case. We now see that this is true also under scaling.

Proposition 3.4.2. Let $\gamma$ be a centered Gaussian measure on a separable Hilbert space, not concentrated on any finite dimensional subspace of $H$. Then there exist a sequence $\{x^*_n\}$ of linear functional such that for almost every $y \in H$

$$\frac{1}{n} \sum_{i=1}^{n} x^*_i(y)^2 \to 1$$

as $n \to \infty$. In particular, $\gamma$ is concentrated on the set $C_\gamma$ defined as

$$C_\gamma = \left\{ x \in H \mid \frac{1}{n} \sum_{i=1}^{n} x^*_i(y)^2 \xrightarrow{n \to \infty} 1 \right\}$$

Proof. By Lemma 3.4.1, $H^*_\gamma$ is infinite dimensional. The space $H^*$ is dense in $H^*_\gamma$ in the $L^2(\gamma)$ norm by definition of $H^*_\gamma$ and so we can take a sequence $\{x^*_n\}_{n \in \mathbb{N}} \subseteq H^*$ orthonormal in $L^2(\gamma)$.

The $x^*_n$ are orthogonal Gaussian random variables, and so by Proposition 3.1.9 they are independent. From the fact that $\gamma$ is centered and that the $x^*_n$ have variance equal to 1, it follows that they are also equally distributed.
3.4. MISCELLANEOUS FACTS

So we have a sequence of independent, identically distributed random variables. Apply the law of large numbers to their square and get

\[
\frac{1}{n} \sum_{i=1}^{n} x^*_n(y)^2 \to \int_H (x^*_1)^2 \, d\gamma = 1
\]

almost surely.

The set \( C_\gamma \) could be seen as a kind of “ellipsoid”, in sense that for every \( \lambda \neq \pm 1 \), the set \( \lambda C_\gamma \) is disjoint from \( C_\gamma \). This is the key observation to prove the next (surprising) corollary.

**Corollary 3.4.3.** Let \( \gamma \) be a centered Gaussian measure on a separable Hilbert space \( H \), not concentrated on any finite dimensional subspace, and denote by \( f_\lambda: H \to H \) the scaling by \( \lambda \)

\[
f_\lambda(x) = \lambda x .
\]

Then for every \( \lambda \neq \pm 1 \) the measures \( f_\lambda \gamma \) and \( \gamma \) are mutually singular.

**Proof.** Let \( C_\gamma \) be defined as in Proposition 3.4.2. Consider a point \( x \in C_\gamma \). By linearity of the \( x_n^* \),

\[
\frac{1}{n} \sum_{i=1}^{n} x_i^* \left( \frac{1}{\lambda} x \right)^2 = \left( \frac{1}{\lambda} \right)^2 \frac{1}{n} \sum_{i=1}^{n} x_i^*(x)^2 \to \frac{1}{\lambda^2}
\]

and so, if \( \lambda \neq \pm 1 \), \( \frac{1}{\lambda} x \notin C_\gamma \) and \( x \notin \lambda C_\gamma \).

It follows that \( C_\gamma \) and \( \lambda C_\gamma \) are disjoint, but \( \gamma \) is concentrated on \( C_\gamma \) and \( f_\lambda \gamma \) is concentrated on \( \lambda C_\gamma \) and so the two measures are mutually singular. \( \square \)
Chapter 4

Probability measures on manifolds

In this chapter we finally deal with probability measures on manifolds. We have seen in Chapter 2 a possible application of this study.

In the current literature many different approaches can be found to endow a finite dimensional manifold with a measure. We concentrate on three particular approaches and try to generalize them to the infinite dimensional case. The tractation here is by no means exhaustive and much work can still be done in studying other approaches.

We do not pay attention to the regularity of the manifolds we are dealing with. For simplicity, we assume manifolds are $C^\infty$ submanifolds of some Hilbert space, and all the examples satisfy this hypothesis. However, the $C^\infty$ regularity is not essential and many results still work with far less regularity.

4.1 Hausdorff measures

If $M$ is an $n$ dimensional manifold embedded in $\mathbb{R}^m$ for some $m \in \mathbb{N}$, a natural way to put a (possibly non finite) measure on it is to restrict the Hausdorff measure $\mathcal{H}^n$. If the Hausdorff measure of the whole $M$ is finite, for example when the manifold is compact, the Hausdorff measure can be scaled to become a probability measure.

In Section 1.1 we defined a measure as a finite measure. So the Hausdorff measure is not properly a measure and should be referred to as a non finite measure. However in this section, when there is no risk of confusion, we just say measure.

We now briefly recall the definition of the Hausdorff measure $H^n$ and some of his properties, referring to [12] for other results and proofs.

Let $E \subseteq \mathbb{R}^m$ be a Borel set. To define the Hausdorff measure, one considers the coverings $\{B_i\}_{i \in \mathbb{N}}$ of $E$, where $B_i$ are subsets of $\mathbb{R}^m$ of “small” diameter
and define the measure as

\[ H^n(E) = \sup_{\varepsilon > 0} \inf_{E \subseteq \bigcup B_i, \text{diam}(B_i) < \varepsilon} \sum_{i=1}^{+\infty} \frac{\omega_n}{2^n} \text{diam}(B_i)^n \]

where \( \text{diam}(B_i) \) denotes the diameter of \( B_i \), i.e.

\[ \text{diam}(B_i) = \sup_{x, y \in B_i} |x - y|, \]

and \( \omega_n \) is a scale factor, usually chosen equal to the Lebesgue measure of the unit ball in \( \mathbb{R}^n \).

