

The Hitchhiker's Guide to Cameron-Martin theory

Philipp Wacker, FAU Erlangen-Nürnberg

April 24, 2017

1 Introduction

The theory of Gaussian measures in infinite dimensional vector spaces is a prerequisite for both the theory of Bayesian inverse problems and general analysis and probability in infinite dimensions. It is also a fascinating subject in its own right but might initially seem austere and forbidding to the aspiring student (here the author (over-)generalizes his own experience). This document is supposed to serve as an appetizer and an introduction: A first course in functional analysis and probability theory should suffice in order to allow the reader to savour some of the lower-hanging fruit. The main goal of this summary is to build intuition for infinite-dimensional Gaussian measures and what pitfalls are there. This means that some of the proofs are only sketched, measure theory is not used and we will refer to more rigorous sources in the text. This exposition relies heavily on two sources: The lecture notes of the 19th Internet seminar on Infinite dimensional analysis, [3], and Nate Eldredge's wonderful lecture notes on analysis and probability in infinite dimensions [2]. I have stolen shamelessly from both notes, trying to combine the ideas from both treatises which helped me the most. This document deals only with a small fraction of both sources, i.e. the first few introductory chapters about infinite-dimensional probability measures. [\[Hinweis auf interessante Erklärungen: Gauss in Hilbert spaces, Unterschied zwischen RKHS und Cameron-Martin space\]](#)

2 Gaussian measures in \mathbb{R}^d

A Gaussian measure γ on \mathbb{R}^d is a probability measure with characteristic function (Fourier transform)

$$\hat{\gamma}(f) := \int_{\mathbb{R}^d} e^{if^T x} \gamma(dx) = \exp \left\{ i \cdot f^T a - \frac{1}{2} f^T \Sigma f \right\}$$

for some $m \in \mathbb{R}^d$ (we call this the mean) and $\Sigma \in \mathbb{R}^{d \times d}$ symmetric and positively semi-definite (we call this the covariance matrix). In short, $\gamma = N(m, \Sigma)$. We can also obtain mean and covariance by integration:

$$m = \int_{\mathbb{R}^d} x \gamma(dx), \quad \text{or} \quad f^T m = \int_{\mathbb{R}^d} f^T x \gamma(dx) \text{ for } f \in \mathbb{R}^d$$

and

$$\Sigma = \int_{\mathbb{R}^d} (x - m) \cdot (x - m)^T \gamma(dx), \quad \text{or} \quad f^T \Sigma g = \int_{\mathbb{R}^d} f^T (x - m) \cdot g^T (x - m) \gamma(dx).$$

The latter definitions in each of the last two expressions allow indirect calculation of m and Σ by “testing” with suitable $f, g \in \mathbb{R}^d$. The best method for that is choosing $f = (1, 0, \dots, 0) = e_1$, then $f = (0, 1, 0, \dots, 0) = e_2$ etc. when calculating $f^T m$ and all combinations for f and g for the calculation of $f^T \Sigma g$ and then $m = (e_1^T m, e_2^T m, \dots, e_d^T m)$ and $\Sigma = (\Sigma_{i,j})_{i,j=1}^d$ with $\Sigma_{i,j} = e_i^T \Sigma e_j$.

It may look strange to calculate mean and variance in that way if there is an explicit formula for m and Σ themselves, but the “testing method” generalizes more easily to infinite dimensions without the use of Bochner integrals (note that the expressions for m and Σ are actually vector-valued and matrix-valued)!

In finite dimensions we have an expression for the density of the measure with respect to the Lebesgue measure on \mathbb{R}^d if Σ is non-singular:

$$\frac{d\gamma}{d\lambda}(x) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp \left\{ -\frac{x^T \Sigma^{-1} x}{2} \right\}.$$

This is nice for calculations in finite dimensions but completely worthless in the context of infinite dimensions as there is no analogue of Lebesgue measure there. So better forget this immediately and focus on the characterization via its Fourier transform. Another definition that generalizes easily to the infinite-dimensional case is via projection on one component: Given any $f \in \mathbb{R}^d$, a vector $x \in \mathbb{R}^d$ is distributed according to a Gaussian measure if and only if $f^T x$ is a one-dimensional Gaussian measure with mean $f^T m$ and variance $f^T \Sigma f$.

3 Gaussian measures in vector spaces

Let X be a topological vector space (think of a Banach space – or a Fréchet space if you’re feeling adventurous). A Gaussian measure is a probability measure on X if and only if for each continuous linear functional $f \in X^*$ the push-forward measure $\gamma \circ f^{-1}$ is a Gaussian measure on \mathbb{R} .

We can define a mean and a covariance by setting

- $a_\gamma(f) = \int_X f(x) \gamma(dx)$, so $a_\gamma : X^* \rightarrow \mathbb{R}$, and hence $a_\gamma \in (X^*)'$.
- $B_\gamma(f, g) = \int_X [f(x) - a_\gamma(f)] \cdot [g(x) - a_\gamma(g)] \gamma(dx)$, so $B_\gamma \in (X^* \times X^*)'$.

So far we only know that a_γ is a linear map from X^* (the so-called topological dual) to \mathbb{R} , i.e. a_γ is an element of the “algebraical dual of the topological dual”, because we don’t know yet whether a_γ is continuous (i.e. bounded). The exponential Gaussian moments we obtain from Fernique’s theorem prove exactly this, and thus actually $a_\gamma \in X^{**}$. This is still kind of impractical: We can interpret $a_\gamma(f)$ as the mean of the measure “in direction of f ”, so we would like to view a_γ as the “general mean” of the measure. But we think of the mean as an object in the probability space (think of sampling from a measure; the empirical mean will lie in the same space X as the samples), but here we can only state $a_\gamma \in X^{**}$. Because $X \subsetneq X^{**}$, we need to do some work in order to identify a_γ with an actual element in X .

What’s the deal with $X \subsetneq X^{**}$?

Every $x \in X$ can also be interpreted as an object in X^{**} in the following sense: Recall that $\phi \in X^{**} = (X^*)^*$ takes an element of the dual space $f \in X^*$ and maps it linearly and continuously into \mathbb{R} . So how can we – given just an element $x \in X$ which is no map whatsoever – map $f \in X^*$ continuously into \mathbb{R} ? The following works: For $x \in X$ define $\phi_x \in X^{**}$ by $\phi_x(f) = f(x)$ with $f \in X^*$.

- This is linear in f : Take $f, g \in X^*$ and $\alpha \in \mathbb{R}$, then $\phi_x(\alpha f + g) = (\alpha f + g)(x) = \alpha f(x) + g(x) = \alpha \phi_x(f) + \phi_x(g)$.
- This is a continuous (i.e. bounded) mapping: $|\phi_x(f)| = |f(x)| \leq \|f\|_{X^*} \cdot \|x\|_X$ and thus $\|\phi\|_{X^{**}} \leq \|x\|_X$. We can actually show $\|\phi\|_{X^{**}} = \|x\|_X$, i.e. the mapping $J : X \rightarrow J(X) \subset X^{**}, x \mapsto \phi_x$ is an isometry!

In general, J is not surjective, i.e. $X \subsetneq X^{**}$. By common laziness, we like to identify $x \in X$ with $J(x) = \phi_x \in X^{**}$ and just use the notation $x \in X^{**}$. This leads to confusing looking terms like $f(x) = x(f)$ for $f \in X^*$ (especially later when we talk about the map R_γ). We need to keep in mind how this is to interpreted.

When is an element $\phi \in X^{**}$ also in X ?

We established that X is a proper subset of X^{**} , so we can’t expect to be able to find an element $x_\phi \in X$ for any $\phi \in X^{**}$ such that $\phi(f) = f(x_\phi)$ for any $f \in X^*$ (note that we have reversed the notation as we are looking for x given ϕ). There is a powerful criterion when this is actually true:

Proposition 1. *Let $\phi : X^* \rightarrow \mathbb{R}$ be a linear functional, i.e. $\phi \in (X^*)'$. If ϕ is continuous for the weak* topology, there exists a $x_\phi \in X$ such that for all $f \in X^*$*

$$\phi(f) = f(x_\phi).$$

- Note that we start with an element $\phi \in (X^*)'$ instead of in $(X^*)^*$, i.e. we only claim linearity.
- Recall that in a separable Banach space it is enough to show sequential weak* continuity, i.e. if $\lim_{n \rightarrow \infty} \phi(f_n) = \phi(f)$ for every sequence $f_n \xrightarrow{*} f$, i.e. $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for every $x \in X$.
- Don't forget: weak* continuity is a stronger condition than (regular) continuity! So it is not enough to show that $\phi \in X^{**}$ to obtain an element x_ϕ as needed. This is obvious from the fact that this would show $X^{**} \subset X$ which is false in general (for non-reflexive spaces). Another way to see this is that weak* continuity means that ϕ is continuous with respect to all weakly converging sequences. As weak* convergence is a weaker criterion than strong convergence (don't mix this up with weak* continuity), there are more sequences that are weakly* convergent than strongly convergent and thus the functional ϕ has to behave even better to be continuous with this bigger set of critical sequences.

The main point to take away from this proposition is that any linear functional $\phi : X^* \rightarrow \mathbb{R}$ on a separable Banach space X has a corresponding $x_\phi \in X$ if it is sequentially weak*-continuous, which is just a fancy way of saying that for any pointwise converging sequence $(f_n) \subset X^*$, i.e. $f_n(x) \rightarrow f(x)$, we also have $\phi(f_n) \rightarrow \phi(f)$. **In all our applications here, ϕ will be an integral operator and we can show this criterion by integral convergence theorems like the dominated convergence theorem.**

A main theorem when working with Gaussian measures on Banach spaces is Fernique's theorem which provides bounds for all polynomial moments. The proof is not too difficult but quite technical and so we will not write it here.

