Master Project – Mathematics

EPFL – Lausanne – June 2010

Elements of Statistics

for Inference in Hilbert Spaces

Shahin TAVAKOLI

Supervisor

Victor M. PANARETOS

Abstract

In this Master Project, two tests statistics for functional data are presented. A review of Hilbert space theory is given and fundamentals of probability in abstract spaces are presented. Then basic aspects of Functional Data Analysis are exposed, followed by some notions from inverse problems. Eventually a test for the mean of random curves, and for the Functional Linear Model with scalar responses are presented, with an application to DNA shape analysis.

I'm not going to do my maths homework. Look at these unsolved problems. Here's a number in mortal combat with another. One of them is going to get subtracted. But why? What will be left of him? If I answered these, it would kill the suspense. It would resolve the conflict and turn intriguing possibilities into boring old facts.

—— Calvin & Hobbes ——

Introduction

Whereas the notions of Statistics in Euclidean space are quite straightforward and accessible, the "simple" problem of drawing inferences for the mean of random variables in infinite-dimensional spaces to necessitates further mathematical background.

This Master Project report is the final milestone of a Master's study in Mathematics – and also a first step in the PhD studies of the author. It has the purpose of introducing the necessary tools to understand two asymptotic tests for Functional Data, namely for:

1. The mean of i.i.d. $\mathcal{L}^2([0,1])$-valued random variables,
2. The Functional Linear Model with Scalar Responses.

The basic idea is to derive a test inspired by the multivariate test of the mean for i.i.d. random vectors,

$$M_n(p)^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(\mathbf{X}_k-\mathbf{m}_0)\right),$$

where $\mathbf{m}_0$ is the mean vector under H_0 and $M_n(p)$ is the empirical covariance matrix. The problem that occurs in the infinite dimensional setting is that the empirical covariance *operator* Γ_n is of finite rank, and moreover, the true covariance operator Γ is trace-class. Therefore, the direct inversion "$\Gamma_n^{-1/2}$" makes no sense and an ill-posed inverse problem naturally arises. The use of regularization techniques is then necessary.

We first need to understand infinite dimensional spaces, especially Hilbert spaces. We will thus begin this report with a review of some Hilbert space theory in Section **??**. First we will present some classical examples of Hilbert spaces, such as the ℓ_2 or L^2 spaces, and we will talk about orthogonality and separability in Hilbert spaces. Then we will introduce the notion of a bounded linear operator, which is central to the study of Hilbert spaces. The adjoint of an operator, the notion of compact operator, and the spectral Theorem for compact and self-adjoint operators will be then presented. Eventually, the singular value decomposition Theorem for compact operators will be given.

Once Hilbert spaces have been seen, we will need to understand how to put a handle on probability measures on such spaces. Especially, we need to understand what a Gaussian random variable in a Hilbert space is. In Section 2, we will first introduce the concepts of measure theory that are

useful for the study of random processes, and provide a Theorem about the existence of processes, due to Kolmogorov. We will then introduce Gaussian random vectors and some of their properties, before presenting Gaussian Processes. This way, we will see the analogies between the finite and infinite dimensional setup. Afterwards, basic notions of probabilities in Banach spaces will be presented, and selected limit Theorems for sequences of Banach space-valued random variables will also be given. We will finally conclude with some remarks about the definition of Gaussian processes.

Next, we will give a review of basic aspects of Functional Data Analysis in Section 3. This section will be central to this report, because it will introduce the techniques that allow to transform raw discrete data into functional datum. First the main ideas for turning discrete data into a functional datum will be introduced, by the mean of basis functions, and the least squares representations will be presented. Then we will introduce the notion of roughness penalization, which is of particular importance due to the infinite dimensional nature of the data we are dealing with. A short presentation of registration of functional data will follow, before an introduction to functional principal component analysis (PCA) and regularized PCA. Eventually, we will present the Functional Llinear Model, which is the infinite dimensional analogue of a regression model.

At that point, we would have seen some basic Hilbert space theory, fundamentals of probability in abstract spaces, and also an introduction to functional data. Section 4 will present the last preparation that will be needed in order to understand tests for functional data. We will begin by introducing the notion of an ill-posed inverse problem, and will present the simplest inverse problems: the Fredholm equations of the first kind. We will then expose two methods for resolving such inverse problems: Spectral Truncation, which is based on the SVD decomposition of compact operators, and Tikhonov regularization, which is the infinite dimensional analogue to Ridge regression. We will then end this section by presenting a generalization of Tikhonov reguralization.

Thereafter, we will turn our attention to statistical inference for functional data. In Section 5, we will first start by giving a review of Mas (2007), an article presents a test statistic for the mean of sample curves. We will see that an inverse problem occurs when we try to "free" the centered sample mean $\overline{X} - m_0$ from dependencies on the unknown distribution of the data, and that Tikhonov regularization is necessary. Then, we will present a paper of Cardot et al. (2003), which gives a test for the functional linear model with scalar responses. Again, an inverse problem will naturally arise, and Spectral truncation will be used.

Eventually, we will end this report by giving an application to DNA Shape

Analysis in Section 6. We will test to see if the mean of some DNA minicircles is indeed a circle.

Contents

List of Symbols

Symbol	Signification
※, ◀, 🍃	Symbols for the end of Definitions, Examples and Remarks (respectively)
$\mathbb{F}$	Generic notation for the field $\mathbb{R}$ or $\mathbb{C}$.
$\mathbb{R}_+$ or $\mathbb{R}^+$	$\{x \in \mathbb{R} : x \geq 0\}$.
$(\lambda_j)_{j=1}^n$	The sequence $\lambda_1, \ldots, \lambda_n$.
(λ_j)	$(\lambda_j)_{j=1}^\infty$
ℓ_2	The space of complex value square summable sequences – see Example 1.1.5.
$C([a,b])$	The space of continuous functions $f : [a,b] \to \mathbb{C}$.
$C^m([a,b])$	The space of functions $f : [a,b] \to \mathbb{C}$ such that $f^{(n)} \in C([a,b])$ for $n = 0, \ldots, m$.
$R([0,1])$	The space of Riemann square-integrable functions $f : [0,1] \to \mathbb{R}$.
${}_R\int_0^1$	The Riemann integral on the interval $[0,1]$.
$\mathcal{L}^2([0,1],\mathbb{F})$	The space of Lebesgue square-integrable functions $f : [0,1] \to \mathbb{F}$.
$L^2([0,1],\mathbb{F})$	The equivalence classes of $\mathcal{L}^2([0,1],\mathbb{F})$ – see Example 1.1.8.
$\mathcal{L}^2([0,1])$, $L^2([0,1])$	See Notation 1.1.9.
$\|\cdot\|_2$	The ℓ_2 or L^2 norm.
$\mathcal{L}(E,F)$	The space of *bounded* (or *continuous*) linear operators between E and F.
$\mathcal{L}_C(E,F)$	The space of *compact* linear operators between E and F.

Symbol	Signification
$\mathcal{L}_{HS}(E,F)$	The space of *Hilbert-Schmidt* linear operators between E and F.
$\mathcal{L}_1(E,F)$	The space of *nuclear* or *trace-class* linear operators between E and F.
$\mathcal{L}_\square(E)$	$\mathcal{L}_\square(E,E)$
$e \otimes f$	The tensor product of two elements of a Hilbert space – see Definition 1.5.9.
$\|\cdot\|_{HS}$	The Hilbert-Schmidt norm.
$\mathrm{trace}\,(T)$	The trace of the operator T.
M_f	The multiplication operator on $L^2([0,1])$ induced by $f \in L^2([0,1])$. See Example 1.3.6
T^*	The adjoint of the operator T.
$\sigma(T)$	The spectrum of the operator T.
$\Gamma(T), \Pi(T), \Pi_0(T)$	Respectively the compression, approximate point and point spectrum of T – see Definition 1.6.7.
$\ker(T), \mathrm{Im}(T)$	The kernel, respectively the image (or range) of the operator T.
$\overline{U}, U^\perp$	The closure, respectively the orthogonal complement of the subspace U.
$U \oplus W$	The direct sum of subspaces U and W.
$(S, \mathcal{A})$	A measurable space: $\mathcal{A}$ is a σ-algebra of subsets of S, and $S \neq$.
$\Pi_{i\in I} S_i$	The cartesian product of sets S_i.
$\bigotimes_{i\in I} \mathcal{A}_i$	The product σ-algebra of the $\mathcal{A}_i$'s – See Definition 2.1.2.
$\mathbb{P}(\cdot), \mathbb{E}[\cdot]$	The probability measure and the expectation, respectively.
$\mathcal{B}_E$	The Borel σ-algebra of the topological space E.

Symbol	Signification
$L^2(\Omega, \mathcal{O}, \mathbb{P})$	The space of measurable functions $f : (\Omega, \mathcal{O}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\int_\Omega \lvert f\rvert^2 \mathrm{d}\mathbb{P} < \infty$.
B^*	The topological dual of a Banach space over $\mathbb{F}$: the set of all continuous linear applications $\varphi : B \to \mathbb{F}$.
B-r.v.	Banach space-valued random variable.
φ_X	The characteristic functional of a random variable X.
$\lVert \cdot \rVert_{L^2(\mathbb{P})}$	The $L^2(\mathbb{P})$-norm, whose definition depends on the context. For real-valued random variables, see Definition 2.7.3. For B-r.v. , see Notation 2.6.1.
$\xrightarrow{w}$	Weak convergence of measures.
$\xrightarrow{a.s.}$	Almost sure convergence of random variables.
$\xrightarrow{p}$	Convergence in probability of random variables.
$\xrightarrow{d}$	Convergence in distribution of random variables.
i.i.d.	Independent and Identically Distributed random variables.
(S)LLN, CLT	(Strong) Law of Large Numbers, Central Limit Theorem.
$\mathbf{A}^\mathsf{T}$	The transpose of the matrix $\mathbf{A}$.
$\underset{f}{\operatorname{argmax}}\,(\cdot)$	The argument f that maximizes the value of $(\cdot)$.
$\mathbb{S}^{p-1}$	The unit sphere in $\mathbb{R}^p$.
$\mathbf{A} \otimes \mathbf{B}$	The Kronecker product of matrices – see Definition 3.6.1.
$\alpha \gg \beta$	α is much bigger that β.
$\alpha \approx \beta$	α is approximatively equal to β.
GCV	Generalized Cross-Validation.

1

A Review of Some Hilbert Space Theory

In order to be able to do statistics in Hilbert spaces, we are going to give a review of some Hilbert space theory in this section. We are first going to give some classical examples of Hilbert spaces, such as the ℓ_2 or L^2 spaces, and we will talk about orthogonality and separability in Hilbert spaces. Then we will introduce the notion of a bounded linear operator, which is central to the study of Hilbert spaces. The adjoint of an operator, the notion of compact operator, and the spectral Theorem for compact and self-adjoint operators will be then presented. Eventually, the singular value decomposition Theorem for compact operators will be given.

1.1 Scalar Products, Norms and Infinite Dimensional Vector Spaces

Hilbert spaces are a special kind of normed vector space, so let us begin with the definition of a norm on a vector space:

Definition 1.1.1. Let V be a vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A *norm* on V is a function $\|\cdot\| : V \to \mathbb{R}_+$ that satisfies:

1. $\|v\| \geq 0$ for all $v \in V$, with equality if, and only if, $v = 0$,
2. $\|\mu v\| = |\mu|\|v\|$, for all $v \in V, \mu \in \mathbb{F}$.
3. $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$. ※

The standard example of a vector space is the Euclidean space $\mathbb{R}^n$, with the norm

$$\|(x_1, \ldots, x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Having a norm on a vector space is nice, because it gives it a topology. However, there are some spaces that have a little more than a norm:

Definition 1.1.2. Let V be a vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A *scalar product* (or inner product) on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ that satisfies, for all $v, w, z \in V$ and $\mu \in \mathbb{C}$:

1. $\langle v, v \rangle \geq 0$, with equality if, and only if, $v = 0$,
2. $\langle \mu v + w, z \rangle = \mu \langle v, z \rangle + \langle w, z \rangle$,
3. $\langle v, \mu w + z \rangle = \overline{\mu} \langle v, w \rangle + \langle v, z \rangle$
4. $\langle v, w \rangle = \overline{\langle w, v \rangle}$,

where $\overline{\mu}$ is the complex conjugate of μ. ※

A scalar product provides much more structure than a norm:

Proposition 1.1.3. *Let V be a vector space with a scalar product. Then*

$$\|v\| = \sqrt{\langle v, v \rangle}, \quad v \in V$$

defines a norm on V. Furthermore, we have the famous Cauchy-Schwarz *inequality:*

$$|\langle v, w \rangle| \leq \|v\| \cdot \|w\|, \tag{1.1}$$

with equality if, and only if v and w are collinear.

Proof. See Axler (1997) for instance. □

Definition 1.1.4. A *Hilbert Space* H is a vector-space (real or complex) with a scalar product $\langle \cdot, \cdot \rangle$ which makes it a complete space. In other terms, if we define $\|v\| = \sqrt{\langle v, v \rangle}$ to be the norm of $v \in H$, then H has the property that any Cauchy sequence $(x_n)_{n=1}^{\infty} \subset H$ converges to an $x \in H$, in the topology induced by the norm. ※

A Hilbert Space is hence a vector space, together with a topological structure (given by the norm). In fact, it has a geometrical structure thanks to the scalar product, which allows to define the angle $\theta \in [0, \pi]$ between two elements $v, w \in H$ by the relation

$$\cos \theta = \frac{\langle v, w \rangle}{\|v\| \|w\|}.$$

When we talk about Hilbert Spaces, we often assume implicitly that they are infinite dimensional vector spaces. Thus things get a lot more complicated than with finite dimensional vector spaces, and that is the reason why we do not study just abstract infinite dimensional vector spaces, but Hilbert Spaces, which allow us the use of analysis techniques. Here is a classical example of a Hilbert space:

Example 1.1.5 (Theorem)**.** Let $\mathbb{C}^{\mathbb{N}} = \{x = (x_1, x_2, \ldots) : x_j \in \mathbb{C}, \forall j = 1, 2, \ldots\}$, and define the space

$$\ell_2 = \left\{ x \in \mathbb{C}^{\mathbb{N}} : \sum_{n=1}^{\infty} |x_n|^2 < \infty \right\}.$$

Then with the scalar product $\langle x, y\rangle = \sum_{n=1}^{\infty} x_n \overline{y_n}$ and the coordinate-wise sum and multiplication by scalars, ℓ_2 is a Hilbert Space.

Indeed, the only non-trivial thing to prove is that it is a complete space.

Proof. Let $(x^k)_{k=1}^{\infty} \subset \ell_2$ be a Cauchy sequence, and let us write the ℓ_2-norm by $\|v\|_2 = \sqrt{\langle v, v\rangle}$. Notice that for all $k = 1, 2, \ldots$, $(x_n^k)_{n=1}^{\infty}$ is a sequence of complex numbers. For simplicity of notation, we shall write $(x^k)_k$ instead of $(x^k)_{k=1}^{\infty}$ from now on.

The idea is to look at the sequence formed by the n^{th} coordinate of each x^k to find a candidate $a = (a_1, a_2, \ldots)$ for the limit of $(x^k)_k$, then show that $a \in \ell_2$ and eventually that it is indeed the desired limit.

Fix $n \geq 1$, and look at $(x_n^k)_k \subset \mathbb{C}$. We claim that it is a Cauchy sequence in $\mathbb{C}$. Indeed, fix $\varepsilon > 0$. Because $(x^k)_k$ is Cauchy in ℓ_2, then there is a $k_0 > 0$ such that for all $j_0, j_1 > k_0$,

$$\sum_{n=1}^{\infty} |x_n^{j_0} - x_n^{j_1}|^2 = \|x^{j_0} - x^{j_1}\|_2^2 < \varepsilon^2,$$

and thus $|x_n^{j_0} - x_n^{j_1}| < \varepsilon, \quad \forall j_0, j_1 > k_0$. Hence $(x_n^k)_k$ is Cauchy in $\mathbb{C}$, and because the latter is a complete metric space, the sequence converges to an $a_n \in \mathbb{C}$. This way, we can construct the sequence $a = (a_1, a_2, \ldots)$.

Let us show that $a \in \ell_2$, and that $\lim_{k\to\infty} \|x^k - a\|_2 = 0$. This will complete our example. To show that $a \in \ell_2$, it is enough to show that $x^k - a \in \ell_2$, because $a = x^k - (x^k - a) \in \ell_2$. Let $N > 0$, fix $\varepsilon > 0$. There exists a $k_0 > 0$ such that for all $j_0, j_1 > k_0$,

$$\sum_{n=1}^{N} |x_n^{j_0} - x_n^{j_1}|^2 \leq \sum_{n=1}^{\infty} |x_n^{j_0} - x_n^{j_1}|^2 < \varepsilon^2.$$

Letting $j_1 \to \infty$ on the left-hand side yields then

$$\sum_{n=1}^{N} |x_n^{j_0} - a_n|^2 < \varepsilon^2.$$

Because the left hand equation is true for all $N > 0$, we can let $N \to \infty$ to get $\|x^{j_0} - a\|_2 < \varepsilon$, and thus we have showed two things: $x^{j_0} - a \in \ell_2$ and also, for all $\varepsilon > 0, \exists N > 0$ such that $\forall j_0 > N, \|x^{j_0} - a\|_2 < \varepsilon$. □

And here is an example of a space which is not a Hilbert Space:

Example 1.1.6. The space $C([0,1])$ of continuous complex-valued function on the compact interval $[0,1]$ (the unit interval) with the scalar product $\langle f,g\rangle = \int_0^1 f(t)\overline{g(t)}\mathrm{d}t$ is *not* a Hilbert Space. For instance, the sequence of continuous functions $(f_n)_n$ defined by

$$f_n(t) = \begin{cases} 0 & \text{if } 0 \le t < 0.5 - 1/(2n) \\ n[t - (0.5 - 1/(2n))] & \text{if } 0.5 - 1/(2n) \le t \le 0.5 + 1/(2n) \\ 1 & \text{if } 0.5 + 1/(2n) < t \le 1 \end{cases} \tag{1.2}$$

is a Cauchy sequence, but it doesn't converge in $C([0,1])$. ◀

So the space of continuous functions on the unit interval is not complete with the above defined scalar product. One could ask if a bigger space, such as the Riemann square-integrable functions on $[0,1]$, is complete under the same scalar product. Well, the answer is no, and here is an example of a non-convergent Cauchy sequence given to me by a collegue (Peter Jossen):

Example 1.1.7. The space of Riemann square-integrable functions

$$R([0,1]) = \left\{ f : [0,1] \to \mathbb{C} \;\middle|\; {}_R\!\int_0^1 |f(t)|^2 \mathrm{d}t < \infty \right\},$$

where ${}_R\!\int_0^1$ denotes the integral in the Riemann sense, with the scalar product defined in example 1.1.6 is not complete, and hence it's *not* a Hilbert Space. Here is an example: let $(a_n)_n$ be an enumeration of the rational numbers $[1/3, 2/3] \cap \mathbb{Q}$, and define the sequence of functions

$$f_n(t) = \begin{cases} 0 & \text{if } |t - a_j| < 10^{-j}, \text{ for a } j = 1, \dots, n; \\ 1 & \text{otherwise.} \end{cases} \qquad n \ge 1.$$

Figure 1.1 shows the graph of f_1, f_2 and f_3 for a certain (a_n).

Assertion (f_n) is a Cauchy sequence in $R([0,1])$, but doesn't converge in $R([0,1])$.

Proof. The core idea is that if a function is Riemann integrable, the Lebesgue integral and the Riemann integral coincide. We will use the notation ${}_R\!\int_0^1$ for the Riemann integral, and $\int$ for Lebesgue's integral. Notice that $\forall t \in$

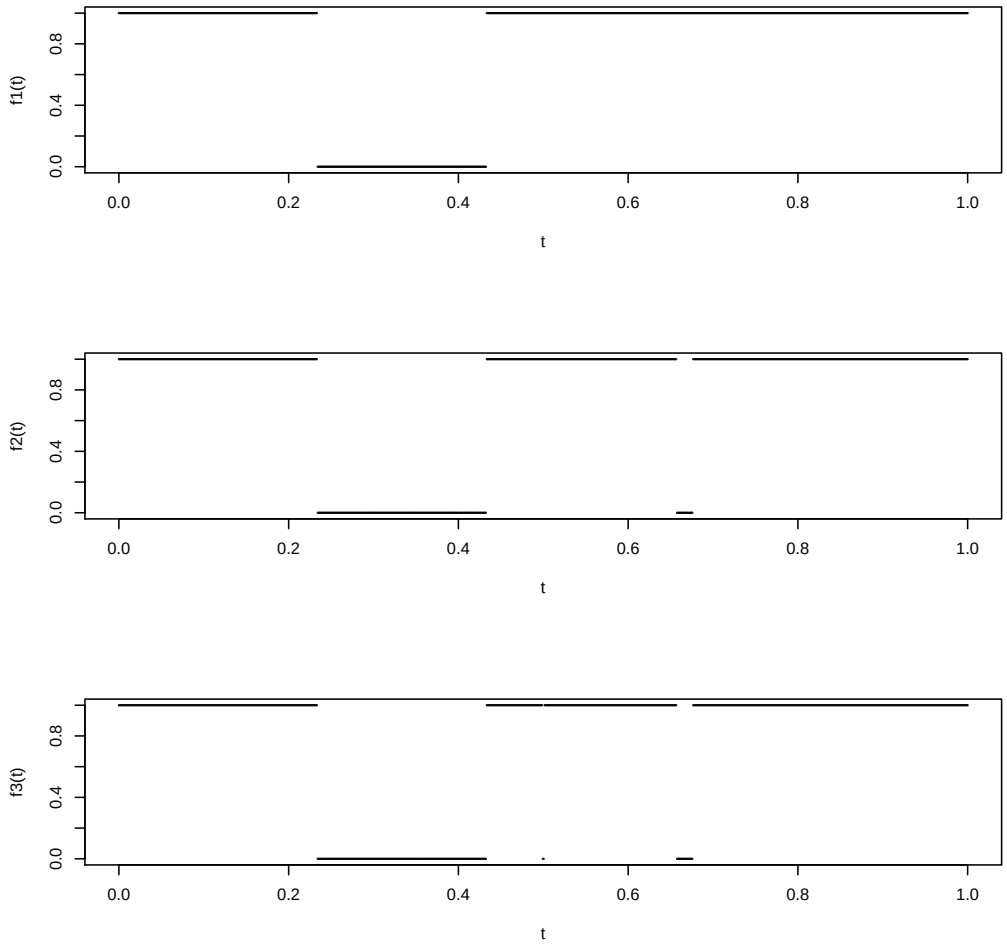

Figure 1.1: From top to bottom: the graphs of f_1, f_2 and f_3, respectively, with $a_1 = 1/3$, $a_2 = 2/3$ and $a_3 = 1/2$. Notice that the interval of length $2 \cdot 10^{-3}$ around $t = 1/2$ in which $f_3(t)$ has value zero.

$[0,1], 1 \geq f_n(t) \geq f_{n+1}(t) \geq f(t)$, where f is the pointwise limit of f_n, explicitly

$$f(t) = \begin{cases} 0 & \text{if } |t - a_j| < 10^{-j}, \text{ for a } j = 1, 2, \ldots; \\ 1 & \text{otherwise.} \end{cases}$$

We have by the dominated convergence Theorem that

$$\int_0^1 f(t)\mathrm{d}t = \lim_{n\to\infty} \int_0^1 f_n(t)\mathrm{d}t \geq 1 - \lim_n \sum_{k=1}^{n} 10^{-k} = \frac{8}{9}.$$

Now suppose there exists an $g \in R([0,1])$ that is the limit of (f_n) in the sense that

$$\lim_n {}_R\int_0^1 (f_n(t) - g(t))^2 \mathrm{d}t = 0.$$

Then by the dominated convergence Theorem, $\int_0^1 (f(t) - g(t))^2 \mathrm{d}t = 0$, and hence $f = g$ a.e. We affirm that $g_{|J} = 0$, where J is a dense subset of $[1/3, 2/3]$. Indeed, if this is not true, then $\exists x_0 \in [1/3, 2/3]$ and $\delta > 0$ such that $g(x) \neq 0$, if $|x - x_0| < \delta$. But there is a rational number $a_k \in (x_0 - \delta, x_0 + \delta)$, hence $f_{|(a_k - 10^{-k}, a_k + 10^{-k})} = 0$. Thus there exists a interval $\tilde{J}$ with non-zero measure such that $g_{|\tilde{J}} \neq 0$ and $f_{|\tilde{J}} = 0$, which leads to a contradiction.