This formula defines a non finite measure on the Borel \( \sigma \)-algebra, \( \mathcal{B}(\mathbb{R}^m) \). It can be restricted to \( \mathcal{B}(M) \).

When \( n = m \) the Hausdorff measure coincide with the Lebesgue measure \( \mathcal{L}^m \) on \( \mathbb{R}^m \).

If instead of allowing the \( B_i \) to be any subset of \( \mathbb{R}^m \) we ask the \( B_i \) to be open, or to be closed, the resulting measure is the same. In general the result is different if the \( B_i \) are asked to be balls. The non finite measure defined on Borel sets \( E \subseteq \mathbb{R}^m \) considering only ball coverings is called spherical Hausdorff measure and denoted by \( \mathcal{S}^n \),

\[ \mathcal{S}^n(E) = \sup_{\varepsilon > 0} \inf_{E \subseteq \bigcup B_i, \text{ball} B_i \text{ ball}} \sum_{i=1}^{+\infty} \omega_n r_i^n \]

where \( r_i \) is the radius of \( B_i \).

Let \( f: \mathbb{R}^n \to \mathbb{R}^m \) be a Lipschitz function, with \( n < m \). The area factor of the function \( f \) is defined as

\[ Jf(x) = \begin{cases} \det \left( \nabla f(x)^T \cdot \nabla f(x) \right) & \text{if } f \text{ is differentiable at } x \\ +\infty & \text{otherwise} \end{cases} \]

where \( \nabla f(x) \) is the differential of \( f \) at \( x \).

The area factor is finite almost everywhere by Rademacher’s theorem, a proof of which can be found in [12], Theorem 3.1.6.

**Theorem 4.1.1 (Rademacher’s theorem).** Let \( U \subseteq \mathbb{R}^n \) be an open set and \( f: U \to \mathbb{R}^m \) a Lipschitz function and \( \mathcal{L}^n \) the Lebesgue measure on \( \mathbb{R}^n \). Then \( f \) is differentiable for \( \mathcal{L}^n \) almost every \( x \in U \).

The area formula relates the integral with respect to the Hausdorff measure to integrals with respect to the Lebesgue measure on the domain of the function. In case \( M \) is a manifold embedded in \( \mathbb{R}^m \), the area formula says that the Hausdorff measure coincide with the measure defined by the canonical volume form, given to \( M \) the Riemannian structure induced by the embedding in \( \mathbb{R}^m \).

This formula is proven in [12], see Theorem 3.2.3 and Section 3.2 for more details.
4.1. HAUSDORFF MEASURES

Proposition 4.1.2 (Area formula). Let $n < m$ be natural numbers, $E \subseteq \mathbb{R}^n$ a Borel set and $f : \mathbb{R}^n \to \mathbb{R}^m$ a Lipschitz function. Then for every integrable function $\varphi : E \to \mathbb{R}$, the function on $\mathbb{R}^m$

$$y \mapsto \sum_{x \in E \atop f(x) = y} \varphi(x)$$

is measurable and integrable and

$$\int_E \varphi(x) Jf(x) \, d\mathcal{L}^n(x) = \int_{f(E)} \sum_{x \in E \atop f(x) = y} \varphi(x) \, d\mathcal{H}^n(y).$$

Is it possible to generalize the Hausdorff measure to subsets of a Hilbert space? Actually the formula that defines the Hausdorff measure in $\mathbb{R}^m$ make sense in a metric space, just adapting the notion of diameter. That formula defines a non finite measure also on the Borel subsets of a metric space, see Section 2.10 of [12].

However the Hausdorff measure $\mathcal{H}^n$ it is not finite on infinite dimensional submanifolds of a Hilbert space.

A generalization of the Hausdorff measure for finite codimension objects in a Hilbert space have been discussed in [13], which deals in general with locally convex spaces. We know expose the main ideas of the construction, using the simplified approach proposed in [1] and restricting ourselves to the definition in Hilbert spaces. The reference for the quite technical proof of well-definiteness and other proofs is [13].

The definition of the $n$ codimension Hausdorff measure is given in an infinite dimensional Hilbert space $H$ with a centered Gaussian measure $\gamma$. Suppose $\gamma$ is non degenerate, or equivalently its support is the whole $H$ (the equivalence is proven in Corollary 3.3.14).

We denote by $\langle \cdot, \cdot \rangle_H$ the scalar product in $H$ and for the rest we adopt the notation of Chapter 3. In particular the dual of $H$ is denoted by $H^*$ and for every $x \in H$, $x^*$ is the element of the dual such that

$$x^*(y) = \langle x, y \rangle_H \quad \text{for each } y \in H.$$

The Cameron-Martin space $H_\gamma$ is defined in Definition 3.3.5, the symbol $H_\gamma^*$ in Definition 3.3.2 and $R_\gamma$ in Lemma 3.3.4. The Cameron-Martin space $H_\gamma$ is a Hilbert space with the scalar product $\langle \cdot, \cdot \rangle_{H_\gamma}$, defined in Corollary 3.3.7.