Theorem 1. *Let γ be a Gaussian measure on a separable Banach space X . Then there is an $\alpha > 0$ such that*

$$\int_X \exp \{ \alpha \|x\|^2 \} \gamma(dx) < \infty.$$

Now we're ready to prove that¹ $a_\gamma \in X$.

Lemma 1. *Let X be a separable Banach space and γ a Gaussian measure on X . Then $a_\gamma \in X$.*

Proof. a_γ is an element of $(X^*)'$ (and actually X^{**} , but we won't need this explicitly). From proposition 1 we know that we only need to prove that a_γ is weak*-continuous. Separability of X means that sequential weak*-continuity is actually enough. Hence we need to prove that for any $f_n \xrightarrow{*} f$, i.e. $f_n(x) \rightarrow f(x)$ for any $x \in X$,

$$a_\gamma(f_n) \rightarrow a_\gamma(f).$$

As $X \subsetneq X^{**}$, this is not true in general, but the special structure of a_γ as a Lebesgue integral functional (and hence we can use the Lebesgue dominated convergence theorem) and the good integrability properties of Gaussian measures – courtesy of Fernique's theorem – help us out. First we realize that any weak*-convergent sequence is bounded, hence $\sup_n \|f_n\|_{X^*} \leq C$. Then pointwise convergence of f_n and

$$|a_\gamma(f_n)| \leq \int_X \|f_n\|_{X^*} \|x\| \gamma(dx) \leq C \int_X \|x\| \gamma(dx) < C'$$

where the last bound follows from Theorem 1. Then Lebesgue's dominated convergence theorem immediately yields the result. \square

¹or rather that for a given γ , there is an element $a \in X$ such that $a_\gamma \in X^{**}$ and $a \in X$ correspond in the sense $a_\gamma(f) = f(a)$.

Example ①: Mean and Covariance in finite dimensions

Take a multivariate Gaussian in d dimensions, i.e. $\gamma = N(m, \Sigma)$ with $m \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$. Then linear functionals $f \in X^*$ can be interpreted as \mathbb{R}^d vectors, with $f(x) = f^T x$. Then

$$a_\gamma(f) = \int_{\mathbb{R}^d} f^T x \gamma(dx) = f^T m.$$

By the duality voodoo explained above, although $a_\gamma \in X^{**}$ (technically), we can say $a_\gamma \in X$ by identifying $a_\gamma(f) = f(a_\gamma) = f^T a_\gamma$, so actually $a_\gamma = m$ as expected and $a_\gamma(f) = f^T m$. One has to be careful not to mix up a_γ (the mean of the Gaussian) and $a_\gamma(f)$ (the mean of the Gaussian in direction f). This is even more true in the general Banach space setting.

The covariance form is

$$B_\gamma(f, g) = \int_{\mathbb{R}^d} [f(x) - a_\gamma(f)] \cdot [g(x) - a_\gamma(g)] \gamma(dx) = \int_{\mathbb{R}^d} f^T (x - m) \cdot g^T (x - m) \gamma(dx) = f^T \Sigma g$$

Example ②: Mean and Covariance for an infinite product of standard Gaussians

Take the Gaussian $\gamma = \bigotimes_{j=1}^{\infty} N(0, 1)$, i.e. a measure on the sequence space $\mathbb{R}^\infty = \{(a_j)_{j \in \mathbb{N}} : a_j \in \mathbb{R}\}$. This is no Banach space but a Fréchet space (visit the appendix for this). A continuous linear functional is of the form $f \in X^* = c_{00}$, i.e. f can be written as a real sequence that is eventually zero: $f = (f_1, f_2, \dots, f_N, 0, 0, \dots)$. Then, $f(x) = \sum_{j=1}^N f_j x_j$ where N depends on the specific f used. An arbitrary sequence (being not eventually zero) is not a continuous linear functional! This statement is proven in lemma 11. Mean and covariance form (set $f, g \in X^* = c_{00}$) are

$$a_\gamma(f) = 0 \in \mathbb{R}^\infty$$

and

$$B_\gamma(f, g) = \sum_{j \in \mathbb{N}} f_j g_j$$

where the sum is actually finite.

Example ③: Mean and Covariance for a degenerate Gaussian in 2-d

An important example is $\gamma = \delta_p \otimes N(q, \sigma^2)$, the product of a one-dimensional Dirac measure (or a degenerate one-dimensional Gaussian) centred at $p \in \mathbb{R}$ and a one-dimensional Gaussian centred at $q \in \mathbb{R}$ with variance σ^2 . This is a (degenerate) Gaussian measure on $X = \mathbb{R}^2$. Linear continuous functionals $f \in X^*$ can be written as vectors in \mathbb{R}^2 via the usual correspondence $f(x) = f^T x$, i.e. $X^* = \mathbb{R}^2$. The mean and covariance are computed as follows:

$$a_\gamma(f) = \int_{\mathbb{R}} \int_{\mathbb{R}} (f_1 x_1 + f_2 x_2) \delta_p(dx_1) N(q, \sigma^2)(dx_2) = f_1 \cdot p + f_2 \cdot q = f^T \begin{pmatrix} p \\ q \end{pmatrix}$$

for $f \in X^* = \mathbb{R}^2$ or more succinctly $a_\gamma = (p, q)^T$. With another $g \in X^* = \mathbb{R}^2$

$$B_\gamma(f, g) = \int_{\mathbb{R}} \int_{\mathbb{R}} (f_1, f_2) \begin{pmatrix} x_1 - p \\ x_2 - q \end{pmatrix} \cdot (g_1, g_2) \begin{pmatrix} x_1 - p \\ x_2 - q \end{pmatrix} \delta_p(dx_1) N(m, \sigma^2)(dx_2) = f_2 g_2 \sigma^2$$

Note that this is a special case of Example ①, if we set $m = (p, q)^T$ and $\Sigma = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix}$.

3.1 Creating a Hilbert space structure on linear functionals from scratch

We are working in a Banach space, so it is a small miracle that we are able to conjure a Hilbert space structure (but at first only in the realm of functionals X^*) out of thin air. Later we will construct an isometry from this Hilbert space into a strict subset of X . This Hilbert subspace is the Cameron-Martin space and has a set of very useful properties. But let's start at the beginning:

It is a fundamental fact that every bounded linear functional has second moments with respect to a Gaussian measure, i.e. $X^* \subset L^2(X, \gamma)$. Even more, the embedding

$$\begin{aligned} j : X^* &\hookrightarrow L^2(X, \gamma) \\ f &\mapsto f - a_\gamma(f) \end{aligned}$$

is continuous. The image of j is not closed in the L^2 topology, so we define:

$$X_\gamma^* = \overline{j(X^*)}^{L^2(X, \gamma)} \text{ is called the reproducing kernel Hilbert space }^2,$$

i.e. X_γ^* consists of all L^2 -limits of sequences of the form $(f_n - a_\gamma(f_n))_n$, with $f_n \in X^*$. **[Characterize X_γ^* further: All Gaussian random variables?]**

This is a definition one easily glances over but there are a few subtleties here that should be pointed out:

- Although each $f \in X^*$ is also a $f \in L^2(X, \gamma)$ (and by extension, $f - a_\gamma(f) \in X_\gamma^*$), there is quite a difference in viewing f as a linear functional in X^* or a square-integrable random variable: The main point is that two “directions” f_1, f_2 that differ only by a component that is degenerate (i.e. the measure puts no mass along this axis) are *different* objects in X^* , but identical as elements in $L^2(X, \gamma)$. Take for example the degenerate Gaussian measure $\delta_0 \otimes N(0, 1)$ in two dimensions: Here, $f_1(x) = x_1 + x_2$ and $f_2(x) = x_2$ are definitely not the same functional, but as random variables they are identical, because $x_1 = 0$ a.s. This amounts to the fact that $\|f_1 - f_2\|_{X^*} = 1$ but $\|f_1 - f_2\|_{L^2(X, \gamma)} = 0$.
- While $f \in X^*$ is a linear functional (i.e. specifically $f(0) = 0$), $j(f)$ isn't anymore, but is rather an affine functional because its mean has been subtracted. On the other hand, $j(f)$ has zero mean, while this is generally not true for $f \in X^*$. This reasoning holds identically for elements $g \in X_\gamma^*$: By definition, $a_\gamma(g) = 0$, but $g(0) \neq 0$ in general. One has to be careful not to mix up those two ideas of “going through the origin”.
- By using the covariance inner product $\langle f, g \rangle_{L^2(X, \gamma)} = B_\gamma(f, g) = \int_X [f(x) - a_\gamma(f)] \cdot [g(x) - a_\gamma(g)] \gamma(dx)$, we obtain a Hilbert space structure in the image of j . This will be of great use later.

Example ①: X_γ^* in finite dimensions

For a multivariate Gaussian $N(m, \Sigma)$ with $m \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$, we have $X = \mathbb{R}^d$ and thus $X^* = \mathbb{R}^d$ in the sense that for $f \in X^*$, the evaluation is scalar multiplication, i.e. $f(x) = f^T x$.

Mean and covariance are $a_\gamma(f) = f^T m$ and $B_\gamma(f, g) = f^T \Sigma g$ for $f, g \in X^*$. This means that X_γ^* consists of objects of the form $j(f) = f - a_\gamma(f)$, hence $j(f) = f(x) - a_\gamma(f) = f^T(x - m)$. In finite dimensions we don't need the L^2 completion, every element of X_γ^* is actually in this form.