So now we know that g has to be equal to 0 on a dense subset J of $[1/3, 2/3]$, and that g is Riemann-integrable. But then ${}_R\int_{\frac{1}{3}}^{\frac{2}{3}} g(t)\mathrm{d}t = 0$, and hence

$${}_R\int_0^1 g(t)\mathrm{d}t \leq \frac{2}{3} = \frac{6}{9} < \frac{8}{9} = \int_0^1 f(t)\mathrm{d}t,$$

which is a contradiction! Therefore the space $R([0,1])$ is not complete. □

One needs to consider the larger space of Lebesgue-integrable function to have a Hilbert-space:

Example 1.1.8. Let $\mathcal{L}^2([0,1], \mathbb{F})$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, be the space of functions $f : [0,1] \to \mathbb{F}$ which are Lebesgue integrable, and such that $\int_0^1 |f(t)|^2 \mathrm{d}t < \infty$ in the Lebesgue sense. For $f, g \in \mathcal{L}^2([0,1], \mathbb{F})$, let $\langle f, g \rangle = \int_0^1 f(t)\overline{g(t)}\mathrm{d}t$. This almost defines a scalar product, except that we can have $\|f\| = 0$ with $f \neq 0$. But if we quotient the space by the equivalence relation

$$f \sim g \text{ if, and only if } f \neq g \text{ only on a set of measure zero,}$$

then the obtained space, which we will denote $L^2([0,1], \mathbb{F})$, is a Hilbert space. For a proof of this statement, you can see Debnath & Mikusiński (2005, p.76). ◀

Notation 1.1.9. *We will use the shortened notation* $\mathcal{L}^2([0,1]) = \mathcal{L}^2([0,1],\mathbb{R})$ *and* $L^2([0,1]) = L^2([0,1],\mathbb{R})$ *from now on.*

1.2 Orthogonality and Separability in Hilbert Spaces

Having a scalar product allows us to introduce the notion of angle between two elements, and more specifically, define what two orthogonal elements are:

Definition 1.2.1. Let H be a Hilbert space. Two elements $x, y \in H$ are said to be *orthogonal* if $\langle x, y\rangle = 0$.

A family $(x_j)_{j\in J}$ is called an *orthogonal system* if $\langle x_j, x_k\rangle = 0$ whenever $j \neq k$; $j, k \in J$. Furthermore, if $\|x_j\| = 1$ for all $j \in J$, then $(x_j)_{j\in J}$ is called an *orthonormal system.*

If the index set J of an orthonormal system is finite or countable, then we will say that $(x_j)_{j\in J}$ is an *orthonormal sequence.* ※

The most canonical example of orthonormal sequence are the vectors $e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1) \in \mathbb{C}^n$, with scalar product $\langle u, v\rangle = \sum_{j=1}^n u_j \bar{v}_j$, where $u, v \in \mathbb{C}^n$. Orthogonal vectors facilitate computations: for instance, if $(x_1, \ldots, x_n) \subset H$ is an orthonormal system, then

$$\|x_1 + \cdots + x_n\|^2 = \|x_1\|^2 + \cdots + \|x_n\|^2.$$

Also, if $v = \sum_{j=1}^n \mu_j x_j$ with $\mu_j \in \mathbb{C}$, then we can easily find the μ_j's through the equality

$$\mu_j = \langle v, x_j\rangle / \|x_j\|^2. \tag{1.3}$$

Consider now an infinite dimensional Hilbert space H, and an orthonormal *sequence* $(e_j)_{j\in\mathbb{N}}$ in H. We would like to know if any vector $v \in H$ can be expressed in terms of a linear combination of the e_j's. If so, then finding the coefficient would be easy, provided equation (1.3) can be generalized to a series. The following proposition assures us that the series $\sum_{j=1}^\infty \langle x, e_j\rangle e_j$ has a meaning when $x \in H$:

Proposition 1.2.2 (Bessel's inequality)**.** *For any orthonormal sequence* $(e_j)_{j\in\mathbb{N}}$ *in a Hilbert space* H*, and for any* $x \in H$*,*

$$\sum_{j=1}^\infty |\langle x, e_j\rangle|^2 \leq \|x\|^2.$$

Therefore, the limit $\sum_{j=1}^\infty \langle x, e_j\rangle e_j$ *is convergent.*

Proof. See Young (1988, p.34). □

So, we would really like to be able to write

$$x = \sum_{j=1}^{\infty} \langle x, e_j \rangle e_j, \quad \forall x \in H.$$

Unfortunately, this need not be the case, because if H is "too big", then we won't be able to fill the whole space with our orthonormal sequence. Here is an example of such a Hilbert space:

Example 1.2.3 (A non-separable Hilbert space)**.** Let

$$X = \left\{ f : \mathbb{R} \to \mathbb{R} \;\middle|\; E_f = f^{-1}(\mathbb{R} - \{0\}) \text{ is countable and } \sum_{x \in E_f} (f(x))^2 < \infty \right\},$$

and for $f, g \in X$, let $\langle f, g \rangle = \sum_{x \in \mathbb{R}} f(x)g(x)$ be the scalar product. Notice that only a countable number of summands are non-zero, thus the sum is actually a countable sum, and verifying that it defines indeed a scalar product is done in the same way as for ℓ_2. This space is a Hilbert space, and although it looks like the ℓ_2 space, it is significantly different from it, as we will see now.

For any orthonormal sequence $(e_n)_{n \in \mathbb{N}}$ in X, let $E = \cup_{n \in \mathbb{N}} e_n^{-1}(\mathbb{R} - \{0\})$ be the subset of $\mathbb{R}$ where at least one of the e_n is non-zero. Then E is countable (being a countable union of countable sets), and we can pick an $\bar{x} \in \mathbb{R} - E$. Define $f \in X$ by $f(\bar{x}) = 1$ and $f(x) = 0$ if $x \neq \bar{x}$. Then f cannot be expressed as a linear combination of the e_n's. ◀

This motivates the following definition:

Definition 1.2.4. Let H be a Hilbert space. An orthonormal sequence $(e_n)_{n \in \mathbb{N}}$ in H is *complete* if any $x \in H$ can be written as

$$x = \sum_{j=1}^{\infty} \langle x, e_j \rangle e_j.$$

A Hilbert space that contains a complete orthonormal sequence is called *separable*. In the sequel, we will sometimes use the term *orthonormal basis* instead of *complete orthonormal sequence*. ※

A separable Hilbert space is thus a space that can be infinite dimensional, but not "too big" (it must be of countable dimension in some sense), and that contains a (countable) sequence of mutually orthonormal vectors, with the property that the closure of their span is the entire Hilbert space. The following formula is important in this setting:

Proposition 1.2.5 (Parseval's formula)**.** *Let H be a separable Hilbert space, and (e_n) be an orthonormal basis. Then*

$$\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \tag{1.4}$$

Proof. See Debnath & Mikusiński (2005, p.115) □

So it seems that separable Hilbert spaces look like sequence spaces. We will now see that this is essentially all there is.

Definition 1.2.6. A linear bijection $U : H \to K$ between two Hilbert spaces is called a *unitary operator* if it preserves the scalar product, that is,

$$\langle x, y\rangle_H = \langle Ux, Uy\rangle_K, \quad \forall x, y \in H.$$

The Hilbert spaces H, K are called *isomorphic* if there is a unitary operator from H to K. ※

The following result is a classification of the separable Hilbert spaces (Debnath & Mikusiński 2005, p.121); recall that in our notation, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Theorem 1.2.7 (Classification of separable Hilbert spaces)**.** *Let H be a separable Hilbert space. Then if H is finite dimensional, it is isomorphic to $\mathbb{F}^n$, where $n = \dim H$, otherwise it is isomorphic to ℓ_2 (with values in $\mathbb{F}$).*

1.3 Bounded Linear Operators

Let us recall the definition of a linear operator:

Definition 1.3.1. A *linear operator* between two vector spaces E, F over the field $\mathbb{F}$ is a mapping $T : E \to F$ such that

$$T(\mu v + \lambda w) = \mu T(v) + \lambda T(w), \qquad \text{for all } \mu, \lambda \in \mathbb{F} \text{ and } v, w \in E. \tag{1.5}$$

A *linear operator* on E is a linear operator from E to E. ※

When the vector spaces E, F are finite dimensional, then the study of linear operators is done through matrix representations, and the theory essentially becomes *linear algebra*. The finite dimensionality of the spaces enables a great deal of intuition and the theory is pretty well behaved. However, letting the spaces E and F be of infinite dimension complicates the theory of linear operators significantly. We thus restrain ourselves to the study of continuous linear operators. In order to talk of continuity, we need to have a notion of topology, which will naturally arise through the introduction of a suitable *norm* on our vector spaces E and F.

Definition 1.3.2. A *Banach Space* B is a vector space together with a norm $\|\cdot\| : B \to \mathbb{R}_+$, under which it is complete. ※

The notion of Banach Space extends the notion of Hilbert Space. Any Hilbert Space is a Banach space, but the converse is not true. In Hilbert spaces, the scalar product gives a notion of geometry and makes life easier. For example, one can talk of orthonormal bases and orthogonal projections in Hilbert space, and these do not exist in Banach spaces.

Notation 1.3.3. *From now on,* H, H' *will denote Hilbert spaces, and* B, B' *will denote Banach spaces.*

We will from now on concentrate on continuous linear operators. This notion can be generalized to metric spaces (in which continuity has a meaning) but we will not need that kind of generality in this report.

Because of linearity, the continuity condition has the following characterization:

Theorem 1.3.4 (Characterization of continuous linear operators)**.** *Let* L *be a linear operator from* E *to* F*, which are both Banach spaces. Then the following properties are equivalent:*

1. L *is continuous,*
2. L *is continuous at* $0 \in E$*,*
3. *There exists a constant* $M > 0$ *such that* $\|Lx\|_F \leq M\|x\|_E$ *for all* $x \in E$*.*

A linear operator satisfying 3. *is said to be* bounded*.*

Proof. See Rudin (1991, p.24). □

Hence for linear operators, boundedness is equivalent to continuity. From now on, we shall use the term *bounded operator* when talking of a continuous linear operator.

The space of bounded operators $\mathcal{L}(E, F)$ is itself a linear space. When $E = F$, we shall write $\mathcal{L}(E) = \mathcal{L}(E, E)$ for simplicity. It happens that we can define a norm on the space of bounded operators:

Proposition 1.3.5 (Definition)**.** *Let* E, F *be Banach spaces (over the field* $\mathbb{R}$*), and* $\mathcal{L}(E, F) = \{L : E \to F : L \text{ is a bounded operator}\}$*. Then with the operations*

$$(\lambda L + \mu S)x = \lambda(Lx) + \mu(Sx), \qquad \forall \mu, \lambda \in \mathbb{R}, \quad \forall x \in E, \tag{1.6}$$

$\mathcal{L}(E,F)$ is a vector space, and

$$\|L\| = \sup\{\|Lx\| : x \in E, \|x\| \leq 1\} \tag{1.7}$$

defines a norm on it, called the operator norm. *Furthermore, if G is another Banach space, then for any $L \in \mathcal{L}(E,F)$ and $S \in \mathcal{L}(F,G)$ we have*

$$\|SL\| \leq \|S\|\|L\| \tag{1.8}$$

and $S \circ L \in \mathcal{L}(E,G)$.

Proof. The only non-trivial thing to show is that (1.7) defines a norm on $\mathcal{L}(E,F)$ and that (1.8) holds. Take $L, S \in \mathcal{L}(E,F)$. Then

$$\begin{aligned} \|L+S\| &= \sup\{\|Lx+Sx\| : x \in E, \|x\| \leq 1\} \\ &\leq \sup\{\|Lx\| + \|Sx\| : x \in E, \|x\| \leq 1\} \\ &\leq \sup\{\|Lx\| : x \in E, \|x\| \leq 1\} \\ &\qquad + \sup\{\|Sx\| : x \in E, \|x\| \leq 1\} \\ &= \|L\| + \|S\|. \end{aligned} \tag{1.9}$$

And of course, if $\lambda \in \mathbb{R}$, then

$$\begin{aligned} \|\lambda L\| &= \sup\{\|\lambda Lx\| : x \in E, \|x\| \leq 1\} \\ &= |\lambda| \sup\{\|Lx\| : x \in E, \|x\| \leq 1\} \\ &= |\lambda|\|L\|. \end{aligned}$$

Now for (1.8), notice that for any $x \neq 0$, $x/\|x\|$ has norm 1 and thus

$$\frac{1}{\|x\|}\|L(x)\| = \left\| L\left(\frac{x}{\|x\|}\right)\right\| \leq \|L\|,$$

hence $\|Lx\| \leq \|L\|\|x\|$ for all $x \in E$. Therefore $\|SLx\| = \|S(Lx)\| \leq \|S\|\|Lx\| \leq \|S\|\|L\|\|x\|$, and (1.8) is established. □

Just like in linear algebra, some important subsets have to be defined: the *kernel* of $L : E \to F$ is the subspace $\ker(L) = \{x \in E : Lx = 0\}$ of E, and the *range* or *image* of L is the subspace $\mathrm{Im}(L) = \{Lx : x \in E\}$ of F. Also, we will say that a linear operator (not necessarily bounded) $L : E \to F$ is invertible if there is a linear operator $S : F \to E$ such that $SL = Id_E$ and $LS = Id_F$, where Id_E is the identity on E,
Now let us give a few example of linear operators:

Examples 1.3.6. Here are some classical examples of bounded operators:

The identity operator Let B be a Banach space. Then the identity operator $Id_B : B \to B$ is a bounded operator with norm $\|Id_B\| = 1$.

A shift operator Consider the operator $S : \ell_2 \to \ell_2$ defined by

$$S(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots)$$

(see example 1.1.5). It is a linear operator, and furthermore, $\|Sx\|_2 = \|x\|_2$ for all $x \in \ell_2$. Therefore $\|S\| = 1$, and S is a bounded operator on ℓ_2. Notice that $ker(S) = \{0\}$, but that S isn't invertible.

A multiplication operator Consider the space $L^2([0,1])$ with the usual scalar product (defined in example 1.1.8) and let $f \in L^2([0,1])$ be a continuous function. We define the operator M_f on $L^2([0,1])$ by

$$(M_f x)(t) = f(t)x(t), \quad x \in L^2([0,1]), t \in [0,1].$$

It is a bounded operator, because

$$\|M_f x\|^2 = \int_0^1 |f(t)x(t)|^2 \mathrm{d}t \leq \|f\|_\infty^2 \int_0^1 |x(t)|^2 \mathrm{d}t = \|f\|_\infty^2 \|x\|^2.$$

where $\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|$. Hence $\|M_f\| \leq \|f\|_\infty$. (In fact, $\|M_f\| = \|f\|_\infty$). ◀

The following example defines a widely used class of linear operators, called integral operators:

Example 1.3.7 (Integral operators)**.** Let $k : [0,1] \times [0,1] \to \mathbb{C}$ be a continuous function, called the *kernel* of the integral operator, and define the operator K on $L^2([0,1])$ by

$$(Kx)(t) = \int_0^1 k(t,s)x(s)\mathrm{d}s, \qquad t \in [0,1]. \tag{1.10}$$

Then, by the Cauchy-Schwarz inequality, we have for any fixed $t \in [0,1]$,

$$|Kx(t)|^2 = \left| \int_0^1 k(t,s)x(s)\mathrm{d}s \right|^2 \leq \int_0^1 |k(t,s)|^2 \mathrm{d}s \int_0^1 |x(s)|^2 \mathrm{d}s = \int_0^1 |k(t,s)|^2 \mathrm{d}s \, \|x\|^2$$

and thus

$$\|Kx\|^2 \leq \left(\int_0^1 \int_0^1 |k(t,s)|^2 \mathrm{d}s\mathrm{d}t \right) \|x\|^2.$$

Because of the continuity of k, the double integral is finite and hence

$$\|M\| \leq \iint_{[0,1]^2} |k(t,s)|^2 \mathrm{d}s\mathrm{d}t < \infty, \tag{1.11}$$

thus operator M is bounded. ◀

Remark 1.3.8. Notice that we could have imposed a weaker condition on k, by asking only that $k \in L^2([0,1] \times [0,1])$, and the resulting integral operator would have still been bounded. 🍃

Unfortunately, not all operators are bounded. And in fact, some very natural and often-encountered operators are unbounded:

Example 1.3.9 (A differential operator)**.** Let $\mathcal{P}$ be the space of formal trigonometric polynomials on the interval $[0, 2\pi]$, that is,

$$\mathcal{P} = \left\{ \sum_{n=1}^{N} (\mu_n \cos(nt) + \nu_n \sin(nt)) : N \in \mathbb{N} \text{ and } \mu_n, \nu_n \in \mathbb{C} \text{ for } n = 1, \ldots, N \right\}.$$

Define the scalar product $\langle f, g \rangle = \int_0^{2\pi} f(t)\overline{g(t)}\mathrm{d}t$ for $f, g \in \mathcal{P}$, and let $\mathcal{P}_1$ be the closure of $\mathcal{P}$ under the norm defined by this scalar product. We define the differential operator $\frac{\mathrm{d}}{\mathrm{dt}}$ on $\mathcal{P}$ which maps $f \in \mathcal{P}$ to $f' \in \mathcal{P}$, its formal derivative. It is a well defined linear operator *on* $\mathcal{P}$, but it cannot be extended to a bounded operator on $\mathcal{P}_1$. Indeed, for all $n = 1, 2, \ldots$ define $f_n(t) = \frac{1}{\sqrt{\pi}} \sin(nt) \in \mathcal{P}$. Then

$$\|f_n\|^2 = \frac{1}{\pi} \int_0^{2\pi} \sin^2(nt)\mathrm{d}t = 1, \quad \forall n \geq 1.$$

But

$$\|\frac{\mathrm{d}}{\mathrm{dt}} f_n\|^2 = \|f_n'\|^2 = \int_0^{2\pi} \frac{n^2}{\pi} \cos^2(nt)\mathrm{d}t = n^2$$

and hence $\|\frac{\mathrm{d}}{\mathrm{dt}}\| = \infty$ and so that $\frac{\mathrm{d}}{\mathrm{dt}}$ is an *unbounded* operator on $\mathcal{P}_1$. ◀

Remark 1.3.10. Of course, the notion of boundedness of an operator is very closely connected to the norm of the spaces. If we define the space $\mathcal{P}_2$ to be the closure of $\mathcal{P}$ under the norm

$$\|f\|_{\mathcal{P}_2}^2 = \int_0^{2\pi} \left(|f(t)|^2 + |f'(t)|^2 \right) \mathrm{d}t,$$

then the differential operator

$$\frac{\mathrm{d}}{\mathrm{dt}} : \mathcal{P}_2 \to \mathcal{P}_1$$

would be a bounded operator! 🍃

1.4 The Adjoint of an Operator

For any Banach space B, an important associated space is the space of mappings $F : H \to \mathbb{F}$ which are linear and continuous (or bounded). We shall denote this space by B^*. It is called the *dual (space)* of B, and its elements are called (continuous) functionals on H.

The dual space B^* is a Banach space, with pointwise algebraic operations and the operator norm (for a proof of this statement if B is a Hilbert space, see Young (1988, p.61). The proof naturally extends to Banach spaces, for it only uses the norm in B and not the scalar product).

The dual of an abstract Banach space can be pretty complicated (for instance, the dual of $L^\infty(\mathbb{R})$ has no simple representation – see Dunford & Schwartz (1988, Theorem 16, p.296)) but for Hilbert spaces, the dual and the space itself can be identified. This important result is called the Riez-Fréchet Theorem.

Theorem 1.4.1 (Riesz-Fréchet)**.** *Let H be a Hilbert space and let F be a functional on H. Then there exists a unique $y \in H$ such that*

$$F(x) = \langle x, y \rangle, \quad \forall x \in H,$$

and $\|F\| = \|y\|$.

Proof. See Young (1988, p.62). □

Just like in linear algebra, where the adjoint A^* of a linear transformation A is defined by $\langle Ax, y \rangle = \langle x, A^* y \rangle$, we can define the adjoint of a bounded operator between two Hilbert spaces:

Theorem 1.4.2 (Definition)**.** *Let $T \in \mathcal{L}(E, F)$, where E, F are Hilbert spaces. Then there exists a unique operator $T^* \in \mathcal{L}(F, E)$ such that*

$$\langle Tx, y \rangle_F = \langle x, T^* y \rangle_E. \tag{1.12}$$

The operator T^* is called the adjoint *of T.*

Proof. See Young (1988, p.76). □

Now let us calculate some adjoints of operators:

Example 1.4.3 (Adjoint of a multiplication operator)**.** The multiplication operator M on $L^2([0,1])$ defined by

$$Mx(t) = f(t)x(t), \quad t \in [0,1],$$

where $f \in C([0,1])$, has an adjoint M^* that is also a multiplication operator. To find it, by definition of the adjoint, we must have

$$\langle Mx, y\rangle = \langle x, M^*y\rangle, \quad \forall x, y \in L^2([0,1]).$$

Rewriting this equation yields

$$\int_0^1 f(t)x(t)\overline{y(t)}\mathrm{d}t = \int_0^1 x(t)\overline{M^*y(t)}\mathrm{d}t. \tag{1.13}$$

Hence

$$\overline{M^*y(t)} = f(t)\overline{y(t)} \text{ almost everywhere.} \tag{1.14}$$

It follows that

$$M^*y(t) = \overline{f(t)}y(t).$$

Thus the adjoint of a multiplication operator is also a multiplication operator. Notice that you can think of it in a "finite dimensional" way: we reduced the integral equation (1.13) to an infinite dimensional analogue of the linear algebra adjoint equation (1.14). We can easily see how the result of this example could be extended to an operator on a vector-valued L^2 space, with multiplication operator given by a matrix A_t. ◀

Example 1.4.4 (Adjoint of an integral operator)**.** Let $k : [0,1] \times [0,1] \to \mathbb{C}$ be continuous and define the integral operator K on $L^2([0,1])$ by

$$Kx(t) = \int_0^1 k(t,s)x(s)\mathrm{d}s.$$

The adjoint of K must satisfy $\langle Kx, y\rangle = \langle x, K^*y\rangle$ for all $x, y \in L^2([0,1])$, that is

$$\int_0^1 \int_0^1 k(t,s)x(s)\overline{y(t)}\mathrm{d}s\mathrm{d}t = \int_0^1 x(t)\overline{(K^*y)(t)}\mathrm{d}t. \tag{1.15}$$

Now Fubini's Theorem allows us to interchange the order of integration to get

$$\int_0^1 x(t)\overline{(K^*y)(t)}\mathrm{d}t = \int_0^1 \int_0^1 x(s)k(t,s)\overline{y(t)}\mathrm{d}t$$

and interchanging the role of t and s yields

$$\begin{aligned} &= \int_0^1 \int_0^1 x(t)k(s,t)\overline{y(s)}\mathrm{d}s\mathrm{d}t \\ &= \int_0^1 x(t)\left(\int_0^1 k(s,t)\overline{y(s)}\mathrm{d}s\right)\mathrm{d}t. \end{aligned}$$

Thus we must have

$$K^*y(t) = \int_0^1 \overline{k(s,t)}y(s)\mathrm{d}s, \quad \text{for almost every } t \in [0,1]. \tag{1.16}$$

K^* is therefore an integral operator, with kernel k^* given by the equation

$$k^*(t,s) = \overline{k(s,t)}.$$

Again, we can draw an analogy with the adjoint in the finite dimensional case: the kernel k^* can be seen as a complex transpose of the kernel k – informally of course. ◀

As we have seen in these examples, an operator and its adjoint are closely related, and the following Theorem gives a relation between them:

Theorem 1.4.5. *Let $T \in \mathcal{L}(E,F)$, where E, F are Hilbert spaces. Then $T^{**} = T$ and $\|T^*\| = \|T\|$.*

Proof. See Young (1988, p.78). □

Some bounded operators are particularly tractable, especially those which are equal to their adjoint:

Definition 1.4.6. Let H be a Hilbert space. An operator $T \in \mathcal{L}(H)$ is called *normal* if $TT^* = T^*T$, that is, if T commutes with its adjoint.

An operator T is called a *self-adjoint* or *Hermitian* operator if $T = T^*$. Notice that a self-adjoint operator is necessarily normal. ※

For example, a multiplication operator is self-adjoint if, and only if, $f(t) = \overline{f(t)}$ almost everywhere. An integral operator will be self-adjoint if, and only if $k^*(t,s) = \overline{k(s,t)}$ almost everywhere.

1.5 Compact Operators

A class of operators that have nice properties (close to those of matrices) is the class of compact operators.