The general plan for the construction is to first slice a Borel set in finite dimensional slices, compute a suitable Hausdorff measure of the slices and then integrate the measure of the slices. This leads to a result which depends on the chosen slices and so in the end a supremum is taken.

Let $F \subseteq R_\gamma(H^*)$ be a finite dimensional subspace of the Cameron-Martin space, contained in $R_\gamma(H^*)$ and of dimension $m > n$. We consider $F$ endowed with the scalar product $\langle \cdot, \cdot \rangle_{H_\gamma}$, This is a key point, keep in mind that orthogonality and norms in $F$ are always referred to the $\langle \cdot, \cdot \rangle_{H_\gamma}$ scalar product, regardless to the fact that $H$ has its own scalar product and norm.
CHAPTER 4. PROBABILITY MEASURES ON MANIFOLDS

The spherical \( m - n \) Hausdorff measure can be defined on Borel subsets \( E \subseteq F \) as above,

\[
\mathcal{H}^{m-n}(E) = \sup_{\varepsilon > 0} \ inf_{E \subseteq \bigcup B_i} \ \sum_{r_i < \varepsilon}^\infty \omega_n r_i^n
\]

where the \( B_i \) are balls of radius \( r_i \) with respect to the \( \| \cdot \|_{H^\gamma} \) norm.

Fix an orthonormal basis \( f_1 \ldots f_m \) of \( F \) and let \( g_i^* = R_{\gamma}^{-1}(f_i) \) for \( i = 1 \ldots m \). Note that \( R_{\gamma}^{-1}(f_i) \in H^* \) since \( F \subseteq R_{\gamma}(H^*) \) and so the notation \( g_i^* \) is justified.

The orthogonal projection from \( H^\gamma \) to \( F \)

\[
\pi_F(x) = \sum_{i=1}^m \langle f_i, x \rangle_{H^\gamma} f_i = \sum_{i=1}^m g_i^*(x) f_i = \sum_{i=1}^m \langle g_i, x \rangle_H f_i
\]

can be extended to a linear continuous projection \( \pi_F : H \to F \) using the right-most formula.

The projection \( \pi_F \) induces a decomposition \( H = F \oplus \ker(\pi_F) \). This is not the orthogonal decomposition that can be expected in a Hilbert space, but decompose the measure \( \gamma \) as

\[
\gamma = \pi_F^* \gamma \otimes (I - \pi_F)^* \gamma
\]

where \( I \) is the identity function. We denote by \( \gamma^\perp \) the measure on \( \ker(\pi_F) \)

\[
\gamma^\perp = (I - \pi_F)^* \gamma
\]

Let \( E \in \mathcal{B}(H) \) be a Borel subset of \( H \). For every \( y \in \ker(\pi_F) \), the slice \( E_y \) of \( E \) is a subset of \( F \) defined as

\[
E_y = \{ x \in F \mid y + x \in E \}.
\]

Let also \( G_m(x) \) be a “Gaussian” density defined on \( H \) as

\[
G_m(x) = \frac{1}{(2\pi)^{\frac{m}{2}}} \exp \left( -\frac{1}{2} \| x \|_H^2 \right).
\]

We can now define a not finite measure on \( \mathcal{B}(H) \) as

\[
\mathcal{H}^{\infty-n}_F(E) = \int_{\ker(\pi_F)} \int_{E_y} G_m(x) \ d\mathcal{H}^{m-n}(x) \ d\gamma^\perp(y).
\]

It can be proven that for \( \gamma^\perp \) almost every \( y \in \ker(\pi_F) \) the set \( E_y \) belongs to \( \mathcal{B}(F) \), that the map \( y \mapsto \int_{E_y} G_m(x) \ d\mathcal{H}^{m-n}(x) \) is \( \gamma^\perp \) measurable and that \( \mathcal{H}^{\infty-n}_F \) is a \( \sigma \)-additive function.

Moreover if \( F_1 \subseteq F_2 \),

\[
\mathcal{H}^{\infty-n}_{F_1} \leq \mathcal{H}^{\infty-n}_{F_2}.
\]

and this allows to define the \( n \) codimension Hausdorff measure on \( \mathcal{B}(H) \) as

\[
\mathcal{H}^{\infty-n}(E) = \sup_F \mathcal{H}^{\infty-n}_F(E).
\]
4.2. PROJECTIONS

4.2 Image 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 put on $M$ the image measure $f_*\mathbb{P}$.

**Example 4.2.1.** 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

$$\pi: \mathbb{R}^{n+1} \setminus \{0\} \to S^n \quad x \mapsto \frac{x}{|x|},$$

which is defined $\gamma$ almost everywhere.

Then the measure $\pi_*\gamma$ on $S^n$ coincides with the Hausdorff measure $\mathcal{H}^n$ restricted to the sphere and normalized.