Example ②: X_γ^* for an infinite product of standard Gaussians

Let $\gamma = \bigotimes_{j \in \mathbb{N}} N(0, 1)$. As linear functionals are represented by sequences “eventually zero”, i.e. $f \in X^* = c_{00}$ and the mean $a_\gamma = 0 \in \mathbb{R}^\infty$, we don't need to shift any f . The L^2 closure in the definition of X_γ^* is more interesting, though: We first note that $\langle f, g \rangle_{L^2(X, \gamma)} = B_\gamma(f, g) = \sum_{j \in \mathbb{N}} f_j \cdot g_j = \langle f, g \rangle_{l^2}$, so the $L^2(X, \gamma)$ inner product is actually the l^2 inner product. As $\overline{c_{00}}^{l^2} = l^2$ (see lemma 16), the reconstructing kernel Hilbert space X_γ^* is the space of square-summable sequences.

²for reasons explained in remark 2

Example ③: X_γ^* for a degenerate Gaussian in 2-d

For, $\gamma = \delta_p \otimes N(q, \sigma^2)$ we established that $X^* = \mathbb{R}^2$. The space X_γ^* now consists of linear functionals which are shifted by their mean, i.e. as in Example ① above if we set $m = (p, q)^T$ and $\Sigma = \text{diag}(0, \sigma^2)$: For any $f \in \mathbb{R}^2$, $j(f) = f_1(x_1 - p) + f_2(x_2 - q)$ and $X_\gamma^* = j(X^*) \subset L^2(X, \gamma)$ (we don't need the L^2 completion in finite dimensions). Note that $j(X^*)$ looks 2-dimensional (as we have two parameters, f_1 and f_2) but this is spurious: Any two f, g that only differ by their first component, i.e. $f - g = (c, 0)$ have $j(f)$ and $j(g)$ lying in the same L^2 equivalence class, i.e. $j(f) - j(g) = j(f - g) = 0 \in L^2(X, \gamma)$.

By L^2 continuity, a_γ , $\hat{\gamma}$ and B_γ can be extended with their domains changed: For any X^* we can also use elements from X_γ^* , while for $g \in X_\gamma^*$, $a_\gamma(g) = 0$, $\hat{\gamma}(f) = \exp\{-\frac{1}{2}\|f\|_{L^2(X, \gamma)}^2\}$ and $B_\gamma(g_1, g_2) = \int_X g_1(x)g_2(x)\gamma(dx)$ for $g_1, g_2 \in X_\gamma^*$. Note that X_γ^* is not a superset of X^* because its elements all need to be centred (as random variables, i.e. $a_\gamma(f) = 0$ for $f \in X_\gamma^*$).

4 Transporting the Hilbert space structure into the probability space

So far, we only have a Hilbert space structure in the domain of linear functionals $f \in X^*$ (strictly speaking: on X_γ^*). We define a functional which will map X_γ^* into X . This functional will actually be an isometry, so we can transport the Hilbert space structure into a subspace of X (this will be the Cameron-Martin space).

We set

$$R_\gamma : X_\gamma^* \rightarrow (X^*)'$$

$$R_\gamma f(g) = \int_X f(x)[g(x) - a_\gamma(g)]\gamma(dx)$$

Remark 1. A few notes on the definition of R_γ :

- Note that $R_\gamma f \in (X^*)'$, but we want it to be in X . This will work along the same lines as the proof that $a_\gamma \in X$ above. We record this in lemma 2
- This functional is very closely related to the covariance inner product B_γ (which is the $L^2(X, \gamma)$ inner product). Actually, $R_\gamma f_1(f_2) = B_\gamma(f_1, f_2)$ for $f_1, f_2 \in X_\gamma^*$ (or $R_\gamma f|_{X_\gamma^*} = B_\gamma(f, \cdot)|_{X_\gamma^*}$). The map R_γ can be thought of as an asymmetric version of B_γ where we straddle “input” from X_γ^* and X^* .
- R_γ needs two inputs (i.e. $f \in X_\gamma^*$ and $g \in X^*$) to be actually explicitly computable, but we will consider the map only with the argument f , i.e. as it is stated as a map $X_\gamma^* \rightarrow (X^*)'$ (where the parameter g is “invisible”). We only care about what R_γ does with an $f \in X_\gamma^*$.
- We don't need to subtract the mean of f (as it is done with g) because f has already mean 0 just by being in X_γ^* .

Lemma 2. R_γ can be represented as a map into X , i.e. for any $f \in X_\gamma^*$ there is an element $r \in X$ such that for all $g \in X^*$

$$R_\gamma f(g) = g(r)$$

or (for brevity)

$$R_\gamma f(g) = g(R_\gamma f).$$

Proof. This is proven almost identically to lemma 1. □

As R_γ can be interpreted as a map $X_\gamma^* \rightarrow X$, we can identify $R_\gamma f$ with an element $r_f \in X$ such that

$$\int_X f(x)[g(x) - a_\gamma(g)]\gamma(dx) = R_\gamma f(g) = g(r_f), \quad (1)$$

but for simplicity we will more sloppily identify $R_\gamma f$ with r_f and write

$$R_\gamma f(g) = g(R_\gamma f).$$

The most important thing of this section is at its end: We define the Cameron-Martin space³ H as the image of R_γ . We put this in a displayed equation to stress its importance.

$$H = \text{ran } R_\gamma$$

R_γ is not onto, so $H \subsetneq X$. There are several senses in which H is “small”: If γ is centred, $\gamma(H) = 0$, i.e. samples of γ will almost surely miss H also, the embedding $H \hookrightarrow X$ is compact.

There is an important equivalent characterization for the Cameron-Martin space H which many authors choose as its definition:

Proposition 2. Define $|h|_H = \sup\{f(h) : f \in X^*, \|j(f)\|_{L^2(X,\gamma)} \leq 1\}$ *[Rolle dieser Norm herausstellen: fuer $f \in X^*$ mit $\|f\|_{X^*}$ diese Einsetzungsnorm waere einfach die Norm von h aber es wird interessant, wenn wir den Bereich der f s ueber die L^2 -Norm steuern], the evaluation norm of $h \in X$. Then the following two statements are equivalent:*

1. $h \in H$, i.e. $h = R_\gamma \hat{h}$.
2. $|h|_H < \infty$.

Also, if $h \in H$ (and thus both conditions are true), $|h|_H = \|\hat{h}\|_{L^2(X,\gamma)}$. This means that $R_\gamma : X_\gamma^* \rightarrow H$ is an isometry, rendering H a Hilbert space with inner product $\langle h, k \rangle_H = \langle \hat{h}, \hat{k} \rangle_{L^2(X,\gamma)}$ for $h = R_\gamma \hat{h}, k = R_\gamma \hat{k}$.

This means that $H = \{h \in X : |h|_H < \infty\}$ (which is a definition often used in the literature).

Proof. We show that (1.) \Rightarrow (2.) and assume that $H \ni h = R_\gamma \hat{h}$ for some $\hat{h} \in X_\gamma^*$. We want to show that the H -norm is finite. Indeed, we recall that $h = R_\gamma \hat{h}$ is actually the short form of $f(h) = R_\gamma \hat{h}(f)$ for any $f \in X^*$ and thus

$$\begin{aligned} f(h) &= R_\gamma \hat{h}(f) = \int_X \hat{h}(x) \cdot [f(x) - a_\gamma(f)] \gamma(dx) = \int_X \hat{h}(x) \cdot j(f)(x) \gamma(dx) \\ &\leq \|\hat{h}\|_{L^2(X,\gamma)} \cdot \|j(f)\|_{L^2(X,\gamma)}, \end{aligned}$$

which means that

$$\begin{aligned} |h|_H &= \sup\{f(h) : f \in X^*, \|j(f)\|_{L^2(X,\gamma)} \leq 1\} \\ &\leq \|\hat{h}\|_{L^2(X,\gamma)} < \infty. \end{aligned}$$

For (2.) \Rightarrow (1.) we define for a given $h \in X$ with $|h|_H < \infty$ the evaluation functional

$$L_h : j(X^*) \rightarrow \mathbb{R}, j(f) \mapsto f(h).$$

The first issue we need to address is well-definedness: As j is not an one-to-one mapping, there might exist $f_1, f_2 \in X^*$ with $j(f_1) = j(f_2)$ and possibly $f_1(h) \neq f_2(h)$. This is impossible, though, as for a given $f \in X^*$

$$|L_h(j(f))| = |f(h)| \leq |h|_H \cdot \|j(f)\|_{L^2(X,\gamma)} \quad (2)$$

and hence for a pair $f_1, f_2 \in X^*$ with $j(f_1) = j(f_2)$ (i.e. $j(f_1 - f_2) = 0$), we obtain $|f_1(h) - f_2(h)| = |(f_1 - f_2)(h)| \leq |h|_H \cdot \|j(f_1 - f_2)\|_{L^2(X,\gamma)} = 0$. Another consequence of (2) is $L^2(X, \gamma)$ -continuity and thus we can extend L_h

³DaPrato/Zabczyk call H the reconstructing kernel Hilbert space (RKHS), a notation we have reserved for X_γ^* (and which remains unnamed in DBZ)

continuously to a (continuous linear) map $L_h : X_\gamma^* \rightarrow \mathbb{R}$. As X_γ^* is a Hilbert space and $L_h \in (X_\gamma^*)^*$, Riesz' representation theorem shows the existence of an element $\hat{h} \in X_\gamma^*$ such that

$$L_h(\phi) = \langle \hat{h}, \phi \rangle_{L^2(X, \gamma)} = \int_X \hat{h}(x) \cdot \phi(x) \gamma(dx) \quad (3)$$

and in particular for any $f \in X^*$ (and thus $j(f) \in X_\gamma^*$):

$$f(h) = L_h(j(f)) = \int_X \hat{h}(x) \cdot j(f)(x) \gamma(dx) = R_\gamma \hat{h}(f) = f(R_\gamma \hat{h}) \quad (4)$$

where the last equality is to be understood in the usual duality sense. As the equation $f(h) = f(R_\gamma \hat{h})$ is true for any $f \in X^*$, we obtain

$$h = R_\gamma \hat{h}$$

and by definition of H , that $h \in H$.