Definition 1.5.1. A *compact operator* $T : E \to F$ between two Banach spaces is a linear operator such that, for every bounded sequence $(x_n)_{n\in\mathbb{N}}$ in E (that is, $\exists K > 0$ such that $\|x_n\| < K, \forall n$), the sequence $(Tx_n)_n \subset F$ has a convergent subsequence. ※

Notice that a compact operator is necessarily bounded, otherwise there would exist a bounded sequence $(x_n)_n$ in E such that $\|Tx_n\| \to \infty$ when $n \to \infty$, and $(Tx_n)_n$ wouldn't have any convergent subsequence. Here is a standard example of a compact operator:

Example 1.5.2 (Finite rank bounded operators)**.** Define the *rank* of an operator $T : E \to F$ between two Banach spaces as the dimension of its image. If it is finite, then we say that T has finite rank. Every bounded finite rank operator is compact (Young 1988, p.89). ◀

However, the most natural operator is not compact on an infinite dimensional Hilbert space:

Example 1.5.3 (Identity operator)**.** Let H be an infinite dimensional Hilbert space, and take $(e_n)_{n\in\mathbb{N}}$ to be an infinite orthonormal sequence (such a sequence always exists). The sequence (e_n) is obviously bounded, however, for all $n \neq m$ we have

$$\|e_n - e_m\|^2 = \|e_n + (-e_m)\|^2 = \|e_n\|^2 + \|e_m\|^2 = 2,$$

hence $(Ie_n)_n$ has no converging subsequence, and the identity operator is not compact. ◀

The corollary to the following Theorem gives a method for constructing compact operators:

Theorem 1.5.4. *Let E, F be Banach spaces. Then the space of compact operators is closed in $\mathcal{L}(E, F)$ with respect to the operator norm.*

Proof. See Young (1988, p.91). □

Corollary 1.5.5. *The limit (with respect to the operator norm) of a converging sequence of bounded finite rank operators is compact.*

The next example gives a characterization of compact diagonal operators:

Example 1.5.6 (Diagonal operators)**.** Let $(\lambda_n)_n$ be a bounded sequence of complex numbers, and let H be a separable Hilbert space (of infinite dimension) with the complete orthonormal sequence $(e_n)_n$. Then the operator $A \in \mathcal{L}(H)$ defined by

$$A\left(\sum_n x_n e_n\right) = \sum_n \lambda_n x_n e_n, \quad x_n \in \mathbb{F},$$

is compact if, and only if $\lambda_n \to 0$.

Proof. See Young (1988, p.90). □

Compactness is a difficult property to verify. There is another property, stronger than compactness, that is easier to verify:

Definition 1.5.7. Let E, F be Hilbert spaces, and $T : E \to F$ be a bounded operator. Then T is called a *Hilbert-Schmidt* operator if there exists a complete orthonormal sequence $(e_n)_n$ in E such that

$$\sum_n \|Te_n\|^2 < \infty. \tag{1.17}$$

Remark 1.5.8. This notion is well defined, because if $\sum_n \|Te_n\|^2 < \infty$ for a complete orthonormal sequence $(e_n)_n$, then for any other complete orthonormal sequence $(f_n)_n$, we have the equality

$$\|Te_n\|^2 = \sum_m |\langle Te_n, f_m\rangle|^2$$

from Parseval's formula (1.4). Hence

$$\sum_n \|Te_n\|^2 = \sum_n \sum_m |\langle Te_n, f_m\rangle|^2 = \sum_m \sum_n |\langle e_n, T^* f_m\rangle|^2 = \sum_m \|T^* f_m\|^2, \tag{1.18}$$

where the change of summation order is justified because the summands are positive. Hence we have

$$\infty > \sum_n \|Te_n\|^2 = \sum_m \|T^* f_m\|^2 = \sum_m \|Tf_m\|^2,$$

where the last equality comes from (1.18) by taking $(e_n)_n = (f_n)_n$. Hence the notion of Hilbert-Schmidt operator is well defined.

Notice also that the condition (1.17) implies that T is bounded, because if $x = \sum_i \lambda_i e_i \in E$ with norm less than one, then by the triangle and Cauchy-Schwarz inequalities,

$$\|Tx\| \leq \sum_i |\lambda_i| \|Te_i\| \leq \sqrt{\sum_i |\lambda_i|^2 \sum_i \|Te_i\|^2} = \|x\| \|T\|_{HS} < \infty,$$

and thus $\|T\| \leq \|T\|_{HS}$.

We can define a norm on the space of Hilbert-Schmidt operators. But first, let us introduce the tensor product notation:

Definition 1.5.9. Let E, F be Hilbert spaces. For $e \in E, f \in F$, we define $e \otimes f \in \mathcal{L}(E, F)$ by

$$e \otimes f(v) = \langle e, v \rangle f, \quad v \in E.$$

Furthermore, the operator norm of $e \otimes f$ is $\|e \otimes f\| = \|e\|\|f\| < \infty$, and $e \otimes f$ is a Hilbert-Schmidt operator. ※

The interesting thing about Hilbert-Schmidt operators is that they form a Hilbert space:

Proposition 1.5.10 (Definition)**.** *Let E, F be separable Hilbert spaces. The* Hilbert-Schmidt *norm of an operator $T : E \to F$, is defined by*

$$\|T\|_{HS} = \sqrt{\sum_n \|Te_n\|^2}, \tag{1.19}$$

for some complete orthonormal sequence (e_n) of E, and the sum is independent of the choice of the complete orthonormal sequence.

We denote by $\mathcal{L}_{HS}(E, F)$ the space of Hilbert-Schmidt operators $T : E \to F$. It is itself a separable *Hilbert space with the scalar product*

$$\langle T, S \rangle_{HS} = \sum_n \langle Te_n, Se_n \rangle, \quad T, S \in \mathcal{L}_{HS}(E, F), \tag{1.20}$$

which is independent of the choice of $(e_n) \subset E$. If (f_m) is a complete orthonormal sequence in F, then

$$(e_n \otimes f_m)_{n,m}$$

is a complete orthonormal sequence in $\mathcal{L}_{HS}(E, F)$.

Furthermore, $T^* \in \mathcal{L}_{HS}(F, E)$ if and only if $\mathcal{L}_{HS}(E, F)$ and we have

$$\|T\|_{HS} = \|T^*\|_{HS}. \tag{1.21}$$

Proof. Notice that (1.19) and (1.21) are direct consequences of Remark 1.5.8. For the independence of (1.20) from the choices of the orthonormal sequence, notice that for any complete orthonormal sequence (v_m) of F, we have $Se_n = \sum_m \langle Se_n, v_m \rangle v_m$ and thus

$$\begin{aligned}
\sum_n \langle Te_n, Se_n \rangle &= \sum_n \sum_m \overline{\langle Se_n, v_m \rangle} \langle Te_n, v_m \rangle \\
&= \sum_m \sum_n \overline{\langle T^* v_m, e_n \rangle} \langle S^* v_m, e_n \rangle \\
&= \sum_m \langle S^* v_m, T^* v_m \rangle.
\end{aligned}$$

Here, the permutation of the sums is allowed because the sum is absolutely convergent. Indeed, notice that the Cauchy-Schwarz inequality yields

$$\begin{aligned}\sum_n \left[\sum_m \left|\overline{\langle Se_n, v_m\rangle}\langle Te_n, v_m\rangle\right|\right] &\leq \sum_n \left[\sum_m |\langle Se_n, v_m\rangle|^2 \sum_m |\langle Te_n, v_m\rangle|^2\right]^{1/2} \\ &= \sum_n \|Se_n\|\|Te_n\| \\ &\leq \|S\|_{HS}\|T\|_{HS} < \infty.\end{aligned}$$

Thus if (e'_n) is another complete orthonormal sequence in E, then

$$\sum_n \langle Te_n, Se_n\rangle = \sum_m \langle S^* v_m, T^* v_m\rangle = \sum_n \langle Te'_n, Se'n\rangle,$$

and thus (1.20) is independent of the choice of the complete orthonormal sequence.

Let us now end the proof by showing that $(e_n \otimes f_m)_{n,m}$ is a complete orthonormal sequence in $\mathcal{L}_{HS}(E,F)$. First, notice that for $T \in \mathcal{L}_{HS}(E,F)$, equation (1.20) yields

$$\lambda_{i,j} := \langle T, e_i \otimes f_j\rangle_{HS} = \langle Te_i, f_j\rangle,$$

and thus $\langle e_i \otimes e_j, e_n \otimes e_m\rangle_{HS} = \delta_{i,n}\delta_{j,m}$ shows that the sequence $(e_n \otimes f_m)_{n,m}$ is orthonormal. Notice also that $\sum_{i,j} \lambda_{i,j}^2 = \|T\|_{HS}^2$. Hence, the Hilbert-Schmidt operator $\tilde{T} = \sum_{i,j} \lambda_{i,j} e_i \otimes f_j$ is well defined. It is in fact equal to T because

$$\begin{aligned}\|T - \tilde{T}\|_{HS}^2 &= \sum_n \left\| Te_n - \sum_j \lambda_{n,j} f_j \right\|^2 \\ &= \sum_n \left\| Te_n - \sum_j \langle Te_n, f_j\rangle f_j \right\|^2 \\ &= 0,\end{aligned}$$

since (f_j) is a *complete* orthonormal sequence in F. Thus $(e_n \otimes f_m)_{n,m}$ is a complete orthonormal sequence in $\mathcal{L}_{HS}(E,F)$, and this shows that $\mathcal{L}_{HS}(E,F)$ is indeed a separable Hilbert space. □

The following Theorem characterizes the Hilbert-Schmidt operators on L^2 spaces:

Theorem 1.5.11. *When $H = L^2(M, d\mu)$, with (M, μ) a measure space, then $A \in \mathcal{L}(H)$ is Hilbert-Schmidt if and only if it is an integral operator, with kernel $k \in L^2(M \times M, d\mu \times d\mu)$.*

Proof. See Reed & Simon (1972, p.210). □

Theorem 1.5.12. *Hilbert-Schmidt operators are compact.*

Proof. See Young (1988, p.93) □

Now we can ask ourselves if all compact operators are Hilbert-Schmidt. If our Hilbert space is an $L^2(M, d\mu)$ space, then the question is equivalent to asking whether all compact operators are integral operators. It is unfortunately not the case:

Example 1.5.13. Take a diagonal operator T on a Hilbert space H with $\lambda_n = 1/\sqrt{n}$, $n \geq 1$ and let $(e_n)_n$ be the corresponding complete orthonormal sequence (see 1.5.6 for the definition of a diagonal operator). Then

$$\sum_n \|Te_n\|^2 = \sum_n \frac{1}{n} = \infty,$$

and thus, T is compact but not Hilbert-Schmidt. ◀

Another important type of operators are the nuclear operators, for which the notion of trace can be defined:

Definition 1.5.14. A bounded operator $T \in \mathcal{L}(H)$ on a Hilbert space H is called *nuclear* – or *trace class* – if there exists a complete orthonormal sequence $(e_n)_n$ in H such that

$$\sum_n |\langle e_n, Te_n\rangle| < \infty. \tag{1.22}$$

In this case, we define the trace by the formula

$$\text{trace}\,(T) = \sum_n \langle e_n, Te_n\rangle.$$

Remark 1.5.15. As with the definition of a Hilbert-Schmidt operator, the definition of the trace of a nuclear operator doesn't depend on the choice of the complete orthonormal basis. Suppose T satisfies (1.22) for a particular (e_n). Using Weidmann (1980, Theorem 7.9), we get the decomposition $T = T_1T_2$ with $T_1, T_2 \in \mathcal{L}_{HS}(H)$. Noticing that $\langle e_n, Te_n\rangle = \langle T_1^* e_n, T_2 e_n\rangle$, we directly have, for any other orthonormal basis (v_m)

$$\sum_n \langle e_n, Te_n\rangle = \langle T_1^*, T_2\rangle_{HS} = \sum_m \langle T_1^* v_m, T_2 v_m\rangle = \sum_m \langle v_m, Tv_m\rangle,$$

where the second equality is due to Proposition 1.5.10.

Notice also that a trace class operator is necessarily Hilbert-Schmidt. Indeed, using once again the decomposition $T = T_1T_2$, we get

$$\sum_n \|Te_n\|^2 = \sum_n \|T_1T_2e_n\|^2 \leq \|T_1\| \sum_n \|T_2e_n\|^2 < \infty,$$

where $\|T_1\| < \|T_1\|_{HS} < \infty$. If we denote by $\mathcal{L}_1(H)$ ($\mathcal{L}_C(H)$) the set of all trace class (respectively compact) operators on H, we have the following inclusions:

$$\mathcal{L}_1(H) \subseteq \mathcal{L}_{HS}(H) \subseteq \mathcal{L}_C(H) \subseteq \mathcal{L}(H).$$

1.6 The Spectral Theorem for Compact Self-Adjoint Operators

The Spectrum

When studying linear operators on finite dimensional spaces, it is equivalent to say that λ is an eigenvalue of A and that the operator $A - \lambda I$ is not invertible. Indeed, if T is an operator on a finite dimensional space, the following conditions are equivalent:

1. $\ker(T) = \{0\}$, that is, T is injective.
2. T is invertible.

However, this isn't true for infinite dimensional vector spaces: an operator can be injective without being invertible! For instance, the shift operator S on ℓ_2 defined by

$$S(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots)$$

in injective, but not invertible. Nevertheless, if an operator is not injective, it cannot be invertible. Thus for infinite dimensional spaces, instead of using only the set of eigenvalues, we define the *spectrum* of an operator, which contains the set of eigenvalues of the operator:

Definition 1.6.1. Let B be a Banach space and $A \in \mathcal{L}(B)$. The *spectrum* of A is the set

$$\sigma(A) = \{\lambda \in \mathbb{F} : A - \lambda I \text{ is not invertible}\}.$$

The eigenvalues of A is the set of $\lambda \in \mathbb{F}$ such that $Av = \lambda v$ for a non-zero $v \in B$. ※

It is possible for an operator to have no eigenvalue, even if it is compact:

Example 1.6.2. Let $H = L^2([0,1])$. The Volterra operator on H defined by

$$Kf(t) = \int_0^t f(s)\mathrm{d}s = \int_0^1 \chi_{[0,t]}(s) f(s)\mathrm{d}s$$

is a (bounded) integral operator, with $k(t,s) = \chi_{[0,t]}(s)$, where χ_A is the indicator function on the set A. It is a Hilbert-Schmidt operator, and hence is compact, but it has no eigenvalue. Indeed, suppose there exists a non-zero element $x \in H$ and $\lambda \in \mathbb{R}$ such that

$$\int_0^t x(s)\mathrm{d}s = \lambda x(t), \quad \forall t \in [0,1]$$

Then by differentiating with respect to t, we get the differential equation $x(t) = \lambda \dot{x}(t)$, with the initial condition $x(0) = 0$. If $\lambda = 0$, then $x = 0$, so $\lambda = 0$ is not an eigenvalue. If $\lambda \neq 0$, the general solution of this differential equation is $x(t) = ke^{t/\lambda}$, and the initial condition forces us to set $k = 0$, and thus $x = 0$. In conclusion, the Volterra operator K has no eigenvalue. ◀

One may wonder why we do not just talk about the eigenvalues of an operator, and we bother ourselves with the spectrum. I posed the question to Professor Nicolas Varopoulos, and he replied that the motivation was the development of symbolic calculus for operators. For an operator $T \in \mathcal{L}(H)$, we make sense of T^2 by composition of T with itself: $T^2 = T \circ T$. In the same way, we can define T^n for any $n \in \mathbb{N}$, with the convention $T^0 = I$. Hence we can give a meaning to $p(T)$, where $p \in \mathbb{C}[x]$ is a polynomial with complex coefficients. The next question that naturally arises is to give a meaning to $f(T)$ for a more general function $f : \mathbb{C} \to \mathbb{C}$. Consider the case where $f : \mathbb{C} \to \mathbb{C}$ is holomorphic on a set $D - \{z_0\}$, with $D \subset \mathbb{C}$ being open and connected. Then if the Laurent series of f around z_0 is $\sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$, we could define $f(T)$ by

$$f(T) = \sum_{n=-\infty}^{\infty} a_n (T - z_0 I)^n.$$

Hence for such a definition to make sense, when not all $a_{-m}, m \geq 1$ are zero, we need to ask for the operator $(T - z_0 I)$ to be invertible, and this motivates the definition of the spectrum of an operator.

Now let us continue our investigation of the spectrum. It is straightforward that the set of eigenvalues of an operator is contained, but need not be equal, to its spectrum. This will be illustrated in the next example:

Example 1.6.3 (from Young (1988))**.** Let $f \in L^2([0,1])$ be continuous, and let M_f be the induced multiplication operator on $L^2([0,1])$, defined in

example 1.3.6. Let $I = f([0,1])$, and let us show that $\sigma(M_f) = I$. Indeed, if $\lambda \notin I$, then $M_{\tilde{f}_\lambda}$ is the inverse of $M_f - \lambda I$, where $\tilde{f}_\lambda(t) = (f(t) - \lambda I)^{-1}$ is continuous. However, if $\lambda \in I$, then $(M_f - \lambda I)$ has no bounded inverse. Indeed, suppose such an inverse, which we will call T, exists. We know that $f(t_0) = \lambda$ for a $t_0 \in [0,1]$, thus if we define the intervals J_n about t_0 of length $\delta_n > 0$ such that $|f(t) - \lambda| < 1/n$ for $t \in J_n$, and define $g_n(t) = \delta^{-1/2}\chi_{J_n}(t)$, then $\|g_n\| = 1$ for all $n > 0$,

$$y_n := (M_f - \lambda I)g_n \to 0, \quad n \to \infty,$$

but $T(M_f - \lambda I)g_n = g_n$, which violates the continuity of T (recall that continuity is equivalent to boundedness for linear operators).

So we have that $\sigma(M_f) = f([0,1])$. However, if we take the particular case $f(t) = t$, then trying to solve the eigenvalue equation

$$tx(t) = \lambda x(t), \quad \lambda \in [0,1] = f([0,1]),$$

yields $x = 0$. Thus a bounded operator can have a spectrum but no eigenvalue. ◀

The following results about invertibility, taken from Halmos (1957), will help us better understand the spectrum:

Definition 1.6.4. An operator $T \in \mathcal{L}(B, B')$, where B and B' are Banach spaces, is said to be *bounded from below* if there exists a real $\alpha \geq 0$ that satisfies

$$\|Ax\|_{B'} \geq \alpha \|x\|_B, \quad \forall x \in B.$$

Proposition 1.6.5. *Let B, B' be Banach spaces and $T \in \mathcal{L}(B, B')$ be bounded from below. Then the image of T is closed.*

Proof. Let $y_n = Tx_n$ be any Cauchy sequence in $\mathrm{Im}(T) \subset B'$, and let y be its limit. We need to show that $y \in \mathrm{Im}(T)$.

From the Cauchy property, for any $\varepsilon > 0$, we have for n, m larger than $N = N_y(\varepsilon) > 0$ that

$$\varepsilon \geq \|y_n - y_m\| = \|T(x_n - x_m)\| \geq \alpha \|x_n - x_m\|,$$

and hence $(x_n)_n$ is Cauchy with $N_x(\varepsilon) = N_y(\varepsilon/\alpha)$. Therefore $(x_n)_n$ is Cauchy in B, with limit x. By the continuity of T,

$$y = \lim_n Tx_n = T\left(\lim_n x_n\right) = Tx \in \mathrm{Im}(T).$$

□

We can now give a characterization of invertibility for operators between Banach spaces:

Theorem 1.6.6 (Characterization of invertibility in Banach spaces)**.** *An operator $T \in \mathcal{L}(B, B')$, where B and B' are Banach spaces, is invertible if and only if the two following conditions hold:*

(i) T *is bounded from below*

(ii) $\mathrm{Im}(T)$ *is dense in* B'*.*

Proof. Suppose T is invertible. Then (ii) is trivial, and we only need to show (i). Let $S = T^{-1}$ be the inverse of T. For any $x \in B$, we have

$$\|x\| = \|STx\| \leq \|S\|\|Tx\|,$$

so taking $\alpha = \|S\|^{-1} < \infty$ yields (i).

Conversely, suppose (i) and (ii) hold. Then by the previous proposition, $\mathrm{Im}(T) = B'$, and if $Tx = 0$, then

$$0 = \|Tx\| \geq \alpha\|x\|$$

so $x = 0$ and T is injective. Hence T is bijective, and we can define $S : B' \to B$ by

$$Sy = x \text{ if, and only if } Tx = y, \quad \forall y \in B'.$$

S is the just the set-inverse of T. To show that it is the inverse of T, we need to show that it is linear and bounded. Linearity is straightforward to check, and its boundedness is a direct consequence of (i):

$$\|Sy\| = \|x\| \leq \frac{1}{\alpha}\|Tx\| = \frac{1}{\alpha}\|y\|$$

for all $y \in B'$, hence $\|S\| \leq \frac{1}{\alpha} < \infty$ and T is invertible, with $T^{-1} = S$. The proof is complete. □

Now we can apply these results to the spectrum. We know that $\lambda \in \sigma(A)$ if and only if $(A - \lambda I)$ is not invertible, which in the light of Theorem 1.6.6 occurs if and only if one of the two following conditions hold:

(i) $(A - \lambda I)$ does not have a dense range,

(ii) $(A - \lambda I)$ is not bounded from below.

So now we can classify the points of the spectrum:

Definition 1.6.7. For an operator $T \in \mathcal{L}(B)$ on a Banach space B, we define the *compression spectrum* $\Gamma(T)$ of T to be the set of the $\lambda \in \sigma(T)$ for which $(T - \lambda I)$ does not have a dense range, and we also define the *approximate point spectrum* $\Pi(T)$ to be the set of $\lambda \in \sigma(T)$ such that $(T - \lambda I)$ is not bounded from below.

Notice that $\lambda \in \Pi(T)$ if, and only if, there exists a sequence $(v_n)_n$ of unit vectors such that $(T-\lambda I)v_n \to 0$ when $n \to \infty$. There might exist a non-zero vector v such that $(T - \lambda I)v = 0$, in which case $Tv = \lambda v$ and the action of the operator T on v is very simple: it only scales v by a factor λ. Such a vector is called an *eigenvector* of T, and λ is the associated *eigenvalue*. The set of eigenvalues of T is sometimes called the *point spectrum* of T, and will be denoted by $\Pi_0(T)$. Notice that it is a subset of $\Pi(T)$. Figure 1.2 illustrates these concepts. ※

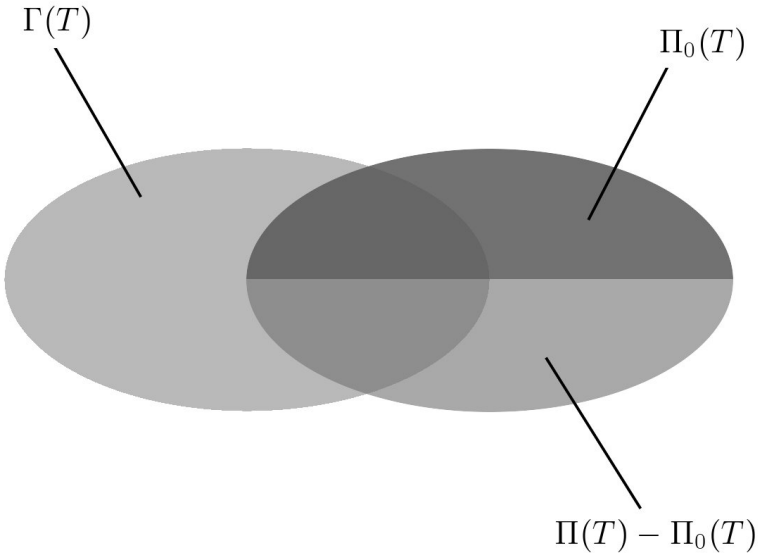

Figure 1.2: The spectrum of an operator T. $\Gamma(T)$ is the *compression spectrum* of T, $\Pi_0(T) \subset \Pi(T)$ is the set of *eigenvalues* of T, or *point spectrum* of T, and $\Pi(T)$ is the approximate point spectrum of T.

Here is a general fact about the spectrum:

Theorem 1.6.8. *If $T \in \mathcal{L}(B)$, then the spectrum $\sigma(T)$ is a compact subset of $\mathbb{F}$ contained in the closed disc of radius $\|T\|$ around the origin.*

Proof. See Young (1988, p.81) for a proof in the case $\mathbb{F} = \mathbb{C}$. The proof for the case $\mathbb{F} = \mathbb{R}$ is exactly the same. □

The Spectrum of Compact Operators

If we assume compactness of an operator, then we have the following nice result:

Theorem 1.6.9. *Let $T \in \mathcal{L}(H)$ be compact. Then*

$$\Pi_0(T) \supseteq \Pi(T) - 0, \tag{1.23}$$

that is, every non-zero element of the approximate point spectrum is in fact an eigenvalue.

Proof. By definition, $\Pi_0(T) \subset \Pi(T)$. Let $\lambda \in \Pi(T)$ be different from zero. We have to show that λ is an eigenvalue of T.

There exists a sequence $(x_n)_n$ of unit vectors such that $(T - \lambda I)x_n \to 0$. By the compactness of T, we can assume – without loss of generality – that Tx_n converges to y. But then, we must have that

$$\lambda \lim_n x_n = \lim_n Tx_n = y,$$

and thus $y \neq 0$ (recall that $\lambda \neq 0$ and that the x_n's are unit vectors). Now by continuity of T, we have

$$Ay = \lambda \lim_n Ax_n = \lambda y,$$

and hence λ is an eigenvalue of T, and the proof is complete. □

Compactness also gives a sort of bound on the cardinality of the spectrum:

Theorem 1.6.10. *If the spectrum $\sigma(T)$ of a compact operator $T \in \mathcal{L}(B)$ is infinite, then $\sigma(A)$ is countable and has exactly one accumulation point, namely, zero.*

Proof. See Dunford & Schwartz (1988, p.579). □

Since we already know that the spectrum is compact, the following result is direct:

Corollary 1.6.11. *For a compact operator, without loss of generality, we can suppose that $\sigma(T) = (\lambda_n)_n \cup \{0\}$ such that:*

(i) $|\lambda_{n+1}| \leq |\lambda_n| \leq \|T\|$, *for all* $n \in \mathbb{N}$,

(ii) $\lim_n \lambda_n = 0$

Let us now see some properties of the spectrum for normal operators.