The above example can be generalized to a manifold embedded in $\mathbb{R}^m$, provided that we define a projection on the manifold. In the following, let $M$ be a manifold embedded and closed in $\mathbb{R}^m$, $d$ the euclidean distance on $\mathbb{R}^m$

$$d(x, y) = |x - y|$$

and $d_M: \mathbb{R}^m \to \mathbb{R}$ the distance from the manifold, defined by

$$d_M(x) = \inf_{y \in M} d(x, y). \quad (4.1)$$

Since $M$ is closed and $\mathbb{R}^m$ is locally compact, the infimum is a minimum, and then for all $x \in \mathbb{R}^m$ 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 in $M$ is unique, the projection $\pi$ can be defined by

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

Let $E_M$ be the set of all points for which the projection is not well defined, namely one could find $y_1, y_2 \in M$, $y_1 \neq y_2$ such that

$$d(x, y_1) = d(x, y_2) = d_M(x).$$

We show that $E_M$ is negligible with respect to the Lebesgue measure $\mathcal{L}^m$, and then the measure $\pi_*\gamma$ on $M$ make sense for all $\gamma$ absolutely continuous with respect to $\mathcal{L}^m$.

This fact is true in a much more general setting. Indeed, when considering the distance from a closed subset in a $m$ dimensional Riemannian manifold, the set of points for which the closest point is not unique has Hausdorff dimension at most $m - 1$. See [20] for more details. Here we prove only the special case stated above.
Lemma 4.2.2. Let \( x \in \mathbb{R}^m \) and \( y \in M \) points such that
\[
d(x, y) = d_M(x).
\]
Then for all \( \lambda \in [0, 1] \) the distance from the manifold \( M \) of the convex combination \( c_\lambda(x, y) = \lambda x + (1 - \lambda)y \) is
\[
d_M(c_\lambda(x, y)) = d(c_\lambda(x, y), y) = \lambda d_M(x)
\]
\[
\text{Proof.} \text{ The second equality can be easily proven, in fact for all } \lambda
\[
d(c_\lambda(x, y), y) = |\lambda x + (1 - \lambda)y - y| = \lambda |x - y| = \lambda d_M(x).
\]
Moreover the inequality
\[
d_M(c_\lambda(x, y)) \leq d(c_\lambda(x, y), y)
\]
follows from the definition of \( d_M \). Suppose by contradiction that they are not equal, that is there exist \( \lambda \in [0, 1] \) and \( z \in M \) such that
\[
d(c_\lambda(x, y), z) < d(c_\lambda(x, y), y).
\]
Then, for the triangular inequality,
\[
d(x, z) \leq d(x, c_\lambda(x, y)) + d(c_\lambda(x, y), z) <
\]
\[
d(x, c_\lambda(x, y)) + d(c_\lambda(x, y), y) = d(x, y) = d_M(x)
\]
and this is a contradiction because \( d_M(x) \) is the minimum of the distance from a point in \( M \).

Proposition 4.2.3. Let \( E_M \) be the set of points \( x \in \mathbb{R}^m \) such that there is more than one point in \( M \) that realize the minimum of the distance from \( x \). Then \( E_M \) is contained in the set of points at which the distance function \( d_M \) is not differentiable.

\[
\text{Proof.} \text{ Let } x \in \mathbb{R}^m \text{ and } y_1, y_2 \in M \text{ such that}
\]
\[
d_M(x) = d(x, y_1) = d(x, y_2) \quad \text{and} \quad y_1 \neq y_2.
\]
Consider the derivative at \( x \) of the function \( d_M \) along the direction \( x - y_1 \). It is the derivative at \( \lambda = 1 \) of the function \( d_M(\lambda x + (1 - \lambda)y_1) \).

By Lemma 4.2.2, if \( \lambda \in [0, 1] \) the value of that function is \( \lambda d_M(x) \) and so the derivative, if it exists, should be \( d_M(x) \).

On the other side, if \( \lambda > 1 \), estimating \( d_M \) with the distance from \( y_2 \),
\[
d_M(\lambda x + (1 - \lambda)y_1) \leq d(\lambda x + (1 - \lambda)y_1, y_2) = |\lambda x - (\lambda - 1)y_1 - y_2| =
\]
\[
\sqrt{(\lambda - 1)^2 |x - y_1|^2 + 2(\lambda - 1) \langle x - y_1, x - y_2 \rangle + |x - y_2|^2} =
\]
\[
|x - y_2| + \frac{\langle x - y_1, x - y_2 \rangle}{|x - y_2|} (\lambda - 1) + o(\lambda - 1) =
\]
\[
d_M(x) + \frac{\langle x - y_1, x - y_2 \rangle}{d_M(x)} (\lambda - 1) + o(\lambda - 1)
\]
4.2. PROJECTIONS

The derivative is less than \( \frac{\langle x - y_1, x - y_2 \rangle}{d(x, y_2)} \).

By Cauchy inequality,
\[
\frac{\langle x - y_1, x - y_2 \rangle}{d_M(x)} \leq \frac{|x - y_1| |x - y_2|}{d_M(x)} = d_M(x)
\]
with equality if and only if \( x - y_1 \) and \( x - y_2 \) are linearly dependent and co-oriented, but this could not be, since \( x - y_1 \) and \( x - y_2 \) have the same norm and \( y_1 \neq y_2 \). It follows that the two derivatives are different and \( d_M \) is not differentiable at \( x \).

\[\square\]

**Lemma 4.2.4.** The function \( d_M \) is a Lipschitz function.

**Proof.** Let \( x_1, x_2 \) be two points in \( \mathbb{R}^m \), and \( y_1 \in M \) a point such that
\[
d(x_1, y_1) = d_M(x_1).
\]
Then, by definition of \( d_M \) and triangular inequality,
\[
d_M(x_2) - d_M(x_1) \leq d(x_2, y_1) - d_M(x_1) \leq d(x_2, y_1) - d(x_1, y_1) \leq d(x_1, x_2).
\]
Exchanging \( x_1 \) and \( x_2 \) one gets,
\[
|d_M(x_2) - d_M(x_1)| \leq d(x_1, x_2)
\]
and so 1 is a Lipschitz constant for the function \( d_M \).

\[\square\]

Now the fact that \( E \) is negligible follows from the well known Rademacher’s theorem, stated before as Theorem 4.1.1.