It remains to show that $|h|_H = \|\hat{h}\|_{L^2(X, \gamma)}$ if any (and thus both) of the equivalent conditions hold. Indeed, for $h = R_\gamma \hat{h} \in H$, from (4)

$$\begin{aligned} \sup\{f(h) : f \in X^*, \|j(f)\|_{L^2(X, \gamma)} \leq 1\} &= \sup\left\{\langle \hat{h}, j(f) \rangle_{L^2(X, \gamma)} : f \in X^*, \|j(f)\|_{L^2(X, \gamma)} \leq 1\right\} \\ &= \sup\left\{\langle \hat{h}, \phi \rangle_{L^2(X, \gamma)} : \phi \in X_\gamma^*\right\} = \|\hat{h}\|_{L^2(X, \gamma)} \end{aligned}$$

where the step from the first to the second line is valid as $j(X^*)$ is dense in X_γ^* . \square

Remark 2. Note that (3) identifies the evaluation map L_h with the element $\hat{h} \in X_\gamma^*$. This is the reason for X_γ^* being called “reconstructing kernel Hilbert space”: The element \hat{h} is the (integral) kernel which reconstructs the value of f in h :

$$f(h) = \langle \hat{h}, f \rangle_{L^2(X, \gamma)}.$$

It is truly remarkable that we are able to evaluate f pointwise by alternatively integrate f over the kernel \hat{h} .

Reconstructing kernel Hilbert spaces

This exposition is borrowed from [4].

Definition 1. A vector subspace of functions $\mathcal{H} = \{f : D \rightarrow \mathbb{R}\}$ on some domain D is called a reconstructing kernel Hilbert space (RKHS) if

- \mathcal{H} is a Hilbert space with some inner product $\langle \cdot, \cdot \rangle$ and
- For any $y \in D$, the evaluation functional $L_y : \mathcal{H} \rightarrow \mathbb{R}$, $f \mapsto f(y)$ is bounded.

Remark 3. Because \mathcal{H} needs to be a vector space of functions, any Lebesgue space cannot be an RKHS (because L^p functions are actually equivalence classes of functions). This makes sense as pointwise evaluation is not well-defined for those functions.

As every linear, bounded functional corresponds to an element in the Hilbert space, there is a $k_y \in \mathcal{H}$ such that $L_y(f) = \langle k_y, f \rangle$. As $k_y(x) = \langle k_x, k_y \rangle = \langle k_y, k_x \rangle = k_x(y)$, this constitutes a symmetric kernel with a remarkable property:

Definition 2. Given an RKHS \mathcal{H} , the symmetric kernel $K : D \times D \rightarrow \mathbb{R}$, $K(x, y) = k_y(x)$ is called the *reconstructing kernel*. In this notation, it holds that for any $f \in \mathcal{H}$,

$$f(y) = \langle K(\cdot, y), f \rangle.$$

H_0^1 as an RKHS: Consider $H_0^1 = \{f \in L^2([0, 1]) : Df \in L^2([0, 1]), f(0) = f(1) = 0\}$ where Df is the weak derivative^a. This is an RKHS with kernel

$$K(x, y) = \begin{cases} x(1-y) & \text{for } x \leq y \\ (1-x)y & \text{for } x > y \end{cases}.$$

By standard theory, H_0^1 is a Hilbert space with inner product $\langle f, g \rangle = \int_0^1 f'(t)g'(t)dt$. Evaluation is bounded, as

$$|f(y)| = \left| \int_0^y f'(t)dt \right| = \left| \int_0^1 f'(t)\chi_{[0,y]}(t)dt \right| \leq \langle f, f \rangle.$$

Thus, H_0^1 is an RKHS with the reconstruction property. This is also easily checked:

$$\langle f, K(y, \cdot) \rangle = \int_0^y f'(x)(1-y)dx + \int_y^1 f'(x)(-y)dx = f(y).$$

The space of band-limited functions as an RKHS With^b the Fourier transform $\hat{f}(\omega) = \int f(x)e^{-i\omega x}dx$ of some $f \in \{f \in L^2(\mathbb{R})\}$, the vector space $\mathcal{H} =: \{f \in L^2(\mathbb{R}) : \text{supp}(\hat{f}) \subset [-a, a]\}$ is an RKHS with kernel $K(x, y) = \frac{\sin(a[y-x])}{\pi[y-x]}$. That \mathcal{H} with the L^2 inner product is an RKHS can be seen from

$$|f(y)| = \frac{1}{2\pi} \int_{-a}^a \hat{f}(\omega)e^{i\omega y}d\omega \leq \sqrt{\frac{a}{\pi}} \cdot \|\hat{f}\|_{L^2(\mathbb{R})} = \sqrt{\frac{a}{\pi}} \cdot \|f\|_{L^2(\mathbb{R})}.$$

That K is actually the reconstructing kernel can be intuitively understood as K is the “bandlimited Dirac”. Indeed $\lim_{a \rightarrow \infty} K(x, y) = \delta_y(x)$ (in the sense of distributions), this makes sense as $f(y) = \int f(x)\delta_y(x)dx$, but the problem here is that the evaluation is not bounded for non-bandlimited functions, so we need $a < \infty$.

^aand the zero Dirichlet conditions are to be interpreted in the usual sense, i.e. $H_0^1 = \overline{C_c^\infty([0, 1])}^{H^1([0, 1])}$

^bfrom https://en.wikipedia.org/wiki/Reproducing_kernel_Hilbert_space

Example ①: R_γ in finite dimensions

For a multivariate Gaussian $N(m, \Sigma)$ with $m \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ invertible, the map R_γ is simply

$$\begin{aligned} R_\gamma : X_\gamma^* &\rightarrow \mathbb{R}^d \\ (x \mapsto f^T \cdot (x - m)) &\mapsto \Sigma \cdot f. \end{aligned}$$

In particular,

$$H = \text{ran } R_\gamma = \mathbb{R}^d$$

as Σ is invertible and thus surjective. For the form of elements in X_γ^* read the blue box on this topic in case you don't remember. **[Interpretation?]** The inverse R_γ^{-1} is then (in the case where Σ is invertible)

$$\begin{aligned} R_\gamma^{-1} : \mathbb{R}^d &\rightarrow X_\gamma^* \\ f &\mapsto (x \mapsto f^T \Sigma^{-1}(x - m)). \end{aligned}$$

That R_γ is actually an isometry can be seen elementarily here: Write $h = R_\gamma \hat{h}$. Note that $h \in \mathbb{R}^d$ but $\hat{h} = (x \mapsto f^T(x - m))$ for some f , so we can't interpret \hat{h} as a column vector! Then (note that $j(f) = f - f^T m$, i.e. $j(f)(x) = f^T(x - m)$),

$$\begin{aligned} |h|_H &= \sup\{f(h) : f \in \mathbb{R}^d, \|j(f)\|_{L^2(X_\gamma)} \leq 1\} = \sup\{f^T h : f \in \mathbb{R}^d, f^T \Sigma f \leq 1\} \\ &= \sup\{g^T \Sigma^{-1/2} h : |g| \leq 1\} = |\Sigma^{-1/2} h| = \sqrt{h^T \Sigma^{-1} h} \end{aligned}$$

And also

$$\begin{aligned}\|\hat{h}\|_{L^2(\mathbb{R}^d, \gamma)} &= \sqrt{\int_{\mathbb{R}^d} \hat{h}(x)^2 \gamma(dx)} = \sqrt{\int_{\mathbb{R}^d} (R_\gamma^{-1} h(x))^2 \gamma(dx)} = \sqrt{\int_{\mathbb{R}^d} (h^T \Sigma^{-1} (x - m))^2 \gamma(dx)} \\ &= \sqrt{\int_{\mathbb{R}^d} ([\Sigma^{-1} h]^T (x - m))^2 \gamma(dx)} = \sqrt{[\Sigma^{-1} h]^T \Sigma [\Sigma^{-1} h]} = \sqrt{h^T \Sigma^{-1} h},\end{aligned}$$

i.e.

$$|h|_H = \|\hat{h}\|_{L^2(\mathbb{R}^d, \gamma)}$$

Example ②: R_γ for an infinite product of standard Gaussians

A calculation shows that $R_\gamma f(g) = \int_X f(x)g(x)\gamma(dx) = \langle f, g \rangle_{l^2}$, thus we can identify $R_\gamma f = f$ and R_γ is the identity. Then the Cameron-Martin space is the image of R_γ , i.e. $H = R_\gamma X_\gamma^* = X_\gamma^* = l^2$, so the Cameron-Martin space is the space of square-summable sequences. It is even more remarkable in this case that we obtain a Hilbert space structure on H , which is a subspace of X , because X is not even a normed space! Here, $|h|_H = \|h\|_{l^2} = \|h\|_{L^2(X, \gamma)}$ (in this case R_γ is the identity, so $h = \hat{h}$).

Example ③: R_γ for a degenerate Gaussian in 2-d

The calculations are similar to Example ①, but we need to take care as Σ is not invertible. R_γ is the same:

$$\begin{aligned}R_\gamma : X_\gamma^* &\rightarrow \mathbb{R}^2 \\ (x \mapsto f^T(x - (p, q)^T)) &\mapsto (0, \sigma^2 f_2)^T\end{aligned}$$

The image of R_γ (the Cameron-Martin space) is $\{0\} \times \mathbb{R} = H$ and the inverse of R_γ is

$$\begin{aligned}R_\gamma^1 : H &\rightarrow X_\gamma^* \\ \begin{pmatrix} 0 \\ f_2 \end{pmatrix} &\mapsto \frac{f_1(x_1 - p) + f_2(x_2 - q)}{\sigma^2}\end{aligned}$$

where f_1 is arbitrary because any value $f_1 \in \mathbb{R}$ will yield the same element in X_γ^* because they will have the same value γ -almost-surely.