The Spectrum of Normal Operators

For normal operators, we have the following result:

Theorem 1.6.12. *If $T \in \mathcal{L}(B)$ is normal, then $\Pi(T) = \sigma(T)$, that is, the spectrum is equal to the approximate point spectrum.*

Proof. See Halmos (1957, p.57). □

The direct consequence, using Theorem 1.6.9 is:

Corollary 1.6.13. *For a normal compact operator $T \in \mathcal{L}(H)$, we have*

$$\Pi_0(T) \supseteq \sigma(T) - 0.$$

Hence compact normal operators behave almost like finite dimensional operators, at least in terms of their spectrum: except for $\lambda = 0$, an element of the spectrum is the same as an eigenvalue.

We now consider the spectrum for self-adjoint operators.

The Spectrum of Self-Adjoint Operators

We know by Theorem 1.6.8 that

$$\sup\{|\lambda| : \lambda \in \sigma(T)\} = \max\{|\lambda| : \lambda \in \sigma(T)\} \leq \|T\|$$

for $T \in \mathcal{L}(B)$, where the supremum is a maximum because of the compactness of the spectrum. For self-adjoint operators, the bound is achieved:

Theorem 1.6.14. *Let $T \in \mathcal{L}(H)$ be self-adjoint. Then*

$$\max\{|\lambda| : \lambda \in \sigma(T)\} = \|T\|.$$

Proof. See Halmos (1957, p. 55) □

Now we know that the spectrum of a self-adjoint operator is never empty. Actually, we even have a stronger result if we use Theorem 1.6.12:

Corollary 1.6.15. *For a self-adjoint operator $T \in \mathcal{L}(H)$, we have the equality*

$$\max\{|\lambda| : \lambda \in \Pi(T)\} = \|T\|. \tag{1.24}$$

We also have another nice result:

Theorem 1.6.16. *The spectrum of a self-adjoint operator is a subset of $\mathbb{R}$.*

Proof. See Rudin (1991, p. 314) □

We will now look at the spectrum of compact and self-adjoint operators. This class of operators will be of great importance in this report.

The Spectrum of Compact Self-Adjoint Operators

Remark 1.6.17. In the latter, we shall use the abbreviation CSA for "Compact and Self-Ajoint". 🍃

Now we can merge our results to get:

Theorem 1.6.18. *Let $T \in \mathcal{L}(H)$ be CSA. Then*

(i) $\|T\|$ *or* $-\|T\|$ *is an eigenvalue of* T,

(ii) $\Pi_0(T) = \sigma(T) - \{0\}$,

(iii) Two eigenvectors corresponding to different eigenvalues of T *are orthogonal,*

(iv) When the spectrum of T *is infinite, we can suppose, without loss of generality, that* $\sigma(T) = \{(\lambda_n)_n\} \cup \{0\}$, *where the* λ_n*'s are eigenvalues of* T *and*

- *(a)* $|\lambda_0| = \|T\|$,
- *(b)* $|\lambda_{n+1}| \leq |\lambda_n| \leq \|T\|$, *for all* $n \in \mathbb{N}$, *and*
- *(c)* $\lim_n \lambda_n = 0$.

Proof. The statements $(i), (ii)$ and (iv) are direct consequences of the Theorems 1.6.14, 1.6.9, 1.6.12 and Corollary 1.6.11.

For the proof of (iii), let $\mu \neq \lambda$ be eigenvalues of T with eigenvectors v, w, respectively. By self-adjointness, we know that the eigenvalues are real numbers. Hence

$$\begin{aligned} 0 &= \langle Tv, w\rangle - \langle v, Tw\rangle \\ &= (\lambda - \bar{\mu})\langle v, w\rangle = (\lambda - \mu)\langle v, w\rangle, \end{aligned}$$

and the desired result is proved. □

So CSA operators are very well-behaved: if $T \in \mathcal{L}(H)$ is CSA, then there exists a (possibly finite) sequence $(\lambda_n, \varphi_n)_n$ where the λ_n's are eigenvalues with corresponding eigenvectors φ_n, and we can assume without loss of generality that all the eigenvectors are orthonormal (indeed, if they have the same eigenvalue, we can just orthogonalize them via the Gram-Schmidt procedure to make then orthonormal). Hence for any vector in the closed

subspace $\overline{\text{span}(\varphi_n)_n}$ the operator T acts in a very nice way, because we can write:

$$v = \sum_n \langle v, \varphi_n \rangle \varphi_n$$
$$Av = \sum_n \lambda_n \langle v, \varphi_n \rangle \varphi_n.$$

The natural question that arises now is wether this expression holds for all $v \in H$. The answer is yes, and it is given by the famous spectral Theorem for CSA operators:

Theorem 1.6.19 (Spectral Theorem for compact and self-adjoint operators)**.** *Let H be a Hilbert space, and $T \in \mathcal{L}(H)$ be a compact and self-adjoint operator. Then T has the expression*

$$Tv = \sum_n \lambda_n \langle v, \varphi_n \rangle \varphi_n, \quad \forall v \in H, \tag{1.25}$$

where $(\varphi_n)_n$ is a finite or infinite sequence or orthonormal eigenvectors with respective real *eigenvalues $(\lambda_n)_n$ that satisfy:*

(a) $|\lambda_0| = \|T\|$,

(b) $|\lambda_{n+1}| \leq |\lambda_n| \leq \|T\|$, *for all $n \in \mathbb{N}$, and*

(c) $\lim_n \lambda_n = 0$, *whenever the sequence $(\lambda_n)_n$ is infinite.*

Proof. See Young (1988, p.99). □

There is a nice and straightforward consequence of this Theorem:

Corollary 1.6.20. *Any CSA operator on a non-separable Hilbert space has a non-trivial kernel.*

1.7 The Singular Value Decomposition of Compact Operators

Let us briefly return to compact operators and state an important decomposition Theorem. Let H be a Hilbert space (supposed real for simplicity), and notice that for a compact operator $T : H \to H$, the operator T^*T is self-adjoint and compact. Indeed, if (x_n) is a bounded sequence in H, then (Tx_n) admits a convergent subsequence, and so does $(T^*(Tx_n))$ by continuity of T^*. Thus the spectral Theorem for compact self-adjoint operators yields

the decomposition $T^*T = \sum_j \lambda_j e_j \otimes e_j$, where $(e_j) \subset H$ is an orthonormal sequence, and $\lambda_j \neq 0, \forall j$. If we let $v_j = Te_j$, we notice that $\langle v_j, v_i \rangle = \lambda_j \delta_{ij}$. Let $S = \sum_j e_j \otimes v_j$, which is a well defined bounded operator on H because if $x \in H, \|x\| < 1$, then

$$Sx = \sum_j \langle e_j, x \rangle Te_j = T\left(\sum_j \langle e_j, x \rangle e_j\right),$$

and $\|S\| \leq \|T\|$. Then S equals T on $\overline{\text{span}(e_j)}$, and if $f \in \overline{\text{span}(e_j)}^{\perp}$, then $T^*Tf = 0$ and thus $0 = \langle T^*Tf, f \rangle = \langle Tf, Tf \rangle$, yielding that

$$\ker(T) = \overline{\text{span}(e_j)}^{\perp} = \ker(S).$$

Thus the operators S and T are equal. Eventually, letting $\mu_j = \|v_j\|$ and $f_j = v_j/\mu_j$, we have the following decomposition Theorem for compact operators:

Theorem 1.7.1 (Singular Value Decomposition for Compact Operators)**.** *Let H be a Hilbert space, and $T : H \to H$ a compact operator. Then there exists orthonormal sequences $(e_j), (f_j) \subset H$ and a decreasing sequence of positive real numbers (μ_j), called* singular values of T*, such that*

$$T = \sum_j \mu_j e_j \otimes f_j,$$

where the convergence is in operator norm. Furthermore, if the sequence (μ_j) is infinite, then $\lim_{j\to\infty} \mu_j = 0$.

Proof. The paragraph preceeding the Theorem contains the essential ideas, except for the last statement which is a consequence of Theorem 1.6.19 and the equality

$$\lambda_j = \langle T^*Te_j, e_j \rangle = \langle Te_j, Te_j \rangle = \|v_j\|^2 = |\mu_j|^2.$$

For a more rigorous proof of the Theorem, see Weidmann (1980, Theorem 7.6) for instance. □

Conclusions

In this section, we have seen how the concepts of vector space and linear transformation become more complicated in infinite dimensional spaces. We have seen that the "smallest" infinite dimensional Hilbert spaces (the separable Hilbert spaces) are like ℓ_2 spaces, and that the link is given by the

choice of a complete orthonormal sequence. This notion resembles that of an orthonormal basis for finite dimensional spaces, but not exactly, because of the use of the limit in the expressions $\sum_{n=0}^{\infty}\langle x, e_i\rangle e_i$. We have also seen several types of continuous operators: the compact operators, for which we have almost the same properties as for matrices. The important results of this section are the spectral Theorem for compact self-adjoint operators (Theorem 1.6.19), and the singular value decomposition (SVD) for compact operators (Theorem 1.7.1). Besides giving us a nice representations of operators, the SVD provides us with a canonical way of associating operators and sequences. For a compact operator $T = \sum_j \lambda_j e_j \otimes f_j$, the decreasing sequence (λ_j) is a sequence of numbers with limit 0 (if it is infinite). Furthermore, if T is Hilbert-Schmidt, the sequence (λ_j) will be in ℓ_2, and if T is trace-class, (λ_j) will be an ℓ_1 sequence.

2

Fundamentals of Probability in Abstract Spaces

The ultimate goal of the present report is to consider basic hypothesis tests for functional data. Often, functional data are assumed to be realisation of some random process taking values in a Hilbert space, such as $L^2([0,1])$ for instance, and we want to carry out inferences for this process. In other words, we would like to do statistics on function spaces (or more generaly on Banach or Hilbert spaces). However, we need first to understand how to put a handle on probability measures on such spaces. Especially, we need to understand what a Gaussian random variable in such spaces is.

In this section, we will first introduce the concepts of measure theory that are useful for the study of random processes, and give a Theorem about the existence of processes, due to Kolmogorov. We will then introduce Gaussian random vectors and some of their properties, before presenting Gaussian Processes. This way, we will see the analogies between the finite and infinite dimensional setup. Afterwards, basic notions of probabilities in Banach spaces will be presented, and selected limit Theorems for sequences of Banach space-valued random variables will be given. We will conclude with some remarks on the definition of Gaussian processes.

2.1 Some Necessary Measure Theory

In order to define the notion of random process, we need to understand what a measurable map $\Omega \to X$ is, where

$$X \subset \mathbb{R}^{[0,1]} := \{f : [0,1] \to \mathbb{R}\},$$

we need to construct a σ-algebra for $\mathbb{R}^{[0,1]}$. A way of doing this is by using ideas similar to the ones used in topology, which we present here.

Recall the definition of the product topology: let (S_i, τ_i) be topological spaces for $i \in I$. On the product space $S = \prod_{i \in I} S_i$ we define the topology τ generated by the sub-basis

$$\left\{ A_j \times \prod_{i \neq j} S_i : j \in I, A_j \in \tau_j \right\}.$$

This topology is called the *product topology*, and we have the following result:

Proposition 2.1.1. *Let A be a topological space, S as above, with the product topology, and let $f_i : A \to S_i$ be set functions.*

Then the mapping $f : A \to S$, defined by $f(a) = (f_i(a))_{i \in I}$, is continuous if, and only if $f_i : A \to S_i$ is continuous $\forall i \in I$.

Proof. See Munkres (2000) □

In a very similar way, we define the product metric space:

Definition 2.1.2. Let $(S_i, \mathcal{A}_i)_{i \in I}$ be measurable spaces, that is, $\mathcal{A}_i$ is a σ-algebra of subsets of S_i. A *column* on the cartesian product

$$S = \prod_{i \in I} S_i$$

is a set of the form

$$A_j \times \prod_{i \neq j} S_i, \quad A_j \in \mathcal{A}_j.$$

We will denote by $\mathcal{A} = \bigotimes \mathcal{A}_i$ the σ-algebra generated by all the columns on S. The pair $(S, \mathcal{A})$ is a measurable space, called the *product measurable space*.

When all the $(S_i, \mathcal{A}_i)$ are the same, say equal to $(S, \mathcal{A})$, then we write $S^I = \prod_{i \in I} S$ and the σ-algebra is $\mathcal{A}^I = \bigotimes_{i \in I} \mathcal{A}$. ※

Notice that this is very similar to the product topology defined above. We also have a result similar to Proposition 2.1.1:

Proposition 2.1.3. *Let Ω be a measurable space, S and the S_i's as in the previous definition, and let $f_i : \Omega \to S_i$ be set functions.*

Then, the mapping $f : \Omega \to S$, defined by $f(\omega) = (f_i(\omega))_{i \in I}$, is $\bigotimes_{i \in I} \mathcal{A}_i$-measurable if, and only if $f_i : A \to S_i$ is measurable $\forall i \in I$.

Proof. See Kallenberg (1997, p.4) □

The following proposition tells us what a measurable set looks like in the product measurable space:

Proposition 2.1.4. *We use here the notation of definition 2.1.2. Let $A \subset S$.*

Then $A \in \bigotimes_{i\in I} \mathcal{A}_i$ if, and only if $A = E \times \prod_{i\in I\setminus J} S_i$, with $J \subset I$ countable and $E \in \bigotimes_{i\in J} \mathcal{A}_i$.

Proof. See Halmos (1974*b*, p.158). □

Our goal here is to define properly what a random process is in order to better understand random elements of abstract spaces. A concrete example of a process is standard Brownian motion, which is a special measurable mapping $\Omega \to \mathbb{R}^{[0,\infty)} = \{f : [0,\infty) \to \mathbb{R}\}$. So a process is a measurable mapping from a probability space Ω into a product measurable space. From now on, we will write T for the index set, instead of I. Secretely, we are thinking of T as of some time index, but we will not assume any kind of structure on T unless specifically mentionned.

Remark 2.1.5. From now on, we will assume that Ω is a probability space with σ-algebra $\mathcal{O}$ and probability measure $\mathbb{P}$.

Definition 2.1.6 (Proposition)**.** Let $(S, \mathcal{A})$ be a measurable space, and T be an index set. A mapping

$$X : \Omega \to U \subset S^T \tag{2.1}$$

is $U \cap \mathcal{A}^T$-measurable if, and only if $X_t : \Omega \to S$ is $\mathcal{A}$-measurable for all $t \in T$, where $X_t = \pi_t \circ X$, with $\pi_t : S^T \to S$ the projection on the t^{th} coordinate.

Such a map is called an S-valued (random) *process* on T with paths in U, or just a random process on T when $S = \mathbb{R}$ and $U = \mathbb{R}^T$.

Proof. See Kallenberg (1997, p.24). □

For any random function $\xi, \eta : \Omega \to S$, we will write $\xi \overset{d}{=} \eta$ to say that ξ and η are *equal in distribution*, that is, $\mathbb{P} \circ \xi^{-1} = \mathbb{P} \circ \eta^{-1}$. For a process $X : \Omega \to S^T$, the *finite dimensional distributions* are given by

$$\{\mathbb{P} \circ (X_{t_1}, \ldots, X_{t_n}) : \ t_1, \ldots, t_n \in T, n \in \mathbb{N}\}. \tag{2.2}$$

The following result is crucial, because it tells us that it is enough to know all finite dimensional distributions of a process in order to define uniquely its distribution:

Proposition 2.1.7. *Let S, T, U be as in Definition 2.1.6. Let X, Y be S-valued processes on T with paths in U.*

Then $X \overset{d}{=} Y$ if, and only if all their finite dimensional distributions agree.

Proof. See Kallenberg (1997, p.25). □

2.2 Existence of Processes

Now we know how to specify the distribution of a process. However, we don't know if, given a collection of finite dimensional distributions, there exists a process with those finite dimensional distributions. We are going to give an answer to this question.

Notation 2.2.1. *Let T be an index set. We will denote by $\hat{T}$ the set of all finite subsets of T.*

For $I \in \hat{T}$, let $S_I = \prod_{i\in I} S_i$, $\mathcal{A}_I = \bigotimes_{i\in I} \mathcal{A}_i$, and let ξ_I be the restriction of the process $(\xi_t)_{t\in T}$ to I.

For $I, J \in \hat{T}$, $I \subset J$, define the coordonate-wise projections

$$\pi_J : S_T \to S_J$$

and

$$\pi_I^J : S_J \to S_I.$$

If a measurable map $\xi : \Omega \to S_T$ has probability distribution μ, then it is straightforward that the restrictions $\mu_J = \mu \circ \pi_J^{-1}$, $J \in \hat{T}$ satisfy the following property:

Definition 2.2.2. A family of probability measures $\mu_J : S_J \to [0,1], J \in \hat{T}$ is called a *projective family* if

(2.3)
$$\mu_I = \mu_J \circ (\pi_I^J)^{-1}, \quad \forall I, J \in \hat{T}, I \subset J. \qquad ※$$

A projective family is a family of "compatible" measures, in the sense that the measures do not contradict themselves. The existence Theorem we aim at involves some kind of topological structure on the measurable spaces S_t.

Definition 2.2.3. Two measurable spaces S and T are called *Borel isomorphic* if there exists a bijection $f : S \to T$ such that f and f^{-1} are both measurable.

A measurable space is called a *Borel space* if it is Borel isomorphic to a Borel subset of $[0,1]$. ※

Now we can state an existence Theorem for processes:

Theorem 2.2.4 (Existence of processes, Kolmogorov)**.** *For any collection of Borel Spaces $S_t, t \in T$, and any projective family of probability measures μ_I on $S_I, I \in \hat{T}$, there exists some measurable function ξ_t on S_t, for all $t \in T$, such that ξ_I has distribution μ_I, for all $I \in \hat{T}$.*

Proof. See Kallenberg (1997, p.92) □

2.3 Gaussian Random Vectors

Before introducing Gaussian random processes, let us recall some facts about Gaussian random vectors. This will then allow us to see the analogies between the Gaussian random vectors – which are random variables taking values in a *finite dimensional* space – and Gaussian random processes, that are random variables taking values in an *infinite* dimensional space.

First, let us introduce some notation:

Notation 2.3.1. *A vector $a \in \mathbb{R}^n$ will always be thought of as a* column *vector with components $a_i \in \mathbb{R}$, that is,*

$$a = (a_1, \dots, a_n)^\mathsf{T} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n.$$

Hence if we write $x, y \in \mathbb{R}^n$, the product $x^\mathsf{T} y$ is a real number and xy^T is a $n \times n$ matrix. Explicitly,

$$x^\mathsf{T} y = (x_1, \dots, x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{j=1}^{n} x_j y_j$$

and

$$xy^\mathsf{T} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} (y_1, \dots, y_n) = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & \ddots & \vdots \\ x_n y_1 & \cdots & x_n y_n \end{pmatrix}$$

Before defining Gaussian processes, we are now going to introduce Gaussian random variables and vectors, and some of their properties. The reader interested in the proofs of the following statements may refer to any mathematical statistics textbook, like Roussas (1997) for instance.

A random variable X *Gaussian* with mean μ and variance σ^2 is a real valued random variable with density $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, or equivalently, with characteristic function $\varphi(t) = \mathbb{E}e^{itX} = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$, where $i \in \mathbb{C}$. We usually write $X \sim \mathcal{N}(\mu, \sigma^2)$.

A *Gaussian random vector* is a random vector $X = (X_1, \dots, X_n)^\mathsf{T}$ such that the random variable $a^\mathsf{T} X$ is Gaussian for all vectors $a \in \mathbb{R}^n$. An equivalent condition is to ask for its characteristic function to be

$$\varphi(t) = e^{it^\mathsf{T}\mu - \frac{1}{2} t^\mathsf{T} \Sigma t}, \quad t \in \mathbb{R}^n,$$

where $\mu \in \mathbb{R}^n$, and Σ is a $n \times n$ symmetric, nonnegative-definite matrix (that is, $\Sigma = \Sigma^\mathsf{T}$, and for all vectors $t \in \mathbb{R}^n$, $t^\mathsf{T}\Sigma t \geq 0$). The vector $\mu \in \mathbb{R}^n$ is the mean of the random vector, $\mu = \mathbb{E}X$, and the matrix Σ is the covariance matrix defined by $\Sigma_{ij} = \operatorname{cov}(X_i, X_j) = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]$. We write $X \sim \mathcal{N}(\mu, \Sigma)$.

Notice that the knowledge of the matrix Σ and the mean μ determines entirely the distribution of X. Also, if A is a $k \times n$ matrix, then $AX \sim \mathcal{N}(A\mu, A\Sigma A^\mathsf{T})$. Each coordinate X_j is therefore Gaussian, and for $j \neq l$, X_j and X_l are independent if, and only if $\operatorname{cov}(X_j, X_l) = 0$.

Notice also that the covariance matrix Σ can be thought as an operator on $\mathbb{R}^n$, acting by left-multiplication and sending $t \in \mathbb{R}^n$ to $\Sigma t \in \mathbb{R}^n$. Now because Σ is symmetric, the associated operator is self-adjoint, and thus the spectral Theorem (for *finite*-dimensional vector spaces) tells us that there are n orthogonal eigenvectors $\varphi_1, \ldots, \varphi_n \in \mathbb{R}^n$ of Σ, with *non-negative* eigenvalues $\lambda_1, \ldots, \lambda_n \geq 0$. Letting Λ being the diagonal matrix with entries $\Lambda_{jj} = \lambda_j$ and U being the orthogonal matrix with the eigenvector φ_j at column j, we have the decomposition $\Sigma = U\Lambda U^\mathsf{T}$. This gives the following decomposition of a Gaussian random vector:

Proposition 2.3.2 (Orthogonal decomposition)**.** *Let $X, \mu, \Sigma, \Lambda, U$ be as in the last paragraph. Then $X \sim \mu + UZ$, where $Z = (Z_1, \ldots, Z_n)$ and the Z_j's are independent Gaussian random variables with variances λ_j. That is, any Gaussian random vector is obtained by rotation and translation of independent Gaussian random variables.*

Proof. Let $Y = X - \mu$, and notice that $Y \sim \mathcal{N}(0, \Sigma)$. Now the decomposition $\Sigma = U\Lambda U^\mathsf{T}$ yields directly $Y \sim U\mathcal{N}(0, \Lambda)$, and the proof is complete. □

Now that we have seen some properties of Gaussian random vectors, let us introduce their infinite dimensional analogue.

2.4 Gaussian Processes

In this section we are going to define what a Gaussian process on $T = [0, 1]$ is.

Definition 2.4.1. A *Gaussian Process* on $[0, 1]$ with paths in $\mathcal{L}^2([0, 1]) = \mathcal{L}^2([0, 1], \mathbb{R})$ is a $\mathbb{R}$-valued random process X (in the sense of definition 2.1.6) with

1. $\mathbb{E}X_t^2 < \infty, \forall t \in [0, 1]$,

2. $X(\omega) \in \mathcal{L}^2([0,1])$ for all $\omega \in \Omega$, and all its finite dimensional distributions are Gaussian. That is, the real-valued random variables

$$\omega \mapsto \sum_{i=1}^{n} \alpha_i X_{t_i}(\omega)$$

are Gaussian, for all $n = 1, 2, \ldots, \alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $t_1, \ldots, t_n \in [0,1]$.

We can then define $m(t) = \mathbb{E}X_t$ the mean function of X and

$$r(s,t) = \operatorname{cov}(X_s, X_t),$$

the covariance function of X. ※

Before going farther in the presentation of Gaussian processes, let us introduce the commonly used notation for process:

Notation 2.4.2. *Let $\pi_t : \mathbb{R}^{[0,1]} \to \mathbb{R}$ be the projection on the coordinate t, that is, $\pi_t(f) = f(t)$, where $f \in \mathbb{R}^{[0,1]}$. Then we write X_t for the random variable*

$$X_t = \pi_t \circ X : \Omega \to \mathbb{R},$$

and hence, if we fix $t \in [0,1]$, the random process X induces a real random variable X_t.

However, if we fix $\omega \in \Omega$, then $t \mapsto X_t(\omega)$ is an $\mathcal{L}^2([0,1])$function. This latter representation allows us to think of the Gaussian process as defining random elements in $\mathcal{L}^2([0,1])$.