We can know recollect what we have proven so far.
Theorem 4.2.5. Let $M$ be a manifold embedded and closed in $\mathbb{R}^m$, and $\mathcal{L}^m$ the Lebesgue measure on $\mathbb{R}^m$. Then for $\mathcal{L}^m$ almost every $x$ there exists a unique point $\pi(x) \in M$ that realizes the minimum of the distance from $x$, i.e.

$$|x - \pi(x)| = \inf_{z \in M} |x - z|$$

Moreover, given a measure $\gamma$ which is absolutely continuous with respect to $\mathcal{L}^m$, the measure $\pi_*\gamma$ is well defined on $M$.

Proof. By local compactness of $\mathbb{R}^m$, there always exists a point $y \in M$ that realizes the minimum. Moreover, by Proposition 4.2.3, we know the set of points $E$ for which there is more than one such point, is contained in the set of non differentiability point of the map $d_M$, defined by Equation 4.1.

Indeed it is sufficient to prove that the map $d_M$ is differentiable $\mathcal{L}^m$ almost everywhere. This follows from Rademacher’s theorem, see Theorem 4.1.1, since $d_M$ is Lipschitz by Lemma 4.2.4.

If we consider an infinite dimensional manifold, embedded in a Hilbert space, an analogous projection on the manifold does not necessarily exists. In this setting compactness fails and so there could be many points for which there is not a point on the manifold at minimal distance.

We first discuss a counterexample, and after show a couple of cases where the projection can be defined.

To do a counterexample to Theorem 4.2.5, one should first fix a reference measure on the Hilbert space, since there is no more a Lebesgue measure. We will use Gaussian measures, but let us talk about that later.

In the following let $H$ be a separable Hilbert space and $\{e_i\}_{i \in \mathbb{N}}$ a fixed orthonormal basis of $H$. If $x \in H$ is a point in the Hilbert space, let also $|x|$ be his norm and $x_i$ the $i$-th coordinate with respect to the basis $\{e_i\}_{i \in \mathbb{N}}$; since $x$ is determined by $\{x_i\}_{i \in \mathbb{N}}$, we would equivalently denote it by $(x_i)_{i \in \mathbb{N}}$.

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

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

Lemma 4.2.6. Consider in the Hilbert space $H$ the ellipsoid $S$ defined by

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

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 $1$

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

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 \( c \) for 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 [19] 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 \( c a_n^{-1} e_n \) for \( n \in \mathbb{N} \),

\[
d_S(0) \leq \inf_{n \in \mathbb{N}} |c a_n^{-1} e_n| = \inf_{n \in \mathbb{N}} c a_n^{-1} = c
\]

since \( a_i \not\to 1 \).

Lemma 4.2.6 shows that, given a Hilbert space \( H \), there exists a submanifold for which the distance from the origin does not have a minimum on the manifold. However it is not quite a counterexample to Theorem 4.2.5, because a single point will often 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 4.2.7.** Let \( \{a_i\}_{i \in \mathbb{N}}, c \) and \( S \) be an ellipsoid and its parameters, satisfying the hypothesis of Lemma 4.2.6. 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 $n \in \mathbb{N}$, saying that the minimum, 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 contained in $S$, i.e.

$$\sum a_i^2 x_i^2 < c^2$$

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 4.2.6, $S = f^{-1}(c^2)$, $f$ 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. $g(y) = |y - x|^2$. 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

$$x_i = (1 - \lambda a_i^2) z_i. \quad (4.2)$$

This equation gives us some information about $\lambda$. Since $x_i$ and $z_i$ are non negative

$$\lambda < \frac{1}{a_i^2} \quad \text{for all } i \text{ such that } x_i \neq 0. \quad (4.3)$$

Suppose that the point $x$ has infinitely many coordinates different from 0. Then, passing to the limit Equation 4.3,

$$\lambda \leq 1.$$
4.2. PROJECTIONS

Compute \( f(z) \) using Equation 4.2 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 \( \bar{y} \in \text{Span}(e_0, \ldots, e_n) \) and \( \hat{y} \in \text{Span}(e_0, \ldots, e_n)^\perp \). 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 \( \{a_{n+1}, a_{n+2}, \ldots\} \) 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 4.2.6 the infimum is equal to

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

The function in the above equation has a global minimum at point \( \tilde{z} \) with coordinates

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

Since \( x \in E_S \), the equation of \( E_S \) gives that \( \tilde{z} \) is such that

\[
f(\tilde{z}) < c^2
\]

and so \( \tilde{z} \) realize the infimum in Equation 4.4.

Now we are nearly done, because if \( g \) has a minimum \( z \) on \( S \) then its first component in the decomposition should be \( \tilde{z} \). The second component should be not null and minimize the distance from the origin on a ellipsoid. This contradicts Lemma 4.2.6 and so \( g \) has no minimum.