5 Shifts of Gaussian measures

The main point of the Cameron-Martin theory is dealing with shifts of Gaussian measures, i.e. given a Gaussian measure γ in some Banach space, we define γ_h by setting $\gamma_h(B) = \gamma(B - h)$ for sets $B \in \mathcal{B}(X)$. Our main goal is to obtain the density of γ_h with respect to γ , i.e. find the function $\rho_h : X \rightarrow \mathbb{R}$ such that

$$\gamma_h(B) = \int_B \rho_h(x) \gamma(dx),$$

or symbolically, $\frac{d\gamma_h}{d\gamma}(x) = \rho_h(x)$. There are two difficulties with this “relative density”: The first problem arises when we try to shift in directions in which the measure γ is degenerate. In this case, we transport our probability mass to an area which had zero probability mass, and thus formally $\frac{d\gamma_h}{d\gamma}(x) = \frac{C}{0}$ and we cannot expect to get a density. The second issue only arises in infinite dimension and is most intuitively explained with looking at the measure $\otimes_{n=1}^J N(0, 1)$ with $J \rightarrow \infty$. We start with our non-degenerate example in finite dimensions as this is illustrative for the general theory:

Example ①: Gaussian Shifts in finite dimensions

We take a non-degenerate Gaussian in d dimensions, i.e. $\gamma = N(m, \Sigma)$ with $\Sigma \in \mathbb{R}^{d \times d}$ non-singular. Define γ_h as $\gamma_h(\cdot) = \gamma(\cdot - h)$ and then (by elementary multivariate Gaussian wizardry or the help of a Computer algebra system)

$$\frac{d\gamma_h}{d\gamma}(x) = \exp \left\{ h^T \Sigma^{-1}(x - m) - \frac{1}{2} h^T \Sigma^{-1} h \right\} = \exp \left\{ (R_\gamma^{-1}(h))(x) - \frac{1}{2} \|h\|_H^2 \right\},$$

It turns out that this form is valid even in infinite dimensions, whenever $|h|_H$ is finite.

The following lemma is useful:

Lemma 3 ([3, Lemma 3.1.4]). *Take a Gaussian measure γ and a $g \in X_\gamma^*$. Recall that the characteristic function of γ is $\hat{\gamma}(f) = \exp(ia_\gamma(f) - \|j(f)\|_{L^2(X, \gamma)}^2/2)$, signifying that a_γ is the mean and $B_\gamma(f, g) = \int_X j(f)(x)j(g)(x)\gamma(dx)$ is the covariance form. Then the measure given by*

$$\frac{d\mu_g}{d\gamma}(x) = \exp \left\{ g(x) - \frac{1}{2} \|g\|_{L^2(X, \gamma)}^2 \right\}$$

is a Gaussian probability measure with characteristic function

$$\hat{\mu}_g(f) = \exp \left\{ i[a_\gamma(f) + R_\gamma g(f)] - \frac{1}{2} \|j(f)\|_{L^2(X, \gamma)}^2 \right\},$$

i.e. is the Gaussian measure with the same covariance structure as γ but with mean shifted by $R_\gamma g$. We can think of the mean as the object $a_\gamma + R_\gamma g$.

Sketch of proof. We show that the characteristic function of μ_g has the form as stated. We do the following formal calculation for $f \in X^*$:

$$\begin{aligned} \hat{\mu}_g(f) &= \int_X e^{if(x)} \mu_g(dx) = \int_X e^{if(x)+g(x)} \mu_g(dx) \cdot e^{-\frac{1}{2} \|g\|_{L^2(X, \gamma)}^2} = e^{-\frac{1}{2} \|g\|_{L^2(X, \gamma)}^2} \cdot \int_X e^{i[f(x)-ig(x)]} \gamma(dx) \\ &= e^{-\frac{1}{2} \|g\|_{L^2(X, \gamma)}^2} \cdot \hat{\gamma}(f - ig) = \exp \left(-\frac{1}{2} \|g\|_{L^2(X, \gamma)}^2 \right) \cdot \exp \left(ia_\gamma(f - ig) - \frac{1}{2} B_\gamma(f - ig, f - ig) \right) \end{aligned}$$

Now note that $a_\gamma(g) = 0$, $B_\gamma(f, g) = R_\gamma g(f)$, $B_\gamma(g, g) = \|g\|_{L^2(X, \gamma)}^2$ and $B_\gamma(f, f) = \|j(f)\|_{L^2(X, \gamma)}^2$, hence

$$\begin{aligned} &= \exp \left(-\frac{1}{2} \|g\|_{L^2(X, \gamma)}^2 \right) \cdot \exp \left(ia_\gamma(f) - \frac{1}{2} \|j(f)\|_{L^2(X, \gamma)}^2 + iR_\gamma g(f) + \frac{1}{2} \|g\|_{L^2(X, \gamma)}^2 \right) \\ &= \exp \left\{ i[a_\gamma(f) + R_\gamma g(f)] - \frac{1}{2} \|j(f)\|_{L^2(X, \gamma)}^2 \right\} \end{aligned}$$

Note that this calculation is not actually valid without further justification as in the step from first to second line we evaluated the characteristic function $\hat{\gamma}$ in the linear continuous functional $f - ig$ which is not real (But we defined the functional only for real linear continuous functionals). This can be done rigorously by a few more lines of arguments from complex analysis, as done in the proof in [3], but the formal calculation yields enough insight for our purpose. \square

Remark 4. • Lemma 3 is *almost* what we want: We obtain a density of a shifted Gaussian with respect to its unshifted reference measure, but we cannot choose the shift directly: For a given $g \in X_\gamma^*$ we obtain a shift of $R_\gamma g$ in the mean. We need to reverse this in order to choose any shift we want.

- We can already see here why not every shift will be feasible: With lemma 3 we can only acquire shifts which are in the range of R_γ – this is exactly the Cameron-Martin space. This means that we can only take shifts $h \in H$ from the Cameron-Martin space if we want to get a measure which – shifted by h – are absolutely continuous w.r.t. their unshifted reference measure (i.e. we have a density). The Cameron-Martin theorem collects all these statements.

Theorem 2 ([3, Theorem 3.1.5]). Let γ be a Gaussian measure on a Banach space X . For $h \in X$, define the shifted measure γ_h by $\gamma_h(\cdot) = \gamma(\cdot - h)$.

- If $h \in H$, then the measure γ_h is equivalent to γ and $\frac{d\gamma_h}{d\gamma}(x) = \rho_h(x)$ with

$$\rho_h(x) = \exp \left\{ R_\gamma^{-1}h(x) - \frac{1}{2}|h|_H^2 \right\}.$$

- If $h \notin H$, then $\gamma_h \perp \gamma$ (the measures are singular with respect to each other).

This means that two Gaussian measures differing only by a shift in the domain are either equivalent (if the shift is in the Cameron-Martin space) or singular. There is no middle ground as in the case of the Lebesgue measure and the Lebesgue measure on $[0, 1]$, where the latter is absolutely continuous w.r.t. the other but not the other way around.

Remark 5. When comparing lemma 3 and 2, $g = R_\gamma^{-1}h$ and equivalently $R_\gamma g = h$, i.e. we see that we can only realize shifts h which are in the range of R_γ , i.e. the Cameron-Martin space. Note that $|h|_H = \|R_\gamma^{-1}h\|_{L^2(X, \gamma)} = \|g\|_{L^2(X, \gamma)}$. The proof relies on lemma 3 for the first statement. To show singularity in the case $h \notin H$, it uses the notion of Hellinger distance (which is a distance between measures and is not covered here) to show that γ and γ_h are “too far away from each other” to have any joint support.

Example ①: Gaussian Shifts in finite dimensions (cont'd)

We already saw that the formula we derived by elementary calculation matches the formula for Theorem 2 in the “equivalence” case. In our case, Σ is invertible, and thus $H = \mathbb{R}^d$, which makes sense: If we shift a non-degenerate multivariate Gaussian, we will obtain an equivalent measure. So we can’t obtain a singular measure by shifting. The next two examples show two reasons why equivalence might *not* work out.

Example ②: Gaussian Shifts for an infinite product of standard Gaussians

We derived before that $H = l^2$, i.e. we can only shift the mean by an element of l^2 . In particular, for $h = (1, 1, \dots)$, $\gamma = \bigotimes N(0, 1)$ and $\gamma_h = \bigotimes N(1, 1)$ are singular with respect to each other (although the shift is not “that big”! For a given feasible shift $h \in l^2$, i.e. $\gamma_h = \bigotimes_{j \in \mathbb{N}} N(h_j, 1)$, the Cameron-Martin theorem yields

$$\frac{d\gamma_h}{d\gamma}(x) = \exp \left\{ \langle h, x \rangle_{l^2} - \frac{1}{2}\|h\|_{l^2}^2 \right\}.$$

Another way to see where the problem comes in when having infinite dimensions is the following: Let’s change our measure to $\gamma = \bigotimes_{j=1}^J N(0, 1)$, i.e. we don’t have infinitely many Gaussians but only “very many”. Then basic intuition (and our finite-dimensional considerations) tell us that for any shift $h \in \mathbb{R}^J$, the measures γ and γ_h should be equivalent. This is indeed the case and the Cameron-Martin formula reduces to

$$\frac{d\gamma_h}{d\gamma}(x) = \exp \left\{ \sum_{j=1}^J h_j x_j - \frac{1}{2} \sum_{k=1}^J h_k^2 \right\}.$$

This is a well-defined term for any fixed J (and thus the density exists), but in the limit $J \rightarrow \infty$, the sum $\sum_{k=1}^J h_k^2$ becomes the l^2 -norm of h and we have to be careful which shift h we choose.