We will assume, without loss of generality, that $m(t) = 0$ for all $t \in [0,1]$, otherwise just define the process $Y_t = X_t - m(t)$, which has mean zero. We shall call such process a *zero-mean* or *centered* process. In this case the covariance function simplifies to $r(s,t) = \mathbb{E}[X_s X_t]$. In order for things to be properly defined, we will assume, from now on, that the covariance function $r(s,t)$ satisfies the condition

$$t \mapsto r(t,t) \text{ is an } \mathcal{L}^2([0,1]) \text{ function.} \tag{2.4}$$

Hence we directly have

$$\mathbb{E}\left[\int_0^1 X_t^2 \mathrm{d}t\right] = \int_0^1 \mathbb{E}X_t^2 \mathrm{d}t = \int_0^1 r(t,t) \leq \infty$$

and also that $t \mapsto r(s,t)$ is an $\mathcal{L}^2([0,1])$ function for all $s \in [0,1]$, for

$$r(s,t) = \mathbb{E}[X_s X_t] \leq \sqrt{\mathbb{E}[X_s^2]\,\mathbb{E}[X_t^2]} \leq \sqrt{r(s,s)}\sqrt{r(t,t)}$$

and $[0,1]$ is compact. Notice that the condition (2.4) is satisfied whenever $r(s,t)$ is a continuous function.

We have the following basic property of the covariance function:

Proposition 2.4.3. *The covariance function $r(s,t)$ is symmetric and non-negative definite, which means that for all $f \in L^2([0,1])$,*

$$\iint_{[0,1]^2} f(s)f(t)r(s,t)\mathrm{d}s\mathrm{d}t \geq 0. \tag{2.5}$$

Proof. The symmetry of r comes from the symmetry of the covariance. For the non-negative definiteness, take $f \in L^2([0,1])$, and notice that the integral (2.5) is absolutely convergent, hence by Fubini's Theorem,

$$\begin{aligned}\iint_{[0,1]^2} f(s)f(t)r(s,t)\mathrm{d}s\mathrm{d}t &= \iint_{[0,1]^2} f(s)f(t)\mathbb{E}\left[X_sX_t\right]\mathrm{d}s\mathrm{d}t \\ &= \mathbb{E}\left[\int_0^1 f(s)X_s\mathrm{d}s\int_0^1 f(t)X_t\mathrm{d}t\right] \geq 0.\end{aligned}$$

□

So for any process X satisfying (2.4), the covariance function of X is symmetric and non-negative definite real function of two variables. Now a question we *should* ask ourselves is whether, given a symmetric, non-negative function $r(s,t)$ safisfying (2.4), there exists a *Gaussian* process with mean zero and covariance function $r(s,t)$. The answer is yes:

Proposition 2.4.4. *For any $r \in L^2([0,1]\times[0,1])$ symmetric and non-negative definite satisfying (2.4), there exists a mean-zero Gaussian process X with values in $\mathcal{L}^2([0,1])$ with covariance function r.*

Proof. First, notice that the covariance function $r(s,t)$ entirely determines the finite dimensinal distributions of a Gaussian process X with mean zero, and this fact tells us that the $r(s,t)$ defines a projective familly of Gaussian measures (with mean zero) on the finite dimensional distributions. Hence by the Kolmogorov existence Theorem 2.2.4, there exists a Gaussian process with mean zero and covariance $r(s,t)$. □

Now let us give the most famous example of Gaussian process, Brownian motion.

Example 2.4.5 (Brownian motion)**.** *Brownian motion*, or the *Wiener process* on $[0,1]$ is an $\mathbb{R}$-valued random process $(W_t, t \in [0,1])$ such that

(i) $W_0 = 0$,

(ii) $\mathbb{P}\{\omega : t \mapsto W_t(\omega) \text{ is continuous}\} = 1$,

(iii) W_t has independent increments, which means that $0 \leq t_1 < t_2 \leq t_3 < t_4 \leq 1$ implies that $W_{t_4} - W_{t_3}$ and $W_{t_2} - W_{t_1}$ are independent,

(iv) $W_t - W_s \sim \mathcal{N}(0, t-s)$ for all $0 \leq s < t \leq 1$.

Let us see that Brownian motion is a Gaussian process. Take $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $t_1, \ldots, t_n \geq 0$. Without loss of generality, we can assume that $t_i < t_{i+1}$. We need to show that the real-valued random variable $Z = \sum_i \alpha_i W_{t_i}$ is Gaussian, and that $\mathbb{E}\left[\int_0^1 W_t^2 \mathrm{d}t\right] < \infty$. Rewriting Z yields

$$Z = \sum_{i=1}^{n} (\alpha_i + \alpha_{i+1} + \cdots + \alpha n)(W_{t_i} - W_{t_{i-1}}),$$

where $t_0 = 0$. But this is a sum of independent Gaussians, and hence is Gaussian. Now we compute

$$\mathbb{E}\left[W_t^2\right] = \mathbb{E}\left[W_t^2\right] = t < \infty,$$

and a simple computation shows that $r(s,t) = \min(s,t)$ is continuous. We have shown that Brownian motion is a Gaussian process. ◀

We would now like to have a result similar to the orthogonal decomposition for Gaussian random vectors given in Proposition 2.3.2. Recall that this decomposition was obtained by looking at the covariance matrix Σ as an *operator*. Hence it is natural to do the same here:

Definition 2.4.6 (Covariance operator)**.** Let $r(s,t) = \mathbb{E}[X_s X_t]$ be the covariance function of a centered (Gaussian) process X. Then we define the *covariance operator* $R_X : L^2([0,1]) \to L^2([0,1])$ by

$$R_X f(s) = \int_0^1 r(s,t) f(t) \mathrm{d}t.$$

※

We will omit the subscript X and just write R for the covariance operator of X, whenever this doesn't create any confusion. The next proposition gives some properties of the covariance operator. It turns out that the covariance operator is very well behaved:

Proposition 2.4.7. *The covariance operator R is a bounded, linear, self-adjoint, compact, Hilbert-Schmidt and nuclear operator.*

Proof. Self-adjointness is a direct consequence of r being symmetric and taking real values. That the covariance operator is Hilbert-Schmidt is a consequence of R being an integral operator (see Theorem 1.5.11) and hence we also have compactness. Now nuclearity is given by Corollary 2.4.9, which is the consequence of the following important Theorem. □

Lemma 2.4.8 (Mercer's Theorem)**.** *Let $r(s,t)$ be a continuous on $[0,1]\times[0,1]$. Then there exists a sequence (φ_n) of* continuous *functions over $[0,1]$, and a decreasing sequence (λ_n) of positive numbers such that*

$$r(s,t)=\sum_{n=0}^{\infty}\lambda_n\varphi_n(s)\varphi_n(t),\quad s,t\in[0,1], \tag{2.6}$$

where the series converges uniformily and absolutely on $[0,1]\times[0,1]$. Moreover, the sequence (φ_n) is orthogonal, in the sense that

$$\int_0^1\varphi_n(s)\varphi_m(s)\mathrm{d}s=\delta_{n,m}=\begin{cases}1 & \text{if } n=m,\\ 0 & \text{if } n\neq m\end{cases} \tag{2.7}$$

and (φ_n) is a sequence of eigenvectors of the covariance operator, with associated eigenvalues (λ_n). That is,

$$\int_0^1 r(s,t)\varphi_n(s)\mathrm{d}s=\lambda_n\varphi_n(t),\quad t\in[0,1], n\in\mathbb{N}. \tag{2.8}$$

Proof. See Zaanen (1953, p.534) □

Corollary 2.4.9. *The covariance operator R is a nuclear operator.*

Proof. Let (λ_n) and (φ_n) be as in Mercer's Theorem, and notice that

$$\lambda_n=\int_0^1\lambda_n\varphi_n(s)\varphi_n(s)\mathrm{d}s,$$

and hence, because the convergence of (2.6) is absolute, by the bounded convergence Theorem:

$$\sum_{n\in\mathbb{N}}\lambda_n=\int_0^1\sum_{n\in\mathbb{N}}\lambda_n\varphi_n(s)\varphi_n(s)\mathrm{d}s=\int_0^1 r(s,s)\mathrm{d}s<\infty.$$ □

This result holds also without the assumption of continuity of the covariance function $r(s,t)$, and will be shown in Proposition 2.5.9.

We may now state a decomposition Theorem for Gaussian Processes, analogous to the decomposition for Gaussian random vectors (given by Proposition 2.3.2). In fact, a similar Theorem holds for more general processes, as will be seen in Remark 2.4.11.

Theorem 2.4.10 (Karhunen-Loève (K.L.) expansion)**.** *Let* $X = (X_t, t \in [0,1])$ *be a Gaussian process with* continuous *covariance function* r*. Then*

$$X_t = \sum_{n=0}^{\infty} z_n \varphi_n(t), \quad t \in [0,1], \tag{2.9}$$

where the convergence is uniform in $L^2(\Omega, \mathcal{O}, \mathbb{P})$*,*

$$\sup_{t\in[0,1]} \mathbb{E}\left(X_t - \sum_{k=0}^{n} z_k \varphi_k(t)\right)^2 \overset{n\to\infty}{\longrightarrow} 0,$$

and where (z_n) *is a sequence of real zero-mean random variables satisfying*

$$\mathbb{E}[z_n z_m] = \lambda_n \delta_{n,m}, \qquad n, m \in \mathbb{N}.$$

The (φ_n) *and* (λ_n) *are given by Mercer's Theorem, and* (z_n) *is given by*

$$z_n = \int_0^1 X_t \varphi_n(t) \mathrm{d}t \quad (a.s.), n \in \mathbb{N}. \tag{2.10}$$

Moreover, the z_n*'s are independent Gaussian variables.*

Proof. See Bosq (2000, p.25) for a proof a the Theorem. The last statement is a consequence of Theorem 2.7.6. □

Remark 2.4.11. The Karhunen-Loève (K.L.) expansion also holds with the only assumption that the covariance function $r(s,t)$ is continuous (Gihman & Skorohod 1974, p.226). We then have the same decomposition

$$X_t = \sum_{n=0}^{\infty} z_n \varphi_n(t), \quad t \in [0,1],$$

where the convergence is uniform in $L^2(\Omega, \mathcal{O}, \mathbb{P})$, except that the z_n's are just uncorrolated, and not necessarily independent.

Now let us show an application of the K.L. expansion to Brownian motion:

Example 2.4.12. Recall that Brownian motion $(W_t, t \in [0,1])$, defined in example 2.4.5, is a Gaussian process with covariance function

$$r(s,t) = min(s,t).$$

We will now give the K.L. expansion of Brownian motion. First, we need to solve the equation

$$\lambda \varphi(t) = \int_0^1 r(s,t)\varphi(s)\mathrm{d}s,$$

which simplifies to

$$\lambda\varphi(t) = \int_0^t s\varphi(s)\mathrm{d}s + t\int_t^1 \varphi(s)\mathrm{d}s.$$

Now we differentiate twice with respect to t and obtain the second order differential equation

$$\lambda\varphi''(t) = -\varphi(t)$$

with general solution $\varphi(t) = A\cos(\eta t) + B\sin(\eta t)$, with $\eta = 1/\sqrt{\lambda}$. With the initial conditions

$$\varphi(0) = \frac{1}{\lambda}\int_0^1 r(s,0)\varphi(s)\mathrm{d}s = 0$$
$$\varphi'(0) = \frac{1}{\lambda}\int_1^1 \varphi(s)\mathrm{d}s = 0,$$

we find

$$\lambda_n = \left(\frac{2}{\pi(2n+1)}\right)^2, \quad n \in \mathbb{N}$$

and

$$\varphi_n(t) = \sqrt{2}\sin\left(\frac{\pi(2n+1)}{2}t\right).$$

Thus the K.L. expansion of Brownian motion on $[0,1]$ is given by

$$W_t = \sum_{n\in\mathbb{N}} z_n\sqrt{2}\sin\left(\frac{\pi(2n+1)}{2}t\right), \tag{2.11}$$

with uniform convergence in $L^2(\Omega, \mathcal{O}, \mathbb{P})$, and where

$$z_n = \int_0^1 W_t\sqrt{2}\sin\left(\frac{\pi(2n+1)}{2}t\right)\mathrm{d}t$$

are independent real-valued Gaussian random variables with mean zero and variance λ_n. ◀

Notice the strength of the K.L. expansion: it can be used to see a Gaussian process as a sum of independent Gaussian random variables, or as an (infinite!) linear combination of continuous functions (φ_n) with random independent Gaussian coefficients. The K.L. expansion can also be used for generating samples of a Gaussian process: for a given continuous covariance function r, provided the φ_n's and λ_n's are known, one only needs to generate

a (large) number N of independent $\mathcal{N}(0,1)$ samples $x_1, \dots, x_N$ in order to obtain an approximate sample

$$W_t(\omega) \approx \sum_{n=1}^{N} \sqrt{\lambda_n} x_n \varphi_n(t). \tag{2.12}$$

However, the problem is that in general, it is impossible to solve the integral equations explicitly so as to find the φ_n's. Nevertheless, the K.L. expansion allows us to *construct* a random (Gaussian) process with desirable smoothness properties.

2.5 Basic Notions of Probabilities in Banach Spaces

Now that we have defined and given examples of Gaussian Processes, let us briefly introduce some basic notions of probability in Banach spaces, following the exposition of Bosq (2000). These notions will be useful for understanding the test for functional data that will be presented in Section 5, and are necessary for the exposition of the stochastic convergence Theorems in Banach spaces, that will be reviewed in the next section.

First, we look at what we already know: for a probability space $(\Omega, \mathcal{F}, P)$, we define a random vector as a measurable function $\Omega \to \mathbb{R}^n$. Now we could imagine doing the same with a Banach space B and its Borel $\sigma-$algebra, but this could lead to some complications (see Ledoux & Talagrand (1991, chapter 2) for instance). However, such problems are solved if we consider only separable Banach spaces:

Definition 2.5.1. Let B be a separable Banach space with norm $\|\cdot\|$ and Borel σ-algebra $\mathcal{B}_B$. Then a *B-valued random variable* (or *B-random variable*) defined on a probability space $(\Omega, \mathcal{F}, P)$ is a measurable function

$$(\Omega, \mathcal{F}) \to (B, \mathcal{B}_B).$$

We will often write B-r.v. instead of B-random variable. ※

Now we will give a criterion for a mapping $(\Omega, \mathcal{F}) \to (B, \mathcal{B}_B)$ to be a B-r.v. Notice that the so-called cylindrical σ-algebra

$$\mathcal{C}_B = \sigma(x^* : x^* \in B^*)$$

is always included in the Borel σ-algebra: $\mathcal{C}_B \subseteq \mathcal{B}_B$, because $x^{*-1}(]a, b[)$ is open in B for any $a < b$ real. The separability of B actually implies that the two σ-algebras are equal, and we have:

Proposition 2.5.2 (Characterization of B-random variables)**.** *Let B be a separable Banach space. Then $X : (\Omega, \mathcal{F}) \to (B, \mathcal{B}_B)$ is a B-random variable if, and only if $x^*(X)$ is measurable for all $x^* \in B^*$, where B^* is the topological dual of B.*

This allows us to define integrals of B-random variables:

Definition 2.5.3. A B-r.v. X is said to be *weakly integrable* if $x^*(X)$ is integrable for all $x^* \in B^*$, and if there exists an element $E_X \in B$ such that

$$\mathbb{E}x^*(X) = x^*(E_X), \quad \forall x^* \in B^*. \tag{2.13}$$

The element E_X is called the *weak expectation* or *weak integral* of X, and is usually written $\mathbb{E}X$. ※

A stronger notion is the following:

Definition 2.5.4. A B-r.v. X is said to be *integrable* (or strongly integrable) if $\mathbb{E}\|X\| < \infty$. ※

As one might guess, strong integrability implies the weak one.

Proposition 2.5.5. *For any integrable B-r.v. X, there exists a sequence $(X_n)_n$ of simple B-r.v. such that*

$$\lim_{n\to\infty} \mathbb{E}\|X - X_n\| = 0,$$

and X is weakly integrable.

Proof. See (Bosq 2000, p.28). □

Hence the following definition makes sense:

Definition 2.5.6. If X is an integrable B-r.v. then $\mathbb{E}X$ is called the *integral* or *expectation* of X and we also denote it by $\int X \mathrm{d}\mathbb{P}$. ※

What does this definition tell us? If we look at random vectors, it coincides with the usual definition of the expectation. So it is simply a generalisation of the notion of expectation to Banach spaces. Notice that the condition (2.13) is a commutativity condition, with our new notation:

$$\mathbb{E}x^*(X) = x^*\left(\mathbb{E}X\right), \quad \forall x^* \in B^*.$$

In order to be able to define what a Gaussian B-r.v. is, we need to define the covariance operator:

Definition 2.5.7. Let X be a B-r.v. such that $\mathbb{E}\|X\|^2 < \infty$ and $\mathbb{E}X = 0$. The *covariance operator* of X is a bounded linear operator $C_X : B^* \to B$ defined by

$$C_X(x^*) = \mathbb{E}\left[x^*(X)X\right], \quad x^* \in B^*.$$

If $\mathbb{E}X \neq 0$, replace X by $(X - \mathbb{E}X)$ in the present definition. ※

In the case $B = H$, a separable Hilbert space, the covariance operator can be written using a tensor product:

Remark 2.5.8. Let H be a separable Hilbert space, and X be an H-valued random variable such that $\mathbb{E}\|X\|^2 < \infty$ and $\mathbb{E}X = 0$. The covariance operator C_X of X can be written as $C_X = \mathbb{E}\left[X \otimes X\right]$, which is defined by

$$\mathbb{E}\left[X \otimes X\right](f) = \mathbb{E}\left[\langle X, f\rangle X\right], \quad f \in H,$$

because of the identification of H^* and H (via a linear isometry). If $\mathbb{E}X \neq 0$, then replace X by $(X - \mathbb{E}X)$ in this formula.

This way, C_X is a linear operator on H, and it is in fact nuclear, as will be shown in the next proposition:

Proposition 2.5.9. *Let H be a separable Hilbert space, and X be an H-valued random variable. If $\mathbb{E}\|X\|^2 < \infty$, then the covariance operator C_X of X is nuclear, and*

$$\operatorname{trace}(C_X) = \mathbb{E}\|X - m\|^2,$$

where $m = \mathbb{E}X$. Furthermore, $\langle C_X f, f\rangle \geq 0$ for all $f \in H$, that is, the covariance operator is non-negative.

Proof. Let (e_i) be a complete orthonormal sequence in H. First, assume that $m := \mathbb{E}X = 0$, then $C_X = \mathbb{E}\left[X \otimes X\right]$ and we have

$$\begin{aligned}
\operatorname{trace}(C_X) &= \sum_i \langle C_X e_i, e_i\rangle \\
&= \sum_i \langle \mathbb{E}\left[\langle X, e_i\rangle X\right], e_i\rangle \\
&= \sum_i \mathbb{E}\langle \langle X, e_i\rangle X, e_i\rangle \\
&= \sum_i \mathbb{E}\langle X, e_i\rangle^2 \\
&= \mathbb{E}\sum_i \langle X, e_i\rangle^2 \\
&= \mathbb{E}\|X\|^2 < \infty,
\end{aligned}$$

where the permutation of the sum and the expectation is allowed because the summands are positive. For the case $m \neq 0$, just repeat the same steps with $X - m$ instead of X, which yields

$$\operatorname{trace}(C_X) = \mathbb{E}\|X - m\|^2 < \infty.$$

To show that C_X is non-negative, first suppose that $m = 0$: let $f \in H$ and notice that

$$\begin{aligned} \langle C_X f, f\rangle &= \langle \mathbb{E}\left[\langle X, f\rangle X\right], f\rangle \\ &= \mathbb{E}\left[\langle X, f\rangle^2\right] \geq 0, \end{aligned}$$

where the second equality is justified by (2.13) because

$$\mathbb{E}\,\|\langle X, f\rangle X\| \leq \|f\| \cdot \mathbb{E}\|X\|^2 < \infty.$$

For the case $m \neq 0$, replacing X by $(X - m)$ in each step yields the same result. □

Remark 2.5.10. The covariance operator is entirely determined by the *covariance function* $c_X : B^* \times B^* \to \mathbb{R}$ defined by

$$c_X(x^*, y^*) = \mathbb{E}\left[x^*(X)y^*(X)\right], \quad x^*, y^* \in B^*.$$

Now let us define the analogue of the characteristic functional for B-r.v.

Definition 2.5.11. Let X be a B-r.v. such that $\mathbb{E}\|X\|^2 < \infty$. The *characteristic functional* of X is a mapping $\varphi_X : B^* \to \mathbb{C}$ defined by

$$\varphi_X(x^*) = \mathbb{E}\left[e^{ix^*(X)}\right], \quad x^* \in B^*.$$

※

The importance of the characteristic functional of a B-r.v. is that it entirely determines its distribution. We can now give the definition of a Gaussian B-r.v.

Definition 2.5.12. A B-r.v. X is said to be *Gaussian* if $\mathbb{E}\|X\|^2 < \infty$ and

$$\varphi_X(x^*) = \exp\left(ix^*(X) - \frac{1}{2}x^*\left[C_X(x^*)\right]\right). \tag{2.14}$$

※

We see that in the case $B = \mathbb{R}^d$, it corresponds to our familiar notion of characteristic function.

2.6 Some Stochastic Convergence Theorems in Banach Spaces

In the study of random variables, the (Strong) Law of Large Numbers ((S)LLN) and the Central Limit Theorem (CLT) are of particular importance, for they give the asymptotic behaviour of

$$S_n = \sum_{k=1}^{n} X, \tag{2.15}$$

for i.i.d. random variables $X_1, \ldots, X_n$. Let us give in this section the analogues for B-r.v. We first start with a weak law of large numbers. But first, let us introduce some notation:

Notation 2.6.1. *For a random variable $X : \Omega \to H$, where H is a separable Hilbert space, let us define its $L^2(\mathbb{P})$-norm*

$$\|X\|_{L^2(\mathbb{P})} = \sqrt{\mathbb{E}\left[\|X\|^2\right]}, \tag{2.16}$$

where $\|X\|^2 = \langle X, X\rangle$ is the norm in H.

Let us also define weak convergence of probability measures for Banach spaces:

Definition 2.6.2. Let $(\mu, \mu_n : n \geq 1)$ be a family of probability measures over $(B, \mathcal{B}_B)$, where B is a separable Banach space. We say that (μ_n) converges *weakly* to μ, written $\mu_n \xrightarrow{w} \mu$ if

$$\mu_n(A) \to \mu(A) \quad \forall A \in \mathcal{B}_B \text{ such that } \mu(\partial A) = 0,$$

where ∂A is the boundary of A. ※

We now define the following types of stochastic convergences for B-random variables:

Definition 2.6.3. Let $\{Y, Y_n : n \geq 1\}$ be a family of B-r.v. defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The (classical) stochastic convergences in B are defined by

1. (Y_n) converges to Y *almost surely*, written $Y_n \xrightarrow{a.s.} Y$, if

$$\mathbb{P}\left\{\omega \in \Omega : \|Y_n(\omega) - Y(\omega)\| \xrightarrow{n\to\infty} 0\right\} = 1.$$

2. (Y_n) converges to Y *in probability*, written $Y_n \xrightarrow{p} Y$, if for all $\varepsilon > 0$,
$$\mathbb{P}\left(\|Y_n - Y\| > \varepsilon\right) \overset{n\to\infty}{\longrightarrow} 0.$$

3. (Y_n) converges to Y *in distribution*, written $Y_n \xrightarrow{d} Y$, if
$$\mathbb{P}_{Y_n} \xrightarrow{w} \mathbb{P}_Y,$$
where $\mathbb{P}_Y = \mathbb{P} \circ Y^{-1} : \mathcal{B}_B \to [0,1]$.

※

The following lemma will be necessary for proving a weak law of large numbers:

Lemma 2.6.4. *Let H be a separable Hilbert space, and $X, Y : \Omega \to H$ be independent random variables s.t. $\|X\|_{L^2(\mathbb{P})} < \infty$, $\|Y\|_{L^2(\mathbb{P})} < \infty$. Then*

(i) $\mathbb{E}\left[\langle X, Y\rangle\right] = \langle \mathbb{E}X, \mathbb{E}Y\rangle$,

Furthermore, if either $\mathbb{E}X = 0$ or $\mathbb{E}Y = 0$, then

(ii) $\|X + Y\|^2_{L^2(\mathbb{P})} = \|X\|^2_{L^2(\mathbb{P})} + \|Y\|^2_{L^2(\mathbb{P})}$.