Now we state and prove that Theorem 4.2.5 is false in infinite dimensional a Hilbert space with a Gaussian measure.
Theorem 4.2.8. 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. Choose an orthonormal basis $\{e_i\}_{i \in \mathbb{N}}$ of $H$. The coordinate functions $x \mapsto x_i$ are Gaussian random variables and, denoted by $\sigma_i$ their variances,

$$\int_H |x|^2 \, d\gamma = \sum_{i=1}^{+\infty} \sigma_i^2.$$

Since the integral is finite by Proposition 3.2.7, then

$$\sum_{i=1}^{+\infty} \sigma_i^2 < +\infty.$$

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 4.2.6. By the same lemma $S$ is a manifold embedded and closed in $H$.

By Lemma 4.2.7 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 $a_i$ to get this integral finite. For this choice of $a_i$, the function $f(x)$ is finite $\gamma$ almost everywhere and, up to negligible sets,

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

so we can choose $c$ in a way that $E_S$ is not negligible for $\gamma$. 

In the converse, the projection on Stiefel manifolds exists, at least almost everywhere for all non degenerate Gaussian measures, and then the approach of considering the projection of a Gaussian measure in the ambient space may be used to define a probability measure on an infinite dimensional Stiefel manifold.
Example 4.2.9. Let $H$ be a separable Hilbert space and $S \subseteq H$ the unit sphere. Then the function

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

is defined in $H$ minus the origin and it is the projection on the nearest point of the sphere $S$. For all Gaussian measures different from the delta in the origin, the projection $\pi$ is defined almost everywhere.

Let $H$ be a separable Hilbert space and $\text{St}(p,H)$ a Stiefel manifold embedded in the separable Hilbert space $H^p$, as defined in Definition 2.2.1. We first characterize the point 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 4.2.10. Let $H$ be a separable Hilbert space, $p \in \mathbb{N}$, $\text{St}(p,H)$ a Stiefel manifold and $x \in H^p$. Then

1. if the components $x_1, \ldots, x_p$ are independent, there exists a unique point $\pi(x) \in \text{St}(p,H)$ 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.

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

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

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

$$v \mapsto \sum_{i=1}^{p} \langle x_i, v_i \rangle .$$

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

First of all we show that the maximum exists. 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 $X = \text{Span}(x_1, \ldots, x_p)$ 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 \( Y^p \cap \text{St}(p, H) \). Using the above basis, this intersection can be written as

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

where

\[
S = \left\{ a \in \mathbb{R}^{p \times (p+q)} \mid \forall j \neq k \sum_{i=1}^{q} a_{j,i}^2|x_i|^2 + \sum_{i=1}^{p} a_{j,i+q}^2 = 1 \right\}.
\]

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

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

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 \( \text{St}(p, H) \) there exists a “nice” subspace \( Y \subseteq H \) such that \( v \in Y^p \) and so the global supremum in \( \text{St}(p, H) \) 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

\[
v_{\min} = \left( \frac{x_1}{|x_1|}, \ldots, \frac{x_p}{|x_p|} \right).
\]

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 \( \text{St}(p, H) \) and then it is the minimum also on the Stiefel manifold.

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

\[
(x x^T)_{ij} = \langle x_i, x_j \rangle
\]

and let \( A \) be an orthonormal matrix that diagonalizes it,

\[
A xx^T A^T = \text{diag}(d_1, \ldots, d_n).
\]
For every \( y \in H^p \) the product \( Ay \in H^p \) can be defined taking linear combinations of the components \( y_1 \ldots y_p \), i.e.

\[
(Ay)_i = \sum_{j=1}^{p} A_{ij} y_j .
\]

Consider the mapping \( f_A : H^p \to H^p \), \( f_A(y) = Ay \). 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,

\[
|Ay|^2 = (Ay)^T A y = y^T (A^T A) y = y^T y = |y|^2 .
\]

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

\[
v v^T = Id \iff A(v v^T) A^T = Id \iff Av(A v)^T = Id .
\]

The components of \( Ax \) are orthogonal because \( A \) diagonalizes \( xx^T \)

\[
Ax(Ax)^T = A xx^T A^T = \text{diag}(d_1, \ldots, d_n) .
\]

By the observation above, there is a unique point \( v_{\text{min}} \) in \( \text{St}(p, H) \) that minimizes the distance from \( Ax \). Because of the proprieties of \( f_A \), the point \( A^T v_{\text{min}} \) is the unique point on \( \text{St}(p, H) \) at minimal distance from \( x \).

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

\[
\langle \tilde{v}_i, \tilde{v}_j \rangle = \langle v_i^x - v_i^\perp, v_j^x - v_j^\perp \rangle = \langle v_i^x, v_j^x \rangle + \langle v_i^\perp, v_j^\perp \rangle = \langle v_i^x + v_i^\perp, v_j^x + v_j^\perp \rangle = \langle v_i, v_j \rangle .
\]

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 \( v \neq \tilde{v} \). Moreover,

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

and then we have at least two minima. \( \square \)

If \( V \) is a vector space and \( p \in \mathbb{N} \) a natural number, we define

\[
\text{Dep}(p, V) = \{(x_1, \ldots, x_p) \in V^p \mid x_1, \ldots, x_p \text{ are linearly dependent} \} .
\]

To prove that the projection is defined almost everywhere, we need the following lemma.
Lemma 4.2.11. Let $p \leq n$ be positive natural numbers and consider the linear space $(\mathbb{R}^n)^p$ with the Lebesgue measure $(\mathcal{L}^n)^p = \mathcal{L}^n \times \cdots \times \mathcal{L}^n$. Then the set $\text{Dep}(p, \mathbb{R}^n) \subseteq (\mathbb{R}^n)^p$ is negligible.