Gaussian shifts for a degenerate Gaussian in 2-d

We know that $H = \{0\} \times \mathbb{R}$, so we can only shift in the second component. This makes sense as the complete probability mass of γ rests on the axis $x_1 = p$, so we can’t move it in x_1 -direction without losing equivalence. The Cameron-Martin formula states that for $h = (h_1, 0)^T \in H$ (recall that $R_\gamma^{-1}(0, h_2)^T = (h_1(x_1 - p) + h_2(x_2 - q))/\sigma^2$)

for arbitrary $h_1 \in \mathbb{R}$,

$$\frac{d\gamma_h}{d\gamma}(x) = \exp\left\{\frac{h_1(x_1 - p)}{\sigma^2}\right\} \cdot \exp\left\{\frac{h_2(x_2 - q)}{\sigma^2} - \frac{h_2^2}{2\sigma^2}\right\}$$

Is the multiplicative factor with arbitrary h_1 a contradiction? No: For $x_1 = p$ this factor is the identity and on $x_1 \neq p$ the measure γ is not supported. We can try to put as much value on the density of γ_h w.r.t. γ , but by courtesy of $C \cdot 0 = 0$ we will obtain the same result.

For simplicity it makes sense to choose $h_1 = 0$ and thus we write

$$\frac{d\gamma_h}{d\gamma}(x) = \exp\left\{\frac{h_2(x_2 - q)}{\sigma^2} - \frac{h_2^2}{2\sigma^2}\right\}$$

6 Interesting infinite-dimensional examples

So far, the infinite product of i.i.d Gaussians was our only infinite-dimensional example. In this section, we will look at some Banach-space-valued examples you are more likely to meet in the wild.

6.1 Gaussian measures in Hilbert space

We record a few facts about operators in Hilbert spaces which we will need. Consult [3] and [1] for proofs and details. Let X be an infinite-dimensional separable Hilbert space with norm $\|\cdot\|$ and inner product $\langle \cdot, \cdot \rangle$. We identify X^* with X via Riesz' representation theorem. We call $\mathcal{L}(X)$ the set of linear and bounded operators $X \rightarrow X$.

- $L \in \mathcal{L}(X)$ is nonnegative iff $\langle Lx, x \rangle \geq 0$ for all $x \in X$.
- $L \in \mathcal{L}(X)$ is called compact iff L is the limit of a sequence of finite rank operators in the operator norm.
- $L \in \mathcal{L}(X)$ is called self-adjoint iff $\langle Lx, y \rangle = \langle x, Ly \rangle$ for any $x, y \in X$.
- A nonnegative, adjoint operator is called *nuclear* or *of trace-class* iff there is an orthonormal basis $\{e_k : k \in \mathbb{N}\}$ of X such that

$$\text{trace}(L) := \sum_{k=1}^{\infty} \langle Le_k, e_k \rangle < \infty.$$

- Every nuclear operator is compact.
- Every operator $L \in \mathcal{L}(X)$ that can be written in the form

$$Lx = \sum_{k=1}^{\infty} \alpha_k \langle x, f_k \rangle f_k$$

for some orthonormal basis $\{f_k : k \in \mathbb{N}\}$ of X is compact.

- For every compact and self-adjoint operator $L \in \mathcal{L}(X)$ there exists an orthonormal basis of X given by $\{e_k : k \in \mathbb{N}\}$ consisting of eigenvectors of L such that

$$Lx = \sum_{k=1}^{\infty} \lambda_k \langle x, e_k \rangle e_k, \tag{5}$$

where $Le_k = \lambda_k e_k$ for $\lambda_k \in \mathbb{R}$. If L is non-negative, all λ_k are non-negative.

On the form of a_γ and B_γ in the Hilbert space setting: Due to Riesz' representation theorem, we have a more direct way of dealing with a_γ and B_γ : Take a Gaussian measure in a Hilbert space (but regard it as a Banach space). Then continuity of $a_\gamma : X^* \rightarrow \mathbb{R}$ and $B_\gamma : X^* \times X^* \rightarrow \mathbb{R}$ suffices (and we don't need weak- $*$ -continuity in this case) that there is an element $a \in X$ such that $a_\gamma(f) = \langle f, a \rangle$ and there is a self-adjoint operator $Q \in \mathcal{L}(X)$ such that $B_\gamma(f, g) = \langle Qf, g \rangle$, hence

$$\langle Qf, g \rangle = \int_X \langle f, x - a \rangle \langle g, x - y \rangle \gamma(dx).$$

We call Q the covariance operator. The measure's characteristic function is

$$\hat{\gamma}(f) = \exp \left\{ i \langle f, a \rangle - \frac{1}{2} \langle Qf, f \rangle \right\} \quad (6)$$

The next theorem states exactly which form Gaussian measures in Hilbert spaces can have.

Theorem 3. 1. If γ is a Gaussian measure on X , then its characteristic function is given by (6) for some $a \in X$ and a self-adjoint, non-negative trace-class operator $Q \in \mathcal{L}(X)$.

2. For any $a \in X$ and a self-adjoint, non-negative trace-class operator $Q \in \mathcal{L}(X)$, the characteristic function in (6) constitutes a Gaussian measure on X .

Remark 6. This theorem and (5) shows that Gaussian measures γ in infinite-dimensional Hilbert spaces are actually very simple: Given the measure's mean a and covariance operator Q , we have a basis of orthonormal eigenfunctions, such that $Qe_k = \lambda_k e_k$, with $\lambda_k \geq 0$, over which the measure diagonalizes. Then $a = \sum_{k=1}^{\infty} a_k \cdot e_k$ and we can say that

$$\gamma = \bigotimes_{k=1}^{\infty} N(a_k, \lambda_k),$$

i.e. a product of infinitely many independent Gaussians with mean a_k and variance⁴ λ_k . The correspondence here is that the k -th factor on the right hand side belongs to direction e_k and equivalently: A Gaussian random variable $x \in X$ is spanned as follows:

$$x = \sum_{k=1}^{\infty} (a_k + \eta_k \cdot \sqrt{\lambda_k}) \cdot e_k, \quad \eta_k \sim N(0, 1) \text{ i.i.d.} \quad (7)$$

Now we are ready to understand what the Cameron-Martin space of a Hilbert-space-valued Gaussian measure looks like: Linear continuous functionals $f \in X^* = X$ are just elements of the Hilbert space with $f(x) = \langle f, x \rangle$ for any $x \in X$, and mean and covariance are given by $a \in X$ and $Q \in \mathcal{L}(X)$ as argued above. Then $j(f) = f - a$ [Das ist vielleicht nicht richtig? Korrigieren, weitermachen.]

6.2 Gaussian processes

[steal from Nate]

7 The Feldman-Hajek theorem

This section is mostly interesting for readers who want to study Bayesian inverse problems.

⁴be careful with squares here: usually we say that σ^2 is a variance, i.e. $\sqrt{\lambda_k}$ is the standard deviation here

8 Fréchet spaces

Let X be a topological vector space with topology \mathcal{T} . A few preliminary definitions:

Definition 3. A set C is called

- balanced if for all $x \in C$ and all $|\lambda| \leq 1$ also $\lambda x \in C$.
- convex if for all $x, y \in C$ and all $0 \leq t \leq 1$, also $tx + (1 - t)y \in C$.
- absorbent if $\bigcup_{t>0} tC = X$.
- absolutely convex, if C is both convex and balanced. This is the same as saying that for every $x, y \in C$ and real numbers $|t| + |s| = 1$, $tx + sy \in C$. The best image for this is that C contains every parallelogram with vertices x and y symmetric to the origin.

Definition 4. • A collection of sets $\mathcal{B} \subset \mathcal{T}$ is called a base of the topology \mathcal{T} if for every open set $U \in \mathcal{T}$ there is a $B \in \mathcal{B}$ with $B \subset U$.

- A collection of sets $\mathcal{B}_x \subset \mathcal{T}$ is called a local base at x of the topology \mathcal{T} if for every neighborhood $U \in \mathcal{T}$ of x (i.e. $x \in U$) there is a $B \in \mathcal{B}_x$ with $B \subset U$.

We are going to compare two definitions of locally convex spaces.

Definition 5. A topological vector space X is called *locally convex* if the origin has a local base of absolutely convex absorbent sets.

Definition 6. A topological vector space X is called *locally convex* if there is a family of seminorms generating the topology.

Lemma 4. *Definitions 5 and 6 are equivalent.*

Proof. *Proof of 5 \Rightarrow 6.* Define the Minkowski gauge of a set $C \in X$ as

$$\mu_C(x) = \inf\{\lambda > 0 : x \in \lambda C\}.$$

Then we can show that the origin's local base of absolutely convex absorbent sets $\mathcal{B}_0 = \{B_a, a \in I\}$ constitute a family of seminorms via their Minkowski gauges $\{\mu_{B_a}, a \in I\}$ and the latter generates the topology of X . XXX

Proof of 6 \Rightarrow 5 Using the family of seminorms we can easily build a local base of absolutely convex absorbent sets for the origin. XXX \square

Remark 7. 1. In Definition 6, note that the family of seminorms is not necessarily countable.

2. If the family of seminorms $\|\cdot\|_k$ generating the topology is *countable*, then this constitutes a pseudometric generating the topology. To see this, define

$$d(x, y) = \sum_{k=1}^{\infty} 2^{-k} \frac{\|x - y\|_k}{1 + \|x - y\|_k}. \quad (8)$$

Note that this pseudo metric is neither unique (so the metrization is not unique) nor homogenous (so we can't define a suitable norm).