Proof. With the notation $P_{X,Y}(A) = \mathbb{P}\left((X,Y) \in A\right)$ for a borel set $A \subset H \times H$, then

$$\begin{aligned}
\mathbb{E}\left[\langle X, Y\rangle\right] &= \iint_{H\times H} \langle x, y\rangle \mathrm{d}P_{X,Y}(x,y) \\
&= \int_H \left(\int_H \langle x, y\rangle \mathrm{d}P_X(x)\right) \mathrm{d}P_Y(y) \\
&= \int_H \mathbb{E}\left[\langle X, y\rangle\right] \mathrm{d}P_Y(y) \\
&= \int_H \langle \mathbb{E}X, y\rangle \mathrm{d}P_Y(y) \\
&= \mathbb{E}\left[\langle \mathbb{E}X, Y\rangle\right] = \langle \mathbb{E}X, \mathbb{E}Y\rangle,
\end{aligned}$$

where the fourth and ultimate equality are justified by (2.13). This proves the assertion (i). The proof of (ii) is then a direct consequence:

$$\begin{aligned}
\|X + Y\|^2_{L^2(\mathbb{P})} &= \mathbb{E}\left[\|X + Y\|^2\right] \\
&= \mathbb{E}\left[\|X\|^2 + \|Y\|^2 + \langle X, Y\rangle + \overline{\langle X, Y\rangle}\right] \\
&= \mathbb{E}\|X\|^2 + \mathbb{E}\|Y\|^2 + \mathbb{E}\left[\langle X, Y\rangle\right] + \overline{\mathbb{E}\left[\langle X, Y\rangle\right]} \\
&= \|X\|^2_{L^2(\mathbb{P})} + \|Y\|^2_{L^2(\mathbb{P})} + \langle \mathbb{E}X, \mathbb{E}Y\rangle + \overline{\langle \mathbb{E}X, \mathbb{E}Y\rangle} \\
&= \|X\|^2_{L^2(\mathbb{P})} + \|Y\|^2_{L^2(\mathbb{P})}.
\end{aligned}$$

□

Now let us give a Weak Law of Large Numbers:

Theorem 2.6.5 (Weak Law of Large Numbers)**.** *Let H be a separable Hilbert space, and $X : \Omega \to H$ be a random variable such that $\mathbb{E}\|X\|^2 < \infty$. If $X_1, \ldots, X_n$ are i.i.d. copies of X, then*

$$\overline{X} \xrightarrow{p} m,$$

where $m = \mathbb{E}X$ and $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Proof. It is enough to show that $\|\overline{X} - m\|_{L^2(\mathbb{P})} \to 0$, because $L^2(\mathbb{P})$-convergence implies convergence in probability. First, notice that the random variables $X_i - m$ are independent, and that

$$\|X_i - m\|^2_{L^2(\mathbb{P})} = \|X - m\|^2_{L^2(\mathbb{P})}.$$

Let us calculate the $L^2(\mathbb{P})$-norm of $\overline{X} - m$:

$$\begin{aligned}
\left\|\overline{X} - m\right\|^2_{L^2(\mathbb{P})} &= \left\|\frac{1}{n}\sum_{i=1}^{n}(X_i - m)\right\|^2_{L^2(\mathbb{P})} \\
&= \frac{1}{n^2}\left\|\sum_{i=1}^{n}(X_i - m)\right\|^2_{L^2(\mathbb{P})} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\|X_i - m\|^2_{L^2(\mathbb{P})} \\
&= \frac{1}{n}\|X - m\|^2_{L^2(\mathbb{P})},
\end{aligned}$$

where the penultimate equality comes from Lemma 2.6.4. We know that $\|X - m\|^2_{L^2(\mathbb{P})} = K < \infty$, and thus

$$\left\|\overline{X} - m\right\|^2_{L^2(\mathbb{P})} \longrightarrow 0,$$

which completes the proof. □

As with the finite-dimensional case, the law of large numbers actually holds without the assumption of finite second moment, and it can be shown that the convergence with probability 1. This is called the *Strong Law of Large Numbers*:

Theorem 2.6.6 (Strong Law of Large Numbers)**.** *Let (X_i) be an i.i.d. sequence of B-r.v. such that $\mathbb{E}\|X_1\| < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}X_1,$$

where S_n is defined in (2.15)*.*

Proof. See Bosq (2000, p. 47) □

The Central Limit Theorem does not hold in general for Banach spaces when only assuming that $\mathbb{E}\left[\|X\|^2\right] < \infty$, and requires more assumptions. However, the CLT always holds in separable Hilbert spaces:

Theorem 2.6.7. *Let (X_i) be an i.i.d. sequence of H-random variables, where H is a separable Hilbert space. Suppose $\mathbb{E}\|X_1\|^2 < \infty$, let $m = \mathbb{E}X_1$ and $C = C_{X_1}$, the covariance operator of X_1. Then*

$$n^{-1/2}\sum_{i=1}^{n}(X_i - m) \xrightarrow{d} N,$$

where $N \sim \mathcal{N}(0, C)$, that is, N is a Gaussian H-r.v. with mean zero and covariance operator C.

Proof. See Bosq (2000, p. 51) □

As an example, let us investigate some asymptotic properties of the sample covariance operator of H-valued i.i.d. random variables:

Example 2.6.8. Let $H = L^2([0,1])$ and for $T \in \mathcal{L}(H)$, define $\|T\|_{\mathcal{H}}$ to be the Hilbert-Schmidt norm of T. Let us define

$$\mathcal{H} = \left\{T : H \to H \text{ linear } : \|T\|_{\mathcal{H}} < \infty\right\},$$

the space of Hilbert-Schmidt operators on H. The space $\mathcal{H}$ is itself a separable Hilbert space, with scalar product

$$\langle A, B\rangle_{\mathcal{H}} = \sum_{i\in\mathbb{N}}\langle Ae_i, Be_i\rangle, \quad A, B \in \mathcal{H},$$

where (e_i) is an orthonormal basis of H – see Proposition 1.5.10. Now suppose that X is an H-valued random variable, such that $\mathbb{E}\|X\|_H^4 < \infty$, where $\|\cdot\|_H$ is the norm on H, and suppose that X has mean zero (for simplicity) and covariance operator $\Gamma = \mathbb{E}\left[X \otimes X\right] \in \mathcal{H}$, where $\mathbb{E}\left[X \otimes X\right](f) = \mathbb{E}\left[\langle X, f\rangle X\right]$ for $f \in H$.

If $X_1, \ldots, X_n$ are i.i.d. copies of X, we define the empirical covariance operator of X by

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes X_i$$

which is a random operator on H, or an $\mathcal{H}$-random variable, defined by

$$\Gamma_n f = \frac{1}{n} \sum_{i=1}^{n} \langle X_i, f \rangle X_i, \quad f \in H.$$

Let us write $Y_i = X_i \otimes X_i$ for $i = 1, \ldots, n$. The Y_i's are i.i.d. $\mathcal{H}$-valued random variables, and if (e_j) is a orthonormal basis of H, then

$$\begin{aligned} \|Y_1\|_{\mathcal{H}}^2 = \sum_{j=1}^{\infty} \|Y_1 e_j\|^2 &= \sum_{j=1}^{\infty} \|(X_1 \otimes X_1) e_j\|^2 \\ &= \sum_{j=1}^{\infty} \langle X_1, e_j \rangle^2 \|X_1\|^2 \\ &= \|X_1\|^2 \sum_{j=1}^{\infty} \langle X_1, e_j \rangle^2 \\ &= \|X_1\|^4, \end{aligned}$$

where the last equality comes from Parseval's formula (1.4). Thus $\mathbb{E}\|Y_1\|_{\mathcal{H}}^2 = \mathbb{E}\|X_1\|^4 < \infty$. Also, $\mathbb{E}Y_1 = \mathbb{E}[X_1 \otimes X_1] = \mathbb{E}[X \otimes X] = \Gamma$, because X_1 has the same distribution as X. Let us denote by $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ the covariance operator of Y_1, which is nuclear by Proposition 2.5.9, and thus a Hilbert-Schmidt operator on the space $\mathcal{H}$. Applying the central limit Theorem yields

$$\sqrt{n}(\Gamma_n - \Gamma) \xrightarrow{d} N,$$

where $N \sim \mathcal{N}(0, \mathcal{C})$.

Let us compute the $L^2(\mathbb{P})$ Hilbert-Schmidt norm of $\Gamma_n - \Gamma$ defined by

$$\|A\|_{L^2(\mathbb{P})} = \sqrt{\mathbb{E}\|A\|_{\mathcal{H}}^2}, \quad \text{for an } \mathcal{H}\text{-random variable } A.$$

Using the notation $Y_{(i)} = X_{(i)} \otimes X_{(i)}$, an examination of the proof of Theo-

rem 2.6.5 gives us

$$\begin{aligned}\|\Gamma_n - \Gamma\|^2_{L^2(\mathbb{P})} &= \left\| \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}Y_i) \right\|^2_{L^2(\mathbb{P})} \\ &= \frac{1}{n}\|Y - \mathbb{E}Y\|^2_{L^2(\mathbb{P})} \\ &= \frac{1}{n}\mathbb{E}\left[\|Y - \mathbb{E}Y\|^2_{\mathcal{H}}\right] \\ &= \frac{1}{n}\mathbb{E}\left[\mathrm{trace}\left((Y - \mathbb{E}Y)\bigotimes(Y - \mathbb{E}Y)\right)\right] \\ &= \frac{1}{n}\mathrm{trace}\left(\mathbb{E}\left[(Y - \mathbb{E}Y)\bigotimes(Y - \mathbb{E}Y)\right]\right),\end{aligned}$$

where for $A, B \in \mathcal{H}$, we define $A \bigotimes B : \mathcal{H} \to \mathcal{H}$ by

$$\left(A \bigotimes B\right) C = \langle A, C\rangle_{\mathcal{H}} B, \quad C \in \mathcal{H}.$$

Using Proposition 2.5.9, we have

$$\begin{aligned}\|\Gamma_n - \Gamma\|^2_{L^2(\mathbb{P})} &= \frac{1}{n}\mathbb{E}\left[\|Y - \mathbb{E}Y\|^2_{\mathcal{H}}\right] \\ &\leq \frac{1}{n}\mathbb{E}\|Y\|^2_{\mathcal{H}} \\ &= \frac{1}{n}\mathbb{E}\|X\|^4 < \infty.\end{aligned}$$

From this, we get the bound

$$\|\Gamma_n - \Gamma\|_{L^2(\mathbb{P})} \leq \frac{1}{\sqrt{n}}\sqrt{\mathbb{E}\|X\|^4},$$

thus the convergence of the empirical covariance operator to the true covariance operator is of order $n^{-1/2}$ with respect to the $L^2(\mathbb{P})$-norm. ◀

2.7 Some Remarks on the Definitions of Gaussian Processes

The careful reader might have seen that the definition of Gaussian processes is not immediately obviously compatible with the definition of Hilbert-space valued Gaussian random variables, which we now recall:

Definition 2.7.1. An H-valued random variable, or *H-random variable* is a measurable function $X : \Omega \to H$, where $(\Omega, \mathcal{A})$ is a measurable space and H is a separable Hilbert space with its Borel σ-algebra.

An H-value random variable X is called *Gaussian* if, for all $h \in H$, the random variable $\langle X, h\rangle$ is a real-valued Gaussian random variable. ※

Now notice that the space $L^2([0,1])$ is an Hilbert space. So an $L^2([0,1])$-valued Gaussian random variable would be defined as:

Definition 2.7.2 (Cramér-Wold definition)**.** An $L^2([0,1])$-valued Gaussian random variable would be a random variable

$$X : \Omega \to H \tag{2.17}$$

such that for all $f \in L^2([0,1])$, the random variable $\Omega \to \mathbb{R}$ defined by

$$\omega \mapsto \langle X(\omega), f \rangle = \int_0^1 X_t(\omega) f(t) \mathrm{d}t \tag{2.18}$$

is Gaussian. ※

Now the natural question that arises is how whether this latter definition is compatible with our previous definition of Gaussian process (see definition 2.4.1). A subtlety is that the definition 2.7.2 would not give you a properly defined process. Indeed, fix $\omega \in \Omega$. Then $X(\omega) \in L^2([0,1])$, hence it is an *equivalence class* of function, and therefore the expression $X_t(\omega)$ has no meaning. But suppose we do not care this, and for each $\omega \in \Omega$, we pick a function in the equivalence class of $X(\omega)$ (the purists will notice that the axiom of choice needs to be used). Now we properly ask the question:

Question: Under which conditions do (2.19) and (2.20) coincide, where

$$\sum_{i=1}^{n} a_i X_{t_i} \text{ is Gaussian } \forall t_i \in [0,1], \forall a_i \in \mathbb{R}, \forall n \geq 1. \tag{2.19}$$

$$\int_0^1 X_t f(t) \mathrm{d}t \text{ is Gaussian } \forall f \in L^2([0,1]). \tag{2.20}$$

We first need a definition:

Definition 2.7.3. A Gaussian process $(X_t, t \in [0,1])$ is called *continuous in the mean* if

$$\lim_{h \to 0} \|X_{t+h} - X_t\|_{L^2(\mathbb{P})} = 0, \quad \forall t \in [0,1]$$

where $\|X_s\|_{L^2(\mathbb{P})} = \sqrt{\mathbb{E}\left[|X_s|^2\right]}$. ※

Notice that continuity in the mean is just the same thing as saying that the function $t \to \Xi$ is continuous with respect to the semi-norm $\|\cdot\|_{L^2(\mathbb{P})}$ on $\Xi = \{\xi : \Omega \to \mathbb{R} \,|\, \xi \text{ is measurable}\}$. Interestingly, continuity in the mean is equivalent to the continuity of the covariance function:

Proposition 2.7.4. *A Gaussian process is continuous in the mean if, and only if its covariance function is continuous.*

Proof. See (Grenander 1981, p. 38). □

We just need a technical result, which states that a random process which is continuous in the mean can be well approximated by "step functions":

Lemma 2.7.5. *Let $T = [0,1]$ and $(\Xi, \|\cdot\|)$ be a normed space. If $X : T \to \Xi$ is continuous with respect to $\|\cdot\|$, then for all $p > 0, \varepsilon > 0$, there exists $0 = t_0 < t_1 < \cdots < t_n = 1$ such that the random variable*

$$Y(t) = \sum_{i=1}^{n} \mathbf{1}_{[t_{i-1},t_i)}(t) X_{t_i - 1}$$

satisties the inequality

$$\int_T \|X_t - Y(t)\|^p \mathrm{d}t < \varepsilon,$$

where $\mathbf{1}_A$ is the indicator function of the set A.

Proof. Because the function $X : T \to \Xi$ is continuous with respect to $\|\cdot\|$ on the compact T, it is uniformly continuous on T. Hence $\exists \delta > 0$ such that $\|X_s - X_t\|^p < \varepsilon$ whenever $|s - t| < \delta, s, t \in T$. Define $t_j = j\delta$ for $j = 0, \ldots, n-1$, with $n - 1 = \lfloor 1/\delta \rfloor$, and let $t_n = 1$. Then $t_j - t_{j-1} \leq \delta$ for all $j = 1, \ldots, n$. Thus letting $Y_i = X_{t_{i-1}}$ and $Y(t) = \sum_{i=1}^{n} \mathbf{1}_{[t_{i-1},t_i)}(t) Y_i$, we have

$$\int_T \|X(t) - Y(t)\|^p = \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \|X(t) - Y_i\|^p \mathrm{d}t \leq \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \varepsilon \mathrm{d}t = \varepsilon,$$

and the proof is complete. □

Just before giving our next Theorem, let introduce some notation: given a stochastic process $(X_t, t \in [0,1])$, we define $L^2(X)$ as the completion of

$$\left\{ \sum_{i=1}^{n} a_i X_{t_i} : t_i \in [0,1], a_i \in \mathbb{R}, n \in \mathbb{N}. \right\}$$

under the norm $\|\cdot\|_{L^2(\mathbb{P})}$ defined in definition 2.7.3. Now a fact we are going to use is that if X is a Gaussian process (with respect to (2.19)), then any $v \in L^2(X)$ is a Gaussian random variable, possibly degenerate. Now we can give a result:

Theorem 2.7.6. *Let X be a process with values in $\mathcal{L}^2([0,1])$ which is continuous in the mean. Then the conditions*

$$\sum_{i=1}^{n} a_i X_{t_i} \text{ is Gaussian}; \quad \forall t_i \in [0,1], \forall a_i \in \mathbb{R}, \forall n \geq 1. \tag{2.21}$$

and

$$\int_0^1 X_t f(t)\mathrm{d}t \text{ is Gaussian}, \quad \forall f \in L^2([0,1]). \tag{2.22}$$

are equivalent.

Proof. First suppose that (2.21) holds. The idea is to show that for any $f \in L^2([0,1])$, $\langle X, f\rangle = \int_0^1 X_t f(t)\mathrm{d}t$ is in $L^2(X)$. We will show that for any $\varepsilon > 0$, there is a "step function" $Y(t) = \sum_{i=1}^n \mathbf{1}_{I_i}(t) X_{t_{i-1}}$ with

$$\|\langle X, f\rangle - \langle Y, f\rangle\|_{L^2(\mathbb{P})} < \varepsilon,$$

and this will complete the proof, for

$$\langle Y, f\rangle = \sum_{i=1}^{n} \left(\int_{I_i} f(t)\mathrm{d}t\right) X_{t_{i-1}} \in L^2(X).$$

The I_i are subintervals of $[0,1]$ of the form $[a,b)$.

Let us begin: take $f \in L^2([0,1])$, and assume without loss of generality that $\|f\| > 0$. Here we will use the notation $\|f\|^2 = \int_0^1 f^2(t)\mathrm{d}t$ for $f \in L^2([0,1])$ and $\|v\|_{L^2(\mathbb{P})} = \sqrt{\mathbb{E}[|v|^2]}$ for v a real-valued random variable. Fix $\varepsilon > 0$ and take $t_0 < t_1 < \cdots < t_n$ and Y as given by lemma 2.7 such that

$$\int_0^1 \|X_t - Y(t)\|^2_{L^2(\mathbb{P})}\mathrm{d}t < \frac{\varepsilon^2}{\|f\|^2}.$$

Then

$$\begin{aligned}
\|\langle X, f\rangle - \langle Y, f\rangle\|^2_{L^2(\mathbb{P})} &= \mathbb{E}\left[\left(\left(\langle X - Y, f\rangle\right)^2\right)\right] \\
&\leq \mathbb{E}\left[\|X - Y\|^2 \|f\|^2\right] \\
&= \|f\|^2 \mathbb{E}\left[\int_0^1 (X_t - Y(t))^2 \mathrm{d}t\right] \\
&= \|f\|^2 \int_0^1 \mathbb{E}\left[X_t - Y(t)\right]^2 \mathrm{d}t \\
&= \|f\|^2 \int_0^1 \|X_t - Y(t)\|^2_{L^2(\mathbb{P})}\mathrm{d}t \\
&\leq \|f\|^2 \frac{\varepsilon^2}{\|f\|^2} = \varepsilon^2,
\end{aligned}$$

where the first inequality comes from the Cauchy-Schwarz inequality, and the inversion of integrals is legitimate because the summand in positive.

Conversely, suppose that (2.22) holds. We are going to use the K.L. expansion (which holds for any process with continuous covariance function) to write

$$X_t = \sum_{n=1}^{\infty} \xi_n \varphi_n(t),$$

where the ξ_n's are uncorrolated random variables, and the convergence is uniform in $L^2(\Omega)$, meaning that

$$\sup_{t\in[0,1]} \|f_N(t) - X_t\|_{L^2(\mathbb{P})} \overset{N\to\infty}{\longrightarrow} 0, \tag{2.23}$$

where we used the notation $f_N(t) = \sum_{n=1}^{N} \xi_n \varphi_n(t)$. Now take $k \in \mathbb{N}$, $t_1, \ldots, t_k \in [0,1]$ and let $B = \sum_{j=1}^{m} \alpha_j X_{t_j}$ and $A_N = \sum_{j=1}^{m} \alpha_j f_N(t_j)$, where the α_j's are real numbers, and notice that

$$\sum_{j=1}^{k} \alpha_j f_N(t_j) = \sum_{n=1}^{N} \left(\sum_{j=1}^{k} \alpha_j \varphi_n(t_j) \right) \xi_n$$

is a Gaussian random variable. Also notice that (2.23) implies that for all N large enough,

$$\left\| \sum_{j=1}^{N} \alpha_j f_N(t_j) \right\|_{L^2(\mathbb{P})} \leq \sum_{j=1}^{N} |\alpha_j| \, \|f_N(t_j)\|_{L^2(\mathbb{P})} < \infty.$$

Hence the limit of A_N under $\|\cdot\|_{L^2(\mathbb{P})}$ exists and is a Gaussian random variable. Finally, we show that this limit is B, and the proof will be complete. Let $\varepsilon > 0$ and take N such that $\sup_t \|f_N(t) - X_t\|_{L^2(\mathbb{P})} < \frac{\varepsilon}{\sum_{j=1}^{m} |\alpha_j|}$, where the supremum is taken over $[0,1]$. Then by the triangle inequality

$$\left\| \sum_{j=1}^{m} \alpha_j f_N(t_j) - \sum_{j=1}^{m} \alpha_j X_{t_j} \right\|_{L^2(\mathbb{P})} \leq \sum_{j=1}^{m} |\alpha_j| \|f_N(t_j) - X_{t_j}\|_{L^2(\mathbb{P})} < \varepsilon,$$

and the proof is complete. □

Conclusions

In this section, we have presented fundamentals of probability in abstract spaces. We have first seen the product measure spaces, and their analogies

with the product topological spaces. This was a necessary step to talk of random processes, and to understand how the finite dimensional distributions define the distribution of a random process, which is in some sense an infinite dimensional random vector. We then reviewed some basic properties of Gaussian random vectors, and introduced Gaussian processes. Remarkably, we saw that the property of orthogonal decomposition of Gaussian random vectors extends naturally to their infinite dimensional analogue, provided some regularity conditions on the process are met – namely *continuity in the mean* of the process. This was given by the Karhunen-Loève expansion, whose use will be central in Section 5.

We then focused on some more abstract concepts, presented some notions of probabilities on Banach spaces, and reviewed some of the principal stochastic convergence Theorems in Banach spaces. We saw that the Strong Law of Large Numbers always holds in Banach Spaces, but that the Central Limit Theorem needs – in general – stronger assumptions than only $\mathbb{E}\|X\|^2 < \infty$. However, this assumption suffices when we restrict ourselves to Hilbert space valued random variables.

Eventually, we concluded with some remarks concerning the definition of Gaussian random processes. The motivation of this part was that Gaussian processes were introduced usually by stating they are random process such that their finite dimensional distributions are Gaussian. However, we can also think of a Gaussian process as a random variable with values in a Hilbert space, say $L^2([0,1])$. Then we could also argue that the use of the Cramér-Wold definition is legitimate (see Definitions 2.5.12 and 2.7.2). We thus attempted to understand the difference between these two definitions. It turned out that under the assumption of *continuity in the mean* of the random process – which is an important assumption of the Karhunen-Loève expansion – the two definitions coincide.

3

Basic Aspects of Functional Data Analysis

Now that we have reviewed some Hilbert space theory, and presented some topics of probability in abstract spaces, let us give a review of basic aspects of functional data analysis, based on material in Ramsay & Silverman (2005). This section is central to this report, because it introduces the techniques that allow to transform raw discrete data into functional datum.

We are first going to introduce the main ideas for turning discrete data into a functional datum: by using basis functions. We will present the least squares representations, and briefly talk about some other representation techniques. Then we will introduce the notion of roughness penalization, which is of particular importance due to the infinite dimensional nature of the data we are dealing with. A short presentation of registration of functional data will follow, before giving an introduction to functional principal component analysis (PCA) and regularized PCA. Eventually, we will present the functional linear model, which is the infinite dimensional analogue of a regression model.

Functional data?

The notion of *functional data* is, by itself, a sort of oxymoron. Indeed, any kind of data that can be gathered today is discrete. Suppose we are measuring the height of a adolescent, from his 13^{th} birthday, until his 20^{th}. Then even if we measure his height every minute, the data collected would still be discrete: we would have $(t_0, y_0), \dots (t_n, y_n)$, where t_i denotes the time at which the height was measured, and y_i is the height at time t_i.

On the other hand, "functional" means that the object considered is a *function* of some *continuous* parameter. For instance, for the height of an adolescent, we would consider $x(t)$ to be his height at time t, varying continuously from the 13^{th} to the 20^{th} birthday. Thus when speaking of functional data, we assume that the data collected has an underlying structure, which is a function with some regularity (we will be more precise soon). We assume

the existence of a function, say $x(t)$, and all we can see are *snapshots* of this underlying function, plus some error. For our adolescent height example, this would translate as

$$y_i = x(t_i) + \varepsilon_i, \quad i = 0, \dots n$$

where ε_i is the measurement error, and $x(t)$ is the underlying height function.

When is data functional?

A necessary condition for discrete data to be considered functional is that the underlying function x should not to be too erratic; otherwise trying to infer the function x would be an effort in vain. Thus we assume that x is continuous, and that it possesses one or more derivatives $Dx, D^2x, \dots, D^mx$. Here D denotes the differential operator: $(Dx)(t) = \frac{\mathrm{d}}{\mathrm{d}t}x(t)$.

In many situations, we will want to infer the underlying function x, as well as some of its derivatives – in such cases, the data is called functional, for we need to assume the existence of x before thinking of its derivatives. The mere consideration of data as multivariate wouldn't allow at all for such inferences. For instance, we would like to estimate the acceleration D^2x of the height of our adolescent, and see when it attains its peak.

One might suggest that if the data are sampled very finely (such as one measurement a minute for the height), we could just use the forward difference $\Delta_j y = (y_{j+1} - y_j)/(t_{j+1} - t_j)$ to estimate the $Dx(t_j)$, and reiterate this to estimate further derivatives $D^mx(t_j)$. However, this method doesn't give good estimates of the derivatives D^mx because the influence of the errors ε is amplified enormously! This phenomenon will be explained in more detail in Section 3.2.