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

Suppose the lemma true for $p - 1$ and decompose $\text{Dep}(p, \mathbb{R}^n)$ as $\text{Dep}(p - 1, \mathbb{R}^n) \times \mathbb{R}^n \cup \left\{ (x_1, \ldots, x_p) \in (\mathbb{R}^n)^p \mid (x_1, \ldots, x_{p-1}) \notin \text{Dep}(p-1, \mathbb{R}^n), x_p \in \text{Span}(x_1, \ldots, x_{p-1}) \right\}$.

The first set is negligible thanks to inductive hypothesis. Moreover for each $x_1, \ldots, x_{p-1} \in \mathbb{R}^n$, 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 $\mathcal{L}^n$-negligible. By Fubini’s theorem, the second set is negligible too, and so also $\text{Dep}(p, \mathbb{R}^n)$ is negligible.

Theorem 4.2.12. Let $H$ be a separable Hilbert space, $p \in \mathbb{N}$ and $\text{St}(p, H)$ a Stiefel manifold. Let also $\gamma$ be a non degenerate Gaussian measure on $H^p$, i.e. not concentrated on any closed subspace of $H^p$.

Then for almost every $x \in H^p$ there exists a unique point $\pi(x) \in \text{St}(p, H)$ that realize the minimum of the distance from $x$, i.e. $d(\pi(x), x) = d_{\text{St}(p, H)}(x)$.

Proof. By Proposition 4.2.10 for all points $x \notin \text{Dep}(p, H)$ there exists a unique point at minimal distance on $\text{St}(p, H)$. So we have to prove that $\text{Dep}(p, H)$ is negligible for the measure $\gamma$.

Fix an orthonormal basis $\{e_i\}_{i \in \mathbb{N}}$ of $H$, consider the projection $f$ on the first $p$ coordinates

$$f: \quad H \rightarrow \mathbb{R}^p \quad x \mapsto (\langle x, e_1 \rangle, \ldots, \langle x, e_p \rangle)$$

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

$$f^p: (x_1, \ldots, x_p) \mapsto (f(x_1), \ldots, f(x_p)).$$

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

By Lemma 4.2.11, $\text{Dep}(p, \mathbb{R}^p)$ is negligible for the Lebesgue measure $(\mathcal{L}^p)^p$. Since $\gamma$ 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 $(\mathcal{L}^p)^p$. It follows that

$$f^\sharp \gamma (\text{Dep}(p, \mathbb{R}^p)) = 0$$

and so $\text{Dep}(p, H)$ is negligible. \qed
4.3 Image of a probability measure under the exponential map

A different way to define a probability measure on a Riemannian manifold $M$ is to choose a point $p \in M$, define a probability measure on the tangent space $T_p M$ in $p$ and then put on $M$ the image of $\gamma$ under the exponential map.

This kind of measures are interesting because they may be used to define discrete stochastic processes on a manifold.

We recall briefly the definition of the exponential map. More details could be found in [19].

The exponential map $\exp_p : T_p M \to M$ is defined as

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

where $\sigma_v$ is the geodesic starting from $p$ with tangent vector $v$. It is a diffeomorphism in a ball around the origin of $T_p M$.

If $M$ is a finite dimensional complete Riemannian manifold, then the exponential map from any point is surjective by Hopf-Rinow theorem (see Theorem 2.8, Chapter 7 of [10]).

The tangent space $T_p M$ is a Hilbert space, so we can consider a Gaussian measure $\gamma$ on it. However 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.

If $M$ is an $n$-dimensional submanifold of $\mathbb{R}^m$, than the tangent spaces are $n$-dimensional subspaces of $\mathbb{R}^m$ and on them is defined the Lebesgue measure $\mathcal{L}^n$. In this case there is a comparison result.

**Theorem 4.3.1.** Let $M$ be a complete $n$-submanifold of $\mathbb{R}^m$, $p \in M$ a point and $T_p M$ the tangent space to $M$ in $p$. Let also $\gamma$ be a measure on $T_p M$, equivalent to the Lebesgue measure $\mathcal{L}^n$.

Then its image under the exponential map is equivalent to the Hausdorff measure restricted to $M$

$$\exp_p \gamma \sim \mathcal{H}^n.$$

**Proof.** The basic idea of this proof is to use the area formula for the exponential map to relate the Lebesgue measure on the tangent space and the Hausdorff measure on the manifold.