3. As translation is continuous in any t.v.s, we can use a base for the origin as a base for any point by translation.

Lemma 5. *Let X be a locally convex TVS with its family of seminorms $(\|\cdot\|_\alpha)_{\alpha \in I}$. Then X is Hausdorff if and only if*

$$\forall \alpha \in I : \|x\|_\alpha = 0 \Leftrightarrow x = 0 \quad (9)$$

Proof. Let X be Hausdorff. We show (9). If for every $\alpha \in I$, $\|x\|_\alpha = 0$ and we assume that $x \neq 0$, we know by the Hausdorff property that there exists an open neighborhood U_x of x and another open neighborhood of 0 such that they do not intersect. As the topology is generated by the seminorms, we can set $U_x = \{u : \|u - x\|_\alpha < \varepsilon \ \forall k \leq K\}$ for some $K \in \mathbb{N}, \varepsilon > 0$. Now obviously $0 \in U_x$ as $\|x\|_\alpha = 0$ for all α . Hence there cannot exist any neighborhood of 0 such that we can separate x and 0 . The other direction of (9) is obvious by the definition of seminorms.

Now suppose (9) holds. If we set $x \neq y$, by (9) we know that there is at least one $\alpha \in I$ such that $\|x - y\|_\alpha = M \neq 0$.

By setting $U_x = \{u \in X : \|u - x\|_\alpha < M/2\}$ and $V_y = \{v \in X : \|v - y\|_\alpha < M/2\}$, we have $U_x \cap V_y = \emptyset$ by the triangle inequality (assume there is a w both in U_x and V_y , then one can show that $\|x - y\|_\alpha \leq \|x - w\|_\alpha + \|w - y\|_\alpha < M$). \square

We are going to compare two definitions of Fréchet spaces.

Definition 7. A Fréchet space X is a topological vector space X such that

1. X is locally convex, i.e. $0 \in X$ has a local base of absorbent and absolutely convex sets.
2. The topology \mathcal{T} can be induced by a translation invariant metric d , i.e. $d(x + a, y + a) = d(x, y)$.
3. (X, d) is a complete metric space.

Definition 8. A Fréchet space X is a topological vector space X such that

1. The topology can be induced by a countable family of seminorms $\|\cdot\|_k$, i.e. U open if and only if for all $u \in U$ there is an integer $K \geq 0$ and an $\varepsilon > 0$ such that

$$\{v : \|v - u\|_k < \varepsilon : k \leq K\} \subset U$$

2. X is Hausdorff.
3. X is complete w.r.t. the family of seminorms $\|\cdot\|_k$, i.e. if $(x_m)_m$ is a Cauchy sequence w.r.t. all seminorms, then there exists an $x \in X$ with $x_n \rightarrow x$ w.r.t. $\|\cdot\|_k$ (and by property 2 even $x_n \rightarrow x$).

Remark 8. 1. Note that the items 1-3 are not all “parallelly equivalent”, but as a set, they are equivalent:

- 7.1 and 7.2 \Rightarrow 8.1 (the countability property of the family of seminorms is stronger than plain local convexity)
- 7.2 \Rightarrow 8.2 (metrizable implies Hausdorff)
- 8.1 and 8.2 \Rightarrow 7.2
- 8.1 \Rightarrow 7.1
- 7.3 \Leftrightarrow 8.3

Lemma 6. *Definitions 7 and 8 are equivalent.*

Proof. **7 \Rightarrow 8.**

ad 1.): From the equivalent definitions of locally convex vector spaces we know that 8.1 is *almost* local convexity. The trouble is that we claim the existence of a *countable* family of seminorms whereas local convexity just gives some family of seminorms. By combining 7.1 (local convexity in its set-theoretic version) and 7.2 (metrizable), we obtain 8.1.

Indeed, the metric d allows us to define small balls $B_n = \{x : d(x, 0) < \frac{1}{n}\}$ for each $n \in \mathbb{N}$. As the topology can be induced by d , we know that there is an open set inside each B_n and by local convexity we know of the existence of an absorbent and absolutely convex set $C_n \subset B_n$. Those sets in turn define via their Minkowski gauges μ_n a countable family of seminorms. Equivalence of the topology of X and the topology generated by the family μ_n follows from the fact that the C_n are a local base of the topology of X .

ad 2.): Every metric (or metrizable) space is Hausdorff.

ad 3.): This is a direct consequence of completeness of d .

8 \Rightarrow 7 ad 1.): As noted above, 8.1 is slightly stronger than 7.1.

ad 2.): By 8.2 we know by lemma 5 that the countable family of seminorms in 8.1 fulfills

$$\forall n \in \mathbb{N} : \|x\|_k = 0 \Leftrightarrow x = 0.$$

This property makes the pseudometric in (8) (which is one possible pseudometrization) a proper metric. The fact that the topology of X is induced by d is equivalent to 8.1. \square

9 The space \mathbb{R}^∞

Lemma 7. *The topological vector space $X = \mathbb{R}^\infty = \{(x_n)_{n \in \mathbb{N}} : x_n \in \mathbb{R} : \forall n \in \mathbb{N}\}$ with the product topology (i.e. the topology generated by all sets of the form $U_1 \times \cdots \times U_m \times \mathbb{R}^\infty$ with all $U_i \subset \mathbb{R}$ open) is a Fréchet space. One possible metrization is*

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \cdot \frac{|x_n - y_n|}{1 + |x_n - y_n|} \quad (10)$$

Proof. We will prove this by showing that \mathbb{R}^∞ fulfills every item in definition 8.

ad 1.): We define $\|x\|_k := |x_k|$, i.e. the absolute value of the k -th item. This is obviously a seminorm on \mathbb{R}^∞ . We show that the product topology can be generated by this family of seminorms.

Let U' be an open set in the product topology. Then U' is a union of sets of the form $U_1 \times \cdots \times U_m \times \mathbb{R}^\infty$. It suffices to consider only open sets out of this basis. Let $U = U_1 \times \cdots \times U_m \times \mathbb{R}^\infty$. Choose a point $u \in U$, i.e. $u = (u_1, \dots, u_m, u_{m+1}, \dots)$ where $u_i \in U_i$ for $i = 1, \dots, m$. Define $\delta = \min_{i=1}^m \text{dist}(u_i, \partial U_i)$, where dist is the distance function on \mathbb{R} . Then the set $\{v : \|v - u\|_k < \delta : k \leq m\}$ is a subset of U , which constitutes the first part of Definition 8.1. The other direction is obvious by definition of the product topology.

ad 2.): By lemma 5 we know that we only need to show that if and only if for given $u \in \mathbb{R}^\infty$ and all $n \in \mathbb{N}$ we have $\|u\|_k = |u_k| = 0$, then $u = 0 \in \mathbb{R}^\infty$. But this is obvious.

ad 3.): This is easily shown by using completeness of \mathbb{R} in each dimension of \mathbb{R}^∞ . \square

Lemma 8. *The space \mathbb{R}^∞ with the product topology is a polish space.*

Proof. \mathbb{R}^∞ is separable as \mathbb{Q}^∞ is countable and dense and it is completely metrizable by definition of Fréchet spaces. \square

Lemma 9. *The product topology of \mathbb{R}^∞ cannot be generated by a norm.*

Proof. Open sets of \mathbb{R}^∞ are necessarily unbounded but balls defined by any norm must be (by definition) bounded. \square

Lemma 10. *The Borel σ -algebra is the same as the product σ -algebra.*

Proof. The Borel- σ -algebra is the σ -algebra generated by open sets (i.e. sets of the form $U_1 \times \cdots \times U_m \times \mathbb{R}^\infty$), which is the same as the product *sigma*-algebra. \square

We write e_i for the element of X with $e_i(j) = \delta_{ij}$. Any element $x \in X$ can be written as $x = \sum_{i=1}^{\infty} x_i e_i$.

Lemma 11. *Every continuous linear functional $f \in X^*$ is of the form*

$$f(x) = \sum_{i=1}^n a_i x_i \quad (11)$$

for some $a_1, \dots, a_n \in \mathbb{R}$ and with the notation $X \ni x = (x_1, x_2, \dots)$. Thus X^* can be identified with c_{00} , the set of all real sequences which are eventually zero.

Proof. Let $f \in X^*$. Define $a_i = f(e_i)$. We show that only finitely many a_i are nonzero. Indeed, assume that for any $N \in \mathbb{N}$ there is an $i_N > N$ such that $a_{i_N} \neq 0$.

As we assumed f to be continuous, for any $\varepsilon > 0$ there is an open neighborhood U of 0 such that

$$\sup_{u \in U} |f(u)| < \varepsilon. \quad (12)$$

By the form of the topology, this U is of the form

$$U = \bigotimes_{i=1}^n U_i \times \mathbb{R}^\infty$$

for some $n \in \mathbb{N}$. Now we can define a sequence $(u_N)_N$ of elements u_N (which are sequences) in X as $u_N = \frac{1}{a_{i_N}} \cdot N \cdot e_{i_N}$, i.e. the N -th element is the sequence which has the value $\frac{N}{a_{i_N}}$ at the i_N -th position and 0 elsewhere. For M high enough, $(u_N)_{N \geq M}$ is a sequence in U . But

$$f(u_N) = \frac{N}{a_{i_N}} \cdot f(e_{i_N}) = N \xrightarrow{N \rightarrow \infty} \infty,$$

hence (12) is violated and f cannot be continuous.

This means that f acts on only a finite number of basis elements of X and as f is linear we can write it in the form (11). \square

We choose the measure μ to be an infinite product of Gaussian measures with variance 1 and mean 0. So the projections of elements $x \in X$ on the coordinates are i.i.d. Gaussians.

Lemma 12. μ is a Gaussian measure. The covariance form of μ is given by

$$q(f, g) = \sum_{i=1}^{\infty} f(e_i)g(e_i)$$

where the sum is actually finite.