3.1 From discrete to smooth functional data, without penalization

Now we are going to see how to transform the raw discrete data into smooth functions. The basic assumption is that x lies in a *finite dimensional subspace* V of a space of functions, say $\mathcal{L}^2([0,1])$ for the sake of formalism. So to infer x, we choose a basis $(\phi_k)_k$ of V and write $x(t) = \sum_{k=1}^{K} c_k\phi_k(t)$, where $K = \dim V$ and $c_k \in \mathbb{R}$. Beware of not thinking that this implies that our problem reduces to multivariate data analysis: the subspace V is unknown! Hence almost all the complexity resides in the choice of V, or similarly in the choice of the basis functions (ϕ_k). This is why we will often be talking of *representations of* x.

The number K of basis functions is also unknown ; ideally, we will want to have a small number K of basis functions, for this will:

1. Allow better interpretation of the coefficients,
2. Necessitate less computations, and
3. Allow to have better tests and confidence intervals.

This basis function approach is advantageous, for it is computationally adapted, and allows us to use the already existing matrix algebra language for solving our problems.

The Choice of Basis Functions

First, we need to choose what kind of basis functions to use for representing x. This choice depends on the nature of the data being analyzed: for instance, if the data is *periodic* in nature – such as monthly temperatures – then a truncated Fourier basis

$$1, \sin \omega t, \cos \omega t, \ldots, sinN\omega t, \cos N\omega t$$

would be appropriate. However, if no periodicity is assumed, then a B-spline basis is often better. But first, let us explain briefly what a spline is.

Spline functions

Definition 3.1.1. Let $t_0 < t_1 < \cdots < t_n$ and $y_0, y_1, \ldots, y_n \in \mathbb{R}$. Then a *spline of order* m passing through (t_j, y_j) is a function $f : [t_0, t_n] \to \mathbb{R}$ such that

1. $f(t_j) = y_j, \quad \forall j,$
2. $f_{|[t_j, t_{j+1}]}$ is a polynomial of *degree* $m-1$, for all j,
3. $f \in C^{m-2}([t_0, t_n])$. ※

Thus, a spline is a piecewise polynomial function, with smooth junctions. That is, for each t_j, the derivatives up to order $m-2$ on the left and on the right are equal. For instance, an order 4 (cubic) spline f would be piecewise a cubic polynomial, with f, f' and f'' continuous.

The *degrees of freedom* of a spline are equal to the order of the polynomials, plus the number of interior points on the interval. Using the notation of definition 3.1.1, we have

$$df(\text{spline}) = m + n - 1. \tag{3.1}$$

The distinct elements t_j are called *breakpoints*. We can generalize the notion of spline to a sequence $\tau_0, \ldots, \tau_N$, where not all τ_j's are distinct. These numbers τ_j are called *knots*. This enables to get splines which don't have the same level of regularity at each point. However, we will refer the interested reader to De Boor (2001) for more details.

For a fixed sequence (t_j) of breakpoints, the set of all splines of order less or equal to m is a real vector space, hence the notion of a basis for splines, called *B-spline basis*, makes sense. As you might have guessed, the dimension of this vector space is equal to the degrees of freedom given in (3.1). An interesting property of the B-spline basis is that a basis function has support in m or less subsequent intervals, which is important for the computation's efficiency.

Notice that for the Fourier basis, there is a canonical way of going from a length K basis to a length $K+1$ basis, and that is by adding a higher frequency function to the basis. Hence for the Fourier basis, augmenting the basis yields necessarily a better (finer) representation of x. However, this need not be true for the B-spline basis. If we take a new basis of length $K+1$ by cutting our interval $[a, b]$ into intervals of same lengths, the representation of x may be worse than with a shorter basis. This is because the $(K+1)$-spaces of splines doesn't contain – in general – the space of K-splines. However, if we just add one breakpoint to a sequence of breakpoints, the generated $(K+1)$-space of splines contains the previous one. Long story made short: **one must be careful when working with splines!**

Representations Through Least Squares

Given a basis (e.g. Fourier or B-spline basis), we would like to find which linear combination of its elements represents the data in the best way. Formally, our goal is to find the best representation of x with a given basis $\phi_1, \ldots, \phi_K$, in the sense that our representation $\hat{x}$ minimizes $\|y - \hat{x}\|^2$, for some measure of fit $\|\cdot\|$.

We assume the collected data $y_1, \ldots, y_n$ follow the model

$$y_j = x(t_j) + \varepsilon_j, \quad j = 1, \ldots, n$$

and we use the vectorial notation

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t) = \mathbf{c}^\mathsf{T} \boldsymbol{\phi}(t),$$

where $\mathbf{c}$ is the vector of coefficients c_k, and $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_K(t))^\mathsf{T}$.

Ordinary Least Squares

One way of finding the best finite dimensional representation of x is to choose $\mathbf{c}$ to minimize the sum of squared errors at each t_j, which we will call `SMSSE` to follow the notation of Ramsay & Silverman (2005). Thus we want to minimize

$$\texttt{SMSSE}(\mathbf{y}|\mathbf{c}) = \sum_j \left(y_j - \sum_k c_k \phi_k(t_j) \right)^2 = (\mathbf{y} - \mathbf{\Phi c})^\mathsf{T} (\mathbf{y} - \mathbf{\Phi c}), \tag{3.2}$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \phi_1(t_1) & \phi_2(t_1) & \cdots & \phi_K(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(t_n) & \phi_2(t_n) & \cdots & \phi_K(t_N) \end{pmatrix}.$$

Differentiation with respect to $\mathbf{c}$ yields that $\texttt{SMSSE}(\mathbf{y}|\mathbf{c})$ is minimized for

$$\hat{\mathbf{c}} = (\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi})^{-1}\mathbf{\Phi}^\mathsf{T}\mathbf{y},$$

provided the matrix $\mathbf{\Phi}$ is of full rank. Thus the vector of fitted values is

$$\hat{\mathbf{y}} = \mathbf{\Phi}(\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi})^{-1}\mathbf{\Phi}^\mathsf{T}\mathbf{y}.$$

Notice that the matrix $\mathbf{\Phi}(\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi})^{-1}\mathbf{\Phi}^\mathsf{T}$ defines an orthogonal projection $\mathbb{R}^n \to \mathbb{R}^n$; thus the least squares representation is the orthogonal projection of the data on the space spanned by the chosen basis.

Weighted Least Squares

A generalization of this method is useful when the errors ε_j are not assumed to be i.i.d, but to have a covariance structure given by the matrix $\mathbf{\Sigma}_\varepsilon$. Then it is better to use a *weighted* version of the `SMSSE` defined by

$$\texttt{SMSSE}(\mathbf{y}|\mathbf{c}) = (y - \mathbf{\Phi c})^\mathsf{T}\mathbf{W}(y - \mathbf{\Phi c}), \tag{3.3}$$

where $\mathbf{W} = \mathbf{\Sigma}_\varepsilon^{-1}$. Then we reach the minimum for

$$\hat{\mathbf{c}} = (\mathbf{\Phi}^\mathsf{T}\mathbf{W}\mathbf{\Phi})^{-1}\mathbf{\Phi}^\mathsf{T}\mathbf{W}\mathbf{y}.$$

However, when the complete matrix $\mathbf{\Sigma}_\varepsilon$ is not estimable, we compromise by assuming it is diagonal.

Kernel Smoothing

Notice that these methods of finding representations of x are linear in the data, that is, $\hat{x}(t) = \mathbf{S}\mathbf{y}$ for some matrix $\mathbf{S}$, where $\hat{x}(\mathbf{t}) = (\hat{x}(t_1), \ldots, \hat{x}(t_n))^\mathsf{T}$. This motivates the notion of *kernel smoothing*: suppose that the true value $x(t_j)$ is influenced more by values $x(t)$ with $|t-t_j|$ small than for $|t-t_j|$ large. Then we can use ideas of non-parametric density estimation to estimate x. Using a kernel function $\kappa(u)$, and defining $S_j(t)$ to be a function of the kernel which is decreasing in $|t-t_j|$, we have the representation

$$\hat{x}(t) = \sum_{j=1}^{n} S_j(t) y_j.$$

That is, the value $x(t)$ of the underlying x at time t is a linear combination (average) of the *local* observations.

Localized Basis Function Representation

Another technique for choosing the representation of x is the *Localized Basis Function Representation*, which consists in minimizing the sum of the errors *around* t, for a fixed t. That is, we wish to minimize

$$\texttt{SMSSE}_t(\mathbf{y}|\mathbf{c}) = \sum_j w_j(t) \left(y_j - \hat{x}(t_j)\right)^2, \quad \hat{x}(t_j) = \sum_k c_k \phi_k(t_j).$$

In matrix notation, this becomes

$$\texttt{SMSSE}_t(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})^\mathsf{T} \mathbf{W}(t) (\mathbf{y} - \mathbf{\Phi c}),$$

where $\mathbf{W}(t)$ is the diagonal matrix with elements $(w_1(t), \ldots, w_n(t))$.

Visualizing this is not straightforward. To get some intuition, imagine that $w_j(t)$ is a kernel function centered at t_j. Now fix $\mathbf{c}$ and pick $\bar{t}$, visualize the plot of the points (t_j, y_j), the kernels $w_j(t)$ as well as the representation $\mathbf{c}^\mathsf{T}\boldsymbol{\phi}(t)$. Now on top of $\bar{t}$, look at the value of each kernel $w_j(\bar{t})$: it tells you the fraction of error $y_j - \hat{x}(t_j)$ that you will take into account in the $\texttt{SMSSE}_t$.

3.2 From discrete to smooth functional data, with roughness penalization

Recall the Stein phenomenon (Stein 1956): for estimating the mean of $n \geq 3$ independent Gaussian random variables $X_1, \ldots, X_n$, the Maximum Likelihood Estimator (MLE) $\hat{\mu} = (X_1, \ldots, X_n)$ is inadmissible under a quadratic

loss function, and a shrunken version of $\hat{\mu}$ dominates the MLE. Even if the X_j's are not independent, the MLE is still dominated by a shrunken version of itself (Hoffmann 2000). Knowing that shrinkage can be seen as the solution to a penalized problem, we find here a heuristic for motivating the present section, in which we are going to find the best representation of the underlying function x, subject to some form of penalization.

Penalization is also crucial when we want to estimate the derivative Dx of the underlying function. The reason lies in the fact that the differential operator D is *unbounded* under the L^2-norm (see example 1.3.9). We could have a representation $\hat{x}$ which is very close to the true x, in the sense that $\|x - \hat{x}\| < \varepsilon$, but such that $\|Dx - D\hat{x}\|$ is huge! Another way of seeing this is in the frequency domain: suppose $\hat{x}(t) = x(t) + a \sin Nt$, with a small, but N very large. That is, x and $\hat{x}$ differ only by a high frequency: $\|x - \hat{x}\|$ is small, but $\|Dx - D\hat{x}\|$ is huge. Therefore we cannot estimate the derivatives of x by using the ordinary least squares estimation, and penalization is necessary.

One other motivation for penalization – which is closely connected with the Stein phenomenon – is the following: suppose we wish to find a representation $\hat{x}$ of x that minimizes the mean square error (MSE) at t, defined by

$$\texttt{MSE}(\hat{x}(t)) = \mathbb{E}\left[(\hat{x}(t) - x(t))^2\right] = \texttt{Bias}^2[\hat{x}(t)] + \texttt{Var}[\hat{x}(t)].$$

If a representation $\hat{x}$ overfits the observations y_j, meaning that $\texttt{Bias}^2[\hat{x}(t)] \simeq 0$, then the variance $\texttt{Var}[\hat{x}(t_j)]$ will be very large. Indeed, the representation $\hat{x}$ will be too "wiggly" and sensitive to a small change in the data. Thus to minimize the MSE, we have to allow for some bias in order to reduce the variance. This is the well-known bias/variance trade-off, and a way of reducing the variance of $\hat{x}$ is through penalization.

Penalized Sum of Square Errors for Estimating the Unknown Function

We would like penalize the representation $\hat{x}$ for being too rough. But what is roughness? Intuitively, the idea is to control how much the function "changes direction" and "how fast" it does it. The direction of a function x, at each point t, can be seen as the value of its derivative at this point: $Dx(t)$. So the variation of the direction would be given by the second derivative $D^2x(t)$, a large value indicating fast changes in direction, and a smaller value (close to 0) meaning the function is almost linear in a neighborhood, and thus that its direction doesn't change locally. So the roughness can be viewed as the cumulative variation of the first derivative of g, and thus, we need to penalize

for large values of

$$\texttt{PEN}_2(x) = \int \left[D^2 x(t)\right]^2 \mathrm{d}t. \tag{3.4}$$

Therefore, the penalized sum of square errors is

$$\texttt{PENSSE}_\lambda(\mathbf{y}|x) = [\mathbf{y} - x(\mathbf{t})]^\mathsf{T}\mathbf{W}[\mathbf{y} - x(\mathbf{t})] + \lambda \cdot \texttt{PEN}_2(x), \tag{3.5}$$

where $\mathbf{W}$ is a weight matrix. The parameter λ is a called a *smoothing parameter*, and controls the trade-off between exact interpolation and a totally smooth representation. If $\lambda = 0$, then (3.5) is just the same as the sum of squared errors (3.3). If we let $\lambda \to \infty$, then the minimizer $\hat{x}$ of (3.5) will be a line $\hat{x}(t) = at + b$, and we would be doing linear regression.

In order to write (3.5) in the matrix algebra language, notice that if we write $x(t) = \boldsymbol{\phi}(t)^\mathsf{T}\mathbf{c}$, then the linearity of the operator D^2, of the integral and the bilinearity of the matrix product yields

$$\texttt{PEN}_2(x) = \mathbf{c}^\mathsf{T}\mathbf{R}\mathbf{c}, \quad (\mathbf{R})_{ij} = \int D^2\phi_i(s) D^2\phi_j(s)\mathrm{d}s.$$

Thus

$$\texttt{PENSSE}_\lambda(\mathbf{y}|\mathbf{c}) = [\mathbf{y} - \boldsymbol{\Phi}\mathbf{c}]^\mathsf{T}\mathbf{W}[\mathbf{y} - \boldsymbol{\Phi}\mathbf{c}] + \lambda\mathbf{c}^\mathsf{T}\mathbf{R}\mathbf{c}, \tag{3.6}$$

which is minimized for

$$\hat{\mathbf{c}} = (\boldsymbol{\Phi}^\mathsf{T}\mathbf{W}\boldsymbol{\Phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\Phi}^\mathsf{T}\mathbf{W}\mathbf{y}. \tag{3.7}$$

We see that this formula resembles the estimator obtained using ridge regression.

A special case of (3.5) is when $\mathbf{W} = \mathbf{I}$, the identity matrix. Remarkably, the minimizer of this problem among all functions x with continuous second derivative is a cubic spline with knots at $t_1, \ldots, t_n$ (Hastie & Tibshirani 1990, p.27).

Penalization for Other Estimators

The idea in the previous section was to penalize the roughness of the representation $\hat{x}$, in the sense of $\texttt{PEN}_2(\hat{x}) = \int (D^2\hat{x}(t))^2 \mathrm{d}t$. But there is no reason why we would restrain ourselves to only this kind of penalization. If we wish to estimate the acceleration D^2x, it will be of no use to penalize $\texttt{PEN}_2(\hat{x})$; instead, it would be preferable to penalize $\texttt{PEN}_4(\hat{x}) = \int (D^4\hat{x}(t))^2 \mathrm{d}t$, that is, the roughness of $D^2\hat{x}$. More generally, we could replace $\texttt{PEN}_2(x)$ in (3.5) by

$\texttt{PEN}_m(x) = \int (D^m \hat{x}(t))^2 \mathrm{d}t$ for a certain m. Even more generally, we could wish to penalize the magnitude of some *functional* $L\hat{x}$ of the representation (L will often be a *differential* operator), and thus we would penalize

$$\int (L\hat{x}(t))^2 \mathrm{d}t = \|L\hat{x}\|^2,$$

where $\|\cdot\|$ denotes the L^2-norm. Notice that this setup generalizes the previous ones.

How to Choose the Smoothing Parameter?

A method for choosing λ is the so-called *cross-validation* method, or *CV* method. To really understand why the method is so usefull, notice that the ideal λ could be known only if we knew the real underlying function x, which we don't know of course. Now suppose we have a representation $\hat{x}$ of x, and we are given a new measurement (t_j, y_j). Then we can see how well our representation $\hat{x}$ would have predicted this new data point. Now recall that our representation depends actually on λ, hence we have a representation $\hat{x}_\lambda$ for each λ. We could thus choose the λ for which $\hat{x}_\lambda$ gives the best prediction *in expectation* of the new data point (t_j, y_j). In essence this is the idea behind cross-validation!

However, we do not often get new data points, so we will leave one observation (t_j, y_j) out of our data set, compute the smoothed representation, and then look at the prediction error for (t_j, y_j). In fact, we will, turn by turn, leave one observation out, and compute for each "incomplete" data sample the smoothed representation, and then the prediction error of the removed observation. If we sum all these prediction errors, and minimize them with respect to λ, we have our method for choosing the smoothing parameter. This will yield the representation that gives the best prediction (in expectation) of new data points.

In a more condensed and explicit form, we have to find λ that minimizes

$$\sum_{j=1}^{n} \left(y_j - \hat{x}_\lambda^{(j)}(t_j) \right)^2, \tag{3.8}$$

where $\hat{x}_\lambda^{(j)}$ is the representation obtained using (3.7) by leaving (t_j, y_j) out of our data set.

More general (and sophisticated) procedures exist to find the smoothing parameter – such as the generalized cross-validation method. We refer the reader to Ramsay & Silverman (2005, Section 5.4.3) for more details.

3.3 Registration of functional data

Remark 3.3.1. From now on, we will suppose that the data is given to us in functional form.

Until now, we have assumed that the functional datum $\hat{x}$ follows the model $\hat{x}(t) = x(t) + \varepsilon(t)$, where x is the true underlying function, and ε is the observation error. A problem that occurs in practice is that the observed *curves* $x_1(t), \ldots, x_n(t)$ may differ not only because of the error ε, but also in a transformation of the parameter t. For instance, if we observe that the height $h_j(t)$ of a few adolescents, the start of their growth peak will very certainly not be at the exact time, and thus we need to change the *tempo* of time in order to be able to retrieve jointly information from the h_j's. In a somewhat more formal setup, we assume that the data follows the model

$$\hat{x}(t) = x(h(t)) + \varepsilon(t), \tag{3.9}$$

where h is some random transformation of time, called the *time-warping function*. The problem of finding the *registered curve* $\hat{x}^*(t) = \hat{x}(h^{-1}(t))$ is called the *registration problem*. Of course, this problem – when considered in the most general setup – is very complicated. So let us start by a simpler model.

The shift registration model is the simplest model in the form (3.9) we could imagine: we suppose $h(t) = t - \delta$, where δ is some random variable. Thus this is the same as saying that

$$x_j(t) = x(t - \delta_j) + \varepsilon(t),$$

with $\delta_j \in \mathbb{R}$ unknown. We are interested in finding $x_j^*(t) = x_j(t+\delta_j)$, and this can be done by a least squares criterion similar to that presented in Ramsay & Silverman (2005, Section 7.2.1), which we reproduce here.

Define $\hat{\mu}(t) = \sum_j x_j(t)/n$, the sample mean curve. Now letting $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$, we define

$$\texttt{REGSSE}(\hat{\mu}, \boldsymbol{\delta}) = \sum_{j=1}^{n} \int_T [x_j^*(t) - \hat{\mu}(t)]^2 \mathrm{d}t, \tag{3.10}$$

where $x_j^*(t) = x_j(t + \delta_j)$ and T is the time interval over which the sample curves and the underlying are considered to be defined. As you certainly have guessed, we will find δ that minimizes (3.10). But this will minimize the difference with the sample mean curve $\hat{\mu}$, which itself was constructed using the shifted sample curves x_j. Hence we need to proceed iteratively:

1. Compute $\hat{\mu}$ from $x_1, \ldots, x_n$,
2. Find $\boldsymbol{\delta}$ that minimizes $\texttt{REGSSE}(\hat{\mu}, \boldsymbol{\delta})$,
3. Set $x_j(t) = x_j(t + \delta_j)$ for all j,
4. Go back to step 1.

Such a method is often called a *Procrustes method*, and in practice, the process usually converges within one or two iterations.

One step towards more general time-warping functions is to use distinguishing or specific features of the curves x_j in order to register them. For instance, in our imaginary adolescent height data curves, we can identify the first time-points t_j at which the height accelerations $D^2 h_j$ are zero, that is

$$t_j = \inf_{t \in T} \left\{ D^2 h_j(t) = 0 \right\},$$

and use these points t_j to register the height curves.

More generally, if m distinguishing landmarks appear in the curves x_j at the times $t_{j,k}, k = 1, \ldots, m$, and if $t_{0,k}$ are the corresponding times in the sample mean curve $\hat{\mu}$, we look for time-warping functions h_i such that

- $h_j(t_{0,k}) = t_{j,k},\ k = 1, \ldots, m$,
- $t \mapsto h_j(t)$ is a strictly increasing function.

The registered curves are thus $x_j^*(t) = x_j(h_j(t))$. Often, linear interpolation is used to define h_j in the intervals $[t_{0,j}, t_{0,j+1}]$.

To push even further into generality, we can use the natural assumption that h should be time increasing, and thus write

$$h(t) = C_0 + C_1 \int_a^t \exp W(s) \mathrm{d}s,$$

where the interval T is assumed to be of the form $[a, b]$. Another idea is to use some sort of principal component analysis technique – we refer the reader to (Ramsay & Silverman 2005, Sections 7.5, 7.6) for more details about this.

3.4 Principal Component Analysis

Principal component analysis (PCA) is a technique for finding the main features of some data set. We first introduce PCA for multivariate data, before presenting PCA for functional data.

PCA for Multivariate Data

The usual multivariate PCA technique is as follows: suppose you have n vector observations $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ which are centered, in the sense that $\sum_j \mathbf{x}_j = \mathbf{0}$, and you wish to find the principal direction in which they are oriented, meaning that we look for the direction $\boldsymbol{\xi}$ for which the sample variance of the random variables $\langle \boldsymbol{\xi}, \mathbf{x}_j \rangle$ is maximized. This can be formulated as the solution to the following maximization problem:

$$\boldsymbol{\xi}_1 = \operatorname*{argmax}_{\boldsymbol{\xi}} \left\{ \frac{1}{n} \sum_j \langle \boldsymbol{\xi}, \mathbf{x}_j \rangle^2 : \boldsymbol{\xi} \in \mathbb{S}^{p-1} \right\}, \tag{3.11}$$

where $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ are the scalar product, respectively the norm, in $\mathbb{R}^p$, and

$$\mathbb{S}^{p-1} = \{ \mathbf{x} \in \mathbb{R}^p : \|x\| = 1 \},$$

is the unit sphere in $\mathbb{R}^p$.

Now we can continue and find the next principal directions of the data, by solving iteratively

$$\boldsymbol{\xi}_k = \operatorname*{argmax}_{\boldsymbol{\xi}} \left\{ \frac{1}{n} \sum_j \langle \boldsymbol{\xi}, \mathbf{x}_j \rangle^2 : \boldsymbol{\xi} \in \mathbb{S}^{p-1}, \langle \boldsymbol{\xi}, \boldsymbol{\xi}_l \rangle = 0 \text{ for } l = 1, \ldots, k-1 \right\}, \tag{3.12}$$

for $k = 2, \ldots, p$. That is, we look for the principal directions in the space orthogonal to $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{k-1}$. The $\boldsymbol{\xi}_k$'s are called the *principal components* of the data.

PCA for Functional Data

A motivation for functional PCA is that it gives a the orthonormal basis that approximates, for a given truncation level K, the sample curves as closely as possible. We can easily generalize the idea of multivariate principal component analysis to curves (functional data) $x_1, \ldots, x_n$. Now our vectors are functions, and hence the first principal component will be a function $\xi_1(t)$ which solves the variational problem

$$\xi_1 = \operatorname*{argmax}_{\xi} \left\{ \frac{1}{n} \sum_j \langle \xi, x_j \rangle^2 : \|\xi\|^2 = 1 \right\}, \tag{3.13}$$

where $\langle \cdot, \cdot \rangle$ is the L^2 scalar product $\langle \xi_1, x_j \rangle = \int \xi_1(t) x_j(t) \mathrm{d}t$ and $\| \cdot \|$ is the norm it defines. Naturally, the other principal components are defined in a way similar to (3.12).