Let $E \subseteq T_p M$ be a negligible set for the Lebesgue measure $\mathcal{L}^n$ on $T_p M$. Then, by the area formula (see Proposition 4.1.2),

$$\int_{f(E)} \# (\exp_p^{-1}(x) \cap E) \, d\mathcal{H}^n(x) = \int_{E} \exp_p(x) \, d\mathcal{L}^n(x) = 0$$

where $\#$ denotes the cardinality of a set. It follows that

$$\mathcal{H}^n(f(E)) = 0$$

since by surjectivity of the exponential map $\#(\exp_p^{-1}E) \geq 1$ on $f(E)$. 
In the proof of the converse, a key role is played by the set
\[ \Sigma_p = \{ x \in T_p M \mid J \exp_p(x) = 0 \} . \]
It is possible to divide the tangent space \( T_p M \) in a countable union of measurable disjoint sets
\[ T_p M = \Sigma \cup \bigcup_{i \in \mathbb{N}} F_i \]
such that the exponential map restricted to each \( F_i \) is bi-Lipschitz. Indeed by the inverse function theorem each point in \( T_p M \setminus \Sigma \) has a neighbourhood in which \( \exp_p \) is a diffeomorphism. Possibly restricting the neighbourhood, \( \exp_p \) is also bi-Lipschitz. Since \( T_p M \) has a countable basis these neighbourhoods can be made countable and then disjoint taking differences.

Let \( E \subseteq M \) be a negligible set for the Hausdorff measure \( \mathcal{H}^n \). Its inverse image \( F \) under the exponential map can be written as
\[ F = \exp_p^{-1}(E) = (\Sigma \cap F) \cup \bigcup_{i \in \mathbb{N}} (F_i \cap F) . \]
Each of the \( F_i \cap F \) is negligible, since \( \exp_p \) restricted to \( F_i \) is bi-Lipschitz and the image of a \( \mathcal{H}^n \)-negligible set under a Lipschitz function is \( \mathcal{H}^n \)-negligible (on \( T_p M \) the measures \( \mathcal{H}^n \) and \( \mathcal{L}^n \) coincide).

It remains to see that the set \( \Sigma \) is negligible. This follows from Theorem 4.4 of [22] and the fact that \( \Sigma_p = \bigcup_{i=0}^{n-1} \{ x \in T_p M \mid \text{dexp}_p \text{ has rank } i \} \). \( \square \)

**Corollary 4.3.2.** Let \( M \) be a complete \( n \)-submanifold of \( \mathbb{R}^m \) and \( p_1, p_2 \) points on \( M \). Let also \( \gamma_1 \) and \( \gamma_2 \) be measures such that \( \gamma_i \) is a measure on \( T_{p_i} M \) equivalent to the Lebesgue measure on \( T_{p_i} M \) for \( i = 1, 2 \).

Then the images of \( \gamma_1 \) and \( \gamma_2 \) under the exponential maps are equivalent
\[ \exp_{p_1} \gamma_1 \sim \exp_{p_2} \gamma_2 . \]

If the manifold \( M \) is infinite dimensional, one can wonder if there could be a similar result. In the finite dimensional case, the comparison between measures on different tangent spaces is made comparing them with the Lebesgue measure, that could be defined in a standard way on all tangent spaces, once embedded in \( \mathbb{R}^m \). The first question to be answered when trying to generalize this result is how to compare measures on different tangent spaces, in order to get equivalent image measures under the exponential maps.

We show a counterexample to the equivalence of measures in the easier setting for comparison, when the tangent spaces coincide as subset of the ambient space.

We first state a lemma, which is useful to prove the counterexample.

**Lemma 4.3.3.** Let \( \gamma \) be a Gaussian measure on a separable Hilbert space \( H \).
Then every sphere has measure zero.
4.3. THE EXPONENTIAL MAP

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 3.2.9. As usual we denote by $e_1^*$ the functional $e_1^*: x \mapsto \langle x, e_1 \rangle$.

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 $\gamma = e_1^* \gamma \times \pi^* \gamma$.

We compute the measure of the sphere using the Fubini's theorem for the product measure $e_1^* \gamma \times \pi^* \gamma$. For every $x \in H'$, there are at most two $x_1$ such that $(x_1, x') \in S_r$. Since $e_1^* \gamma$ 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 $e_1^* \gamma$.

Proposition 4.3.4. 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 could be seen as subsets of $H$, and are both the set $T_p S = T_{-p} S = T = \{x \in H \mid \langle x, p \rangle = 0\}$.

Consider a Gaussian measure $\gamma$ on $T$, not concentrated on a finite dimensional subspace. Then the measures $\exp_p \gamma$ and $\exp_{-p} \gamma$ are mutually singular.

Proof. By Lemma 3.4.1 in the Hilbert space $L^2(\gamma)$ there exists an orthonormal sequence $\{f_i\}_{i \in \mathbb{N}}$ of continuous linear functional on $H$.

The functions $f_i$ could be regarded as real random variables on $H$. Since their laws are Gaussian, orthogonality implies independence and their square $f_i^2$ are a sequence of independent, identically distributed random variables with mean 1. As it has been observed yet in Proposition 3.4.2, by the law of large numbers, $\gamma$ is concentrated on the set

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

Since $f_i$ are linear, for every direction $x \in T$ there exists two or none $\lambda \in \mathbb{R}$ such that $\lambda x \in C$ and, if there are two, they have opposite sign.

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

$$\mu_1 = \exp_p \gamma \quad \mu_2 = \exp_{-p} \gamma,$$

and $C_1, C_2 \subseteq S$ the images of $C$ under $\exp_p$ and $\exp_{-p}$. 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 as

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

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 4.2).

The equator $T \cap S$ is negligible for $\mu_1$ (and also for $\mu_2$), 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 that spheres are negligible by Lemma 4.3.3. \hfill $\square$
Bibliography


67