Proof. We argue by using characteristic functions.

$$\begin{aligned} \int_X e^{if(x)} \mu(dx) &= \int_{\mathbb{R}^\infty} e^{i \sum_{j=1}^n f(e_j)x_j} \bigotimes_{j=1}^{\infty} N(0, 1)(dx_j) = \prod_{j=1}^n \int_{\mathbb{R}} e^{if(e_j)x_j} N(0, 1)(dx_j) \\ &= \prod_{j=1}^n e^{-\frac{f(e_j)^2}{2}} = e^{-\frac{1}{2} \sum_{j=1}^n f(e_j)^2} \end{aligned}$$

and thus the covariance form is $q(f, f) = \sum_{j=1}^n f(e_j)^2$ and hence $q(f, g) = \sum_{j=1}^{\infty} f(e_j)g(e_j)$ where the sum runs until the largest non-zero entry of both f and g . \square

Lemma 13. μ has full support, i.e. there is no open set other than the empty set having zero measure.

Proof. Let U be open in X , i.e. $U = \bigotimes_{i=1}^n U_i \times \mathbb{R}^\infty$ with U_j open in \mathbb{R} . Then $\mu(U) = \prod_{j=1}^n N(0, 1)(U_j) > 0$ as $N(0, 1)$ has full support on \mathbb{R} . \square

Remark 9. q is actually positive definite: The only $f \in X^*$ with $q(f, f) = 0$ is $f = 0$. This means that $i : X^* \hookrightarrow L^2(X, \mu)$ with inner product q is an injection in the sense that $i(X^*) \subset L^2$ is isomorphic to X^* . This is actually cool because normally there may be $0 \neq f \in X^*$ with $0 = f \in L^2$ (for example in spaces with degenerate Gaussians like $X = \mathbb{R}^2$ with $N(0, 1) \otimes \delta_0$). Here, choosing f as the projection on the second coordinate (written as $f = (0, 1)^T$) is not 0 as a linear functional in $X^* = \mathbb{R}^2$ but $f(x) = 0$ μ -almost surely and thus $0 = f \in L^2$ as a square-integrable random variable. So, to reiterate, $X^* \hookrightarrow L^2(X, \mu)$ is an injection and we can view X^* as a **subspace** of L^2 . This is because we don't lose "information" by viewing f as an element in L^2 . In the example above with \mathbb{R}^2 we do lose information because after identifying X^* with a subset of L^2 , we can't distinguish $(0, 1)^T$ and $(0, 0)^T$ as random variables in L^2 , although they are quite distinct in $X^* = \mathbb{R}^2$.

Remark 10. As was pointed out, we can think of X^* as a subset of $L^2(X, \mu)$ with inner product q . But: X^* is not complete in the q inner product. Let's define $X_\gamma^* = \overline{X^*}^{L^2(X, \mu)}$, the L^2 -closure of X^* .

Lemma 14. X_γ^* consists of all functions $f : X \rightarrow \mathbb{R}$ of the form

$$f(x) = \sum_{j=1}^{\infty} a_j x_j \quad (13)$$

with $\sum_{j=1}^{\infty} |a_j|^2 < \infty$.

Proof. As X^* can be identified with c_{00} and $L^2(X, \mu)$ with inner product q can be identified with l^2 , the space of square summable sequences, this is equivalent to $\overline{c_{00}}^{l^2} = l^2$, which is for convenience proven in lemma 16. \square

Remark 11. Note one subtle point about (13): For an arbitrary $x \in X$, the sum may not converge. It does, however, converge for μ -a.e. $x \in X$. Indeed: Sums of independently and identically distributed (i.i.d.) centered random variables (and the $a_j \cdot x_j$ are i.i.d Gaussians) converge a.s. as soon as the sum of their variances converge. Thus, we only need to show

$$\sum_{j=1}^{\infty} a_j^2 \cdot \mathbb{E} x_j^2 < \infty$$

which follows immediately by $\mathbb{E} x_j^2 = 1$ and the condition on $(a_j)_j$.

By the same calculation as in lemma 12, we can show that f is a Gaussian random variable with covariance form $q(f, g) = \sum_{j=1}^{\infty} f(e_j)g(e_j)$ where now the sum may contain infinitely many terms (but still is of finite value by the Cauchy-Schwarz inequality in l^2).

10 Appendix

Lemma 15 ($(\mathbb{R}^\infty, \|\cdot\|_\infty)$ is complete). *Let $(a^{(k)})_{k \in \mathbb{N}}$ be a Cauchy sequence (of sequences) in the space of sequences with the supremum norm, i.e.*

$$\text{for all } \varepsilon > 0 \text{ there is a } K_\varepsilon \text{ s. t. for all } k, l \geq K_\varepsilon : \|a^{(k)} - a^{(l)}\|_\infty = \sup_n |a_n^{(k)} - a_n^{(l)}| < \varepsilon.^5$$

Then $a^{(k)} \rightarrow a$ in $(\mathbb{R}^\infty, \|\cdot\|_\infty)$, i.e. there is a sequence $a \in \mathbb{R}^\infty$ such that

$$\text{for all } \varepsilon > 0 \text{ there is a } M_\varepsilon \text{ s. t. for all } k \geq M_\varepsilon : \|a^{(k)} - a\|_\infty = \sup_n |a_n^{(k)} - a_n| < \varepsilon,$$

i.e. $a^{(k)}$ converges uniformly to a .

Proof. We use completeness of \mathbb{R} when we realize that by assumption for every $\varepsilon > 0$ there is an index N_ε such that for every $n \in \mathbb{N}$ there is a limit a_n such that for all $l \geq N_\varepsilon$,

$$|a_n^{(l)} - a_n| < \varepsilon.$$

Now we fix $\varepsilon > 0$ and we want to bound $|a_n^{(k)} - a_n| < \varepsilon$ for all $n \in \mathbb{N}$ by choosing k large enough. For this, we write

$$|a_n^{(k)} - a_n| \leq |a_n^{(k)} - a_n^{(l)}| + |a_n^{(l)} - a_n|$$

Now first we choose $K = K_{\varepsilon/2}$, $N = N_{\varepsilon/2}$ and $l \geq \max\{K, N\}$. This makes the second part smaller than $\varepsilon/2$. As soon as we choose $k \geq K$, the first part is also smaller than $\varepsilon/2$. Hence we can choose $M_\varepsilon = K_{\varepsilon/2}$ and we are done. \square

⁵We can think of $(a^{(k)})_k$ being a “uniformly Cauchy” sequence.

Lemma 16 ($\overline{c_{00}}^{l^2} = l^2$). Define $c_{00} = \{(a_n)_{n \in \mathbb{N}} : \exists m : \forall n \geq m : a_n = 0\}$ and choose a Cauchy sequence in c_{00} w.r.t. the l^2 -norm, i.e.

$$\text{for all } \varepsilon > 0 \text{ there is a } N_\varepsilon \text{ s. t. for all } k, l \geq N_\varepsilon : \sum_{n=1}^{\infty} |a_n^{(k)} - a_n^{(l)}|^2 < \varepsilon.$$

Then

$$a^{(k)} \rightarrow a \text{ as a sequence in } l^2$$

Proof. We need to show that there is a sequence a with

1. $\|a^{(k)} - a\|_2 \xrightarrow{k \rightarrow \infty} 0$ and
2. $a \in l^2$.

The existence of such an a follows immediately by dropping the summation symbol and using completeness of \mathbb{R} to obtain a limit element a_n for every $n \in \mathbb{N}$, which constitutes a sequence $a = (a_n)_{n \in \mathbb{N}}$.

Now for 1.) we see that for any $R \in \mathbb{N}$

$$\sum_{n=1}^R |a_n^{(k)} - a_n|^2 \leq \sum_{n=1}^R 2 \cdot |a_n^{(k)} - a_n^{(l)}|^2 + 2 \cdot |a_n^{(l)} - a_n|^2.$$

If we choose $l \geq M := M_{\sqrt{\frac{\varepsilon}{4 \cdot R}}}$ (the index from Lemma 15), the second sum is bounded by $\varepsilon/2$. After setting $N := N_{\sqrt{\frac{\varepsilon}{4}}}$ and claiming additionally $k, l \geq N$ (thus $l \geq \max\{M, N\}$), we see that uniformly in R we just need to set $k \geq N$ in order to bound

$$\sum_{n=1}^R |a_n^{(k)} - a_n|^2 < \varepsilon$$

and thus the bound also holds for the infinite sum, which proves 1.).

For 2.) we see that

$$\sum_{n=1}^R |a_n|^2 \leq \sum_{n=1}^R 2 \cdot |a_n - a_n^{(k)}|^2 + 2 \cdot |a_n^{(k)}|^2$$

which is finitely bounded uniformly in R as we can set k such that the first sum is arbitrarily small and the second sum is some finite value (the l^2 -norm of $a^{(k)}$). \square

11 Acknowledgments

Thank Nate Eldredge, mention his lecture notes, the ISEM lecture notes, Bruce Driver's notes, DaPrato/Zabczyk, Brezis for Functional Analysis. Mention distill.pub and Conjugated Gradient without the agonizing pain.

References

- [1] H. BREZIS, *Functional analysis, Sobolev spaces and partial differential equations*, Springer Science & Business Media, 2010.
- [2] N. ELDRIDGE, *Analysis and probability on infinite-dimensional spaces*, arXiv preprint arXiv:1607.03591, (2016).
- [3] A. LUNARDI, M. MIRANDA, AND D. PALLARA, *ISEM 19 lecture notes*, 2016.
- [4] V. I. PAULSEN AND M. RAGHUPATHI, *An introduction to the theory of reproducing kernel Hilbert spaces*, vol. 152, Cambridge University Press, 2016.