PCA : An Eigenanalysis Approach

Now assume that we observe n vector observations $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})$, $j = 1, \ldots, n$, that have mean zero, in the sense that $\sum_j x_{jl}/n = 0$ for all l, and let $\mathbf{X}$ be the $n \times p$ matrix with components $\mathbf{X}_{jl} = x_{jl}$. Then our first PCA problem can be rewritten as

$$\boldsymbol{\xi}_1 = \underset{\boldsymbol{\xi}}{\operatorname{argmax}} \left\{ n^{-1} \boldsymbol{\xi}^\mathsf{T} \mathbf{X}^\mathsf{T} \mathbf{X} \boldsymbol{\xi} : \boldsymbol{\xi} \in \mathbb{S}^{p-1} \right\},$$

or more simply, if one defines the sample covariance matrix $\mathbf{V} = n^{-1}\mathbf{X}^\mathsf{T}\mathbf{X}$, by

$$\boldsymbol{\xi}_1 = \underset{\boldsymbol{\xi}}{\operatorname{argmax}} \left\{ \boldsymbol{\xi}^\mathsf{T} \mathbf{V} \boldsymbol{\xi} : \boldsymbol{\xi} \in \mathbb{S}^{p-1} \right\}.$$

The solution to this problem is found by finding the eigenvector of the matrix $\mathbf{V}$ with largest eigenvalue, that is

$$\boldsymbol{\xi}_1 = \underset{\boldsymbol{\xi}}{\operatorname{argmax}} \left\{ \rho : \mathbf{V}\boldsymbol{\xi} = \rho \boldsymbol{\xi}, \boldsymbol{\xi} \in \mathbb{S}^{p-1} \right\}, \tag{3.14}$$

and similarly for the remaining components. Notice that we are assured that the right-hand-side set of (3.14) is non-empty because $\mathbf{V}$ is a symmetric matrix.

As you might have guessed, in analogy with the multivariate case, the functional PCA is just a truncation of an empirical Karhunen-Loève expansion (see Theorem 2.4.10). That is, define the empirical covariance function

$$v(s,t) = \frac{1}{n} \sum_{j=1}^{n} x_j(s) x_j(t),$$

and the covariance operator $V\xi(s) = \int v(s,t)\xi(t)\mathrm{d}t$, which is the integral operator with kernel function $v(s,t)$. Then the first principal component is

$$\xi_1 = \underset{\xi}{\operatorname{argmax}} \left\{ \rho : V\xi = \rho\xi \right\}. \tag{3.15}$$

In practice, determination of the functional principal components requires one to reduce the problem (3.15) to a matrix eigenanalysis problem. The two main techniques are

1. Discretization of the functions;

2. Basis expansion.

In the first method, the idea is to replace the function ξ by a vector $\boldsymbol{\xi}$ and the integral operator V (covariance operator) by a sort of Riemann sum, and thus reduce the problem to the form (3.14).

The second method involves writing the curves x_i as linear combinations of basis functions, and also reduces our problem to finding an eigenvector of a matrix. We refer the reader to Ramsay & Silverman (2005, Section 8.4) for more details.

3.5 Regularized Principal Component Analysis

As we have done with smoothing functional data, we might apply the idea of penalization to smooth the principal components, in order to avoid having too rough ξ_k's. A motivation for doing so is the following: suppose we believe that the true principal components of the underlying function x have some level of smoothness. Then we would like the estimated principal component to be also smooth, unlike with the standard PCA, which may yield very rough components, as it is only trying to maximize the variance.

We will thus penalize for large values of $\texttt{PEN}_2(\xi)$, similarly to the penalized least squares criterion. Our first principal component is thus the C^2 function ξ_1 with L^2-norm 1 that minimizes

$$\texttt{PCAPSV}(\xi) = \frac{\sum_j \langle \xi, x_j \rangle}{\|\xi\|^2 + \lambda \cdot \texttt{PEN}_2(\xi)}, \tag{3.16}$$

where λ is the smoothing parameter. For $\lambda = 0$, this reduces to the standard functional PCA. Now for the following principal components, we also maximize PCAPSV, but with a modified orthogonality condition: we require that

$$\int \xi_j(s)\xi_k(s)\mathrm{d}s + \int D^2\xi_j(s)D^2\xi_k(s)\mathrm{d}s = \langle \xi_j, \xi_k \rangle + \langle D^2\xi_j, D^2\xi_k \rangle = 0, \tag{3.17}$$

for $j = 1, \ldots, k-1$.

Remark 3.5.1. Notice that the left-hand side of (3.17) actually defines a new inner product on the space of C^2 functions, and that the ξ_j's are not too far from L^2-orthogonality, because

$$|\langle D^2\xi_j, D^2\xi_k \rangle| \leq \|D^2\xi_j\| \|D^2\xi_k\|,$$

by the Cauchy-Schwarz inequality. So, because $\|D^2\xi\|$ cannot be too large (depending on λ of course), the scalar product $|\langle \xi_j, \xi_k \rangle|$ is not too large.

As with the usual PCA, the parameter λ can be chosen by cross-validation, and the regularized PCA also reduces to an eigenanalysis problem.

Once registration, smoothing, and optimal finite-dimensional approximation are established, one turns to the issue of doing statistics on functional data. For instance, we can look at the infinite dimensional analogues of a regression model.

3.6 The Functional Linear Model

The Functional Linear Model (FLM) is just the generalization of a regression model to infinite dimensions. A very general model is

$$y = Lx + \varepsilon, \tag{3.18}$$

where $L : H_1 \to H_2$ is a linear operator between two Hilbert spaces H_1, H_2, and is called the *functional predictor* or *explanatory functional*. The vector $x \in H_1$ is called the *parameter*, $y \in H_2$ is the *response* and $\varepsilon \in H_2$ is the *error*. The model (3.18) is called a *functional* linear model because the parameter $x \in H_1$ is functional.

The model (3.18) is too general to be studied as is. We therefore look at special forms of (3.18), depending on H_2 and on the form of the covariate $L : H_1 \to H_2$. In the sequel, we shall assume that H_1 is a function space, such as a product of $\mathcal{L}^2([0,1])$ spaces for instance. The different cases we are going to consider are:

1. The response is functional and the covariate is a multivariate (a matrix),
2. The response is functional, and the covariate is a (multivariate) multiplication operator,
3. The response is scalar, and the covariate is an integral operator,
4. The response is functional, and the covariate is an integral operator.

Functional Responses, Multivariate Covariates

Let us consider the case where

$$H_2 = \left(\mathcal{L}^2([0,1])\right)^n = \mathcal{L}^2([0,1]) \times \mathcal{L}^2([0,1]) \times \cdots \times \mathcal{L}^2([0,1]),$$

and $H_1 = \left(\mathcal{L}^2([0,1])\right)^p$. The parameter and the response can be seen as vectors $\mathbf{y}, \mathbf{x}$, with coefficients in $\mathcal{L}^2([0,1])$, and if we fix $t \in [0,1]$, then

$\mathbf{y}(t) \in \mathbb{R}^n$ and $\mathbf{x}(t) \in \mathbb{R}^p$. Hence any $n \times p$ matrix $\mathbf{Z}$ defines a linear operator L between H_1 and H_2 by $(L\mathbf{x})(t) = \mathbf{Z} \cdot \mathbf{x}(\mathbf{t})$.

In such cases, the covariate L is said to be *multivariate*, because L is not a truly infinite dimensional operator, and the FLM reduces to

$$\mathbf{y}(t) = \mathbf{Z}\mathbf{x}(t) + \boldsymbol{\varepsilon}(t), \quad \forall t \in [0,1]. \tag{3.19}$$

We seek to minimize the fitting criterion

$$\texttt{LMSSE}(\mathbf{x}) = \int_0^1 [\mathbf{y}(t) - \mathbf{Z}\mathbf{x}(t)]^\mathsf{T}[\mathbf{y}(t) - \mathbf{Z}\mathbf{x}(t)]\mathrm{d}t. \tag{3.20}$$

If no particular assumption is made on the parameter function $\mathbf{x}(t)$, then $\texttt{LMSSE}(\mathbf{x})$ can be minimized by minimizing $\|\mathbf{y}(t) - \mathbf{Z}\mathbf{x}(t)\|^2$ for each t. In practice, we would choose a suitable grid of values of t and compute the pointwise minimizers $\hat{\mathbf{x}}(t)$ at each knot, and then interpolate between these values.

Penalizing the parameter

The idea of regularization can be also applied here to the parameter $\mathbf{x}$, in order to prevent us from incorporating high frequencies which would account for noise. In order to do so, we need some algebraic tools:

Definition 3.6.1 (The Kronecker product)**.** The *Kronecker product* of an $m \times n$ matrix $\mathbf{A}$ with a $p \times q$ matrix B a matrix $\mathbf{A} \otimes \mathbf{B}$ of order $mp \times nq$ consisting of submatrices $a_{kl}\mathbf{B}$, explicitly

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1_n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2_n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{m_n}\mathbf{B} \end{pmatrix}.$$

Let us also introduce the function $\mathrm{vec}\,(\cdot) : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{mn}$ which simply writes a matrix in a column-wise form, namely,

$$\mathrm{vec}\,(\mathbf{A}) = (a_{11}, a_{21}, \ldots, a_{m1}, a_{12}, a_{22}, \ldots, a_{m2}, \ldots, a_{m(n-1)}, a_{mn})^\mathsf{T}.$$

One use of the Kronecker product and the $\mathrm{vec}\,(\cdot)$ function is that it allows writing an equation of the form

$$\mathbf{AXB} = \mathbf{C},$$

which cannot be seen directly as a system of linear equations, in the simpler form

$$(\mathbf{B}^\mathsf{T} \otimes \mathbf{A})\mathrm{vec}\,(\mathbf{X}) = \mathrm{vec}\,(\mathbf{C}).$$

This then yields the solution $\mathrm{vec}\,(\mathbf{X}) = (\mathbf{B}^\mathsf{T} \otimes \mathbf{A})^{-1}\mathrm{vec}\,(\mathbf{C})$, provided the Kronecker product yields a regular matrix.

Notice also that the Kronecker product gives the matrix representation of the tensor product of linear transformations on finite dimensional spaces – See Halmos (1974*a*, §52) for instance. ※

Now we can give explicitly the formulas for finding a minimizer of $\mathtt{LMSSE}(\mathbf{x})$ with the basis approach: suppose without loss of generality the choice of the same basis of functions $\phi_1, \ldots, \phi_{K_y}$ for $y_1, y_2, \ldots, y_n$, the component functions of $\mathbf{y}$, and that we write $\mathbf{y}(t) = \mathbf{C}\boldsymbol{\phi}(t)$, where $\boldsymbol{\phi}$ is the column matrix $(\phi_1, \ldots, \phi_{K_y})^\mathsf{T}$ and $\mathbf{C}$ is the $n \times K_y$ coefficient matrix. Do the same for $\hat{\mathbf{x}}$, the penalized minimizer of $\mathtt{LMSSE}(\mathbf{x})$, by writing $\hat{\mathbf{x}}(t) = \mathbf{B}\boldsymbol{\theta}(t)$, where $\boldsymbol{\theta}$ is the column of basis functions for $\hat{\mathbf{x}}$.

We use a linear operator $\mathcal{R}$ for the roughness

$$\mathtt{PEN}_\mathcal{R}(\mathbf{x}) = \int [\mathcal{R}\mathbf{x}(t)]^\mathsf{T}[\mathcal{R}\mathbf{x}(t)]\mathrm{d}t = \int \|\mathcal{R}\mathbf{x}(t)\|^2\mathrm{d}t,$$

and define the matrices

$$\mathbf{J}_{\phi\phi} = \int \boldsymbol{\phi}\boldsymbol{\phi}^\mathsf{T},\ \mathbf{J}_{\theta\theta} = \int \boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T},\ \mathbf{J}_{\phi\theta} = \int \boldsymbol{\phi}\boldsymbol{\theta}^\mathsf{T}, \tag{3.21}$$

and

$$\mathbf{R} = \int [\mathcal{R}\boldsymbol{\theta}][\mathcal{R}\boldsymbol{\theta}]^\mathsf{T}, \tag{3.22}$$

where the (t) and $\mathrm{d}t$ have been dropped for clarity. The solution to the penalized least squares criterion

$$\mathtt{PENSSE}(\mathbf{y}|\mathbf{x}) = \int \|\mathbf{C}\boldsymbol{\phi} - \mathbf{Z}\mathbf{B}\boldsymbol{\theta}\|^2 + \lambda \int \|\mathcal{R}\mathbf{B}\boldsymbol{\theta}\|^2,$$

is

$$\mathrm{vec}\,(\mathbf{B}) = [(\mathbf{J}_{\theta\theta} + \lambda\mathbf{R}) \otimes (\mathbf{Z}^\mathsf{T}\mathbf{Z})]^{-1}\mathrm{vec}\left(\mathbf{Z}^\mathsf{T}\mathbf{C}\mathbf{J}_{\phi\theta}\right), \tag{3.23}$$

where $\|\cdot\|$ stands for the 2-norm on $\mathbb{R}^n$.

Functional Responses, Functional Covariates in the Concurrent Model

A more general case of the FLM is when we still have $H_2 = \left(\mathcal{L}^2\left([0,1]\right)\right)^n$ and $H_1 = \left(\mathcal{L}^2\left([0,1]\right)\right)^p$, but $L : H_1 \to H_2$ is a multiplication operator, in the sense that

$$(L\mathbf{x})(t) = \mathbf{Z}(t)\mathbf{x}(t),$$

where $\mathbf{Z}(t)$ is a $n \times p$ matrix for each t in the time interval. Hence the model we have is

$$\mathbf{y}(t) = \mathbf{Z}(t)\mathbf{x}(t) + \boldsymbol{\varepsilon}(t), \quad \forall t \in [0,1] \tag{3.24}$$

Such a model is called *concurrent*, because only the values of $\mathbf{x}$ at time t influence the value of $\mathbf{y}(t)$.

The penalized fitting criterion is very similar to the model with multivariate covariates, and an explicit equation can be written for the regularized $\hat{\mathbf{x}}$ when $\mathbf{y}$ and $\mathbf{x}$ are viewed in a basis expansion form. The level of complexity is a little higher than for the multivariate case, and some numerical integration is needed. We leave the interested reader to look at section 14.4 of Ramsay & Silverman (2005).

FLM with Scalar Responses

Another class of functional linear models is the one where $H_2 = \mathbb{R}^n$, $H_1 = (\mathcal{L}^2([0,1]))^p$, and the operator $L : H_1 \to H_2$ is defined by

$$L\mathbf{x} = \int \mathbf{Z}(t)\mathbf{x}(t)\mathrm{d}t,$$

with $\mathbf{Z}(t)$ being an $n \times p$ matrix for each t. In the case $n = 1$, the operator becomes

$$L\mathbf{x} = \int \left[\sum_{j=1}^{p} z_j(t)x_j(t)\right] \mathrm{d}t = \sum_{j=1}^{p} \int z_j(t)x_j(t)\mathrm{d}t.$$

Hence the model we have is

$$\mathbf{y} = \int \mathbf{Z}(t)\mathbf{x}(t)\mathrm{d}t + \boldsymbol{\varepsilon}. \tag{3.25}$$

This model is called the FLM with *scalar responses*.

Notice that in this model, it might be possible that the subspace of functions $\{\mathbf{x} : L\mathbf{x} = \mathbf{y}\}$ is infinite dimensional. For instance, this is the case when $n = 1, p = 1$: $y = \int z(t)x(t)\mathrm{d}t + \varepsilon$ and $\int z^2(t)\mathrm{d}t > 0$. Therefore, regularization of the parameter $\mathbf{x}$ is crucial here!

Regularization with a function basis

For the sake of simplicity, let us assume here that $p = 1$, and hence our model is

$$y_j = \int z_j(t)x(t)\mathrm{d}t, \quad j = 1, \ldots, n. \tag{3.26}$$

We wish to minimize the penalized residual sum of squares

$$\texttt{PENSSE}_\lambda(x) = \sum_{j=1}^{n} \left[y_j - \int z_j(t)x(t)\mathrm{d}t \right]^2 + \lambda \int [\mathcal{R}x(t)]^2 \mathrm{d}t,$$

where $\mathcal{R}$ is a roughness operator (as in section 3.6).

Let us write the parameter with respect to a function basis $\boldsymbol{\theta}(t) = (\theta_1(t), \ldots, \theta_{K_x}(t))^\mathsf{T}$, thus having $x(t) = \boldsymbol{\theta}^\mathsf{T}\mathbf{b}$, where $\mathbf{b}$ is the coefficient vector, and K_x is the length of the basis. We also do the same for the covariates $z_j(t)$ by writing $\mathbf{Z}(t) = \mathbf{C}\boldsymbol{\phi}(t)$, where $\mathbf{C}$ is the coefficient matrix of size $n \times K_z$, and define the matrices $\mathbf{J}_{\phi\theta}$ and $\mathbf{R}$ as in (3.21) and (3.22), respectively. The penalized residual sum of squares is then

$$\texttt{PENSSE}_\lambda(x) = \|\mathbf{y} - \mathbf{C}\mathbf{J}_{\phi\theta}\mathbf{b}\|^2 + \lambda \mathbf{b}^\mathsf{T}\mathbf{R}\mathbf{b},$$

and is minimized for $\mathbf{b}$ solving the equation

$$\left[(\mathbf{C}\mathbf{J}_{\phi\theta})^\mathsf{T}\mathbf{C}\mathbf{J}_{\phi\theta} + \lambda\mathbf{R} \right] \mathbf{b} = \mathbf{Z}^\mathsf{T}\mathbf{y}.$$

FLM for Functional Responses, with 2-parameters

In this model, we look at $H_1 = \left(\mathcal{L}^2([0,1]\times[0,1])\right)^n$, $H_2 = \left(\mathcal{L}^2([0,1])\right)^p$ and we assume that covariate $L : H_1 \to H_2$ is an integral operator with kernel being the parameter $\mathbf{x}(s,t) \in H_1$, defined by

$$L\mathbf{x} = \int [\mathbf{Z}(s)\mathbf{x}(s,t)]\,\mathrm{d}s, \tag{3.27}$$

where $\mathbf{Z}$ is the matrix of covariate functions $z_{ij}(s)$, and our model becomes

$$\mathbf{y}(t) = \int [\mathbf{Z}(s)\mathbf{x}(s,t)]\,\mathrm{d}s + \boldsymbol{\varepsilon}(t). \tag{3.28}$$

We shall call a parameter $\mathbf{x}(s,t)$ a 2-*parameter* to clearly emphasize the fact that it is a vector of functions of *two variables* s and t.

We are going to consider the simpler case of (3.27) in which $p = 1$, obtaining thus the simplified model

$$y_j(t) = \int z_j(s)x(s,t)\mathrm{d}s + \varepsilon_j(t), \quad j = 1, \ldots, n. \tag{3.29}$$

Our fitting criterion is

$$\texttt{LMSSE}(x) = \sum_{j=1}^{n} \int \left[y_j(t) - \int z_j(s)x(s,t)\mathrm{d}s \right]^2 \mathrm{d}t.$$

As with the model (3.26), regularization is crucial in order to avoid overfitting the data. Suppose the function x has a tensor product expansion

$$x(s,t) = \sum_{j=1}^{K_1}\sum_{l=1}^{K_2} b_{kl}\theta_l(s)\eta_j(t) = \boldsymbol{\theta}(s)^\mathsf{T}\mathbf{B}\boldsymbol{\eta}(t),$$

with respect to the basis functions $\eta_j(t), \theta_l(s)$. There are a few ways of conducting regularization to obtain $\hat{x}$:

1. By restricting the number of basis functions K_1, K_2,
2. By penalization of the roughness of the estimate $\hat{x}(s,t)$.

Let us give the computational details for the second method: let $\mathcal{R}_s$ and $\mathcal{R}_t$ be two roughness operators for the first and second variable of x. Then the penalties on x for each variable are given, respectively, by

$$\begin{aligned}
\texttt{PEN}_s(x) &= \iint [\mathcal{R}_s x(s,t)]^2 \mathrm{d}s\mathrm{d}t = \operatorname{trace}\left(\mathbf{B}^\mathsf{T}\mathbf{R}\mathbf{B}\mathbf{J}_{\eta\eta}\right)\\
\texttt{PEN}_t(x) &= \iint [\mathcal{R}_t x(s,t)]^2 \mathrm{d}s\mathrm{d}t = \operatorname{trace}\left(\mathbf{B}^\mathsf{T}\mathbf{J}_{\theta\theta}\mathbf{S}\mathbf{B}\right),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{R} &= \int [\mathcal{R}_s\boldsymbol{\theta}(s)][\mathcal{R}_s\boldsymbol{\theta}(s)]^\mathsf{T}\mathrm{d}s,\\
\mathbf{S} &= \int [\mathcal{R}_t\boldsymbol{\eta}(t)][\mathcal{R}_t\boldsymbol{\eta}(t)]^\mathsf{T}\mathrm{d}t,
\end{aligned}$$

and the matrices $\mathbf{J}_{\eta\eta}, \mathbf{J}_{\theta\theta}$ are defined as in (3.21).

If we further express $\mathbf{y} = \mathbf{C}\boldsymbol{\phi}$, as in section 3.6, then the regularized estimate for $\mathbf{B}$ is

(3.30)
$$\operatorname{vec}(\hat{\mathbf{B}}) = \left[\mathbf{J}_{\eta\eta}\otimes(\check{\mathbf{Z}}^\mathsf{T}\check{\mathbf{Z}}) + \lambda_s\mathbf{J}_{\eta\eta}\otimes\mathbf{R} + \lambda_t\mathbf{S}\otimes\mathbf{J}_{\theta\theta}\right]^{-1}(\mathbf{J}_{\theta\eta}\otimes\check{\mathbf{Z}}^\mathsf{T})\operatorname{vec}(\mathbf{C}),$$

where $\check{\mathbf{Z}} = \int \mathbf{Z}(s)\boldsymbol{\theta}(s)^\mathsf{T}\mathrm{d}s$ and $\mathbf{Z}(s) = (z_1(s),\ldots,z_n(s))^\mathsf{T}$.

Remark 3.6.2. Notice that our expansion of $x(s,t)$ is very specialized, as we considered a tensor product expansion, and it could sometimes be better to use an expansion using more general functions basis functions $\theta_j(s,t)$. Details and further references about these can be found in Ramsay & Silverman (2005, section 16.5).

Conclusions

In this section, we have seen what distinguishes functional data from multivariate data. We have presented the main ways of transforming discrete data points into functional data, through the use of basis functions. Due to the infinite dimensional nature of functional datum, roughness penalization was found to be necessary, especially for the estimation of the derivatives of the underlying function. Then the basic ideas for registration of functional data were presented, which correspond somehow to aligning the data. The functional principal component analysis – that exhibits the main features of a data set – was then introduced, together with its penalized version. This section ended with the presentation of the infinite dimensional analogues of a regression model – the functional linear model – and we saw that the general model was too complicated to be studied on its own; thus a few particular instances where considered.

4

Notions from Inverse Problems Pertinent to Functional Data

So far, we have seen how to fit some basic statistical models, but would like to use probability to conduct statistical inferences. For instance, when is a functional linear model significant? In order to achieve this goal, we have seen some basic Hilbert space theory, fundamentals of probability in abstract spaces, and also given an introduction to functional data. The last preparation we need is to get familiar with the notion of *ill-posed inverse problems*. An ill-posed inverse problem occurs essentially when we try to invert a compact linear operator. Because the singular values of a compact operator tend to zero, the inverse is an unbounded operator, making direct inversion useless. The use of *regularization* techniques is therefore necessary, and the present section is devoted to their introduction.

We will begin by introducing the notion of an ill-posed inverse problem, and will present the simplest inverse problems: the Fredholm equations of the first kind. We will then expose two methods for resolving such inverse problems: Spectral Truncation, which is based on the SVD decomposition of compact operators, and Tikhonov regularization, which is the infinite dimensional analogue to Ridge regression. We will then end this section by presenting a generalization of Tikhonov reguralization.

Let us begin with the definition of an ill-posed problem.

Definition 4.0.3. A problem is called *ill-posed* if it is not well-posed. ※

Of course, this definition is meaningless without the notion of well-posed problem, which has been first suggested by Hadamard (1952). We give here the definition by Tikhonov & Arsenin (1977).

Definition 4.0.4. Let (U, d_U) and (S, d_S) be metric spaces, where U is called the space of *initial data*, and S is the space of possible solutions. Given $u \in U$,

the problem of finding $x \in S$ such that

$$u = G(x),$$

where $G : S \to U$ is some mapping is said to be *well-posed on the pair of metric spaces* $((S, d_S), (U, d_U))$ if the following conditions are satisfied:

1. for every $u \in U$, there exists a *unique* solution $x := R(u) \in S$, where $R : U \to S$ maps each initial condition to its unique solution,

2. the mapping R is *continuous*. ※

Thus a well-posed problem is a problem with unique solutions, and where this solution depends continuously on the data, meaning that a small enough perturbation of the data will not radically change the solution.

Now we understand better the notion of ill-posed problem: it has either multiple solutions for a given initial data, or small changes in the initial data can yield significant changes in the solution. In fact, these two concept are closely related, especially in applied settings. Indeed, discontinuity of the solution translates into non-uniqueness of the solution when numerical computations are made (because of the limited numerical precision).

Nevertheless, we will only talk about this second kind of ill-posedness in this section, and will mainly rely on the exposition given by Kaipio & Somersalo (2005), unless specifically mentioned.

An *inverse problem* is the inverse of a direct problem. Though the notion of direct problem doesn't have a formal definition, a way of having an intuition is through some examples:

1. Finding the dilatation of the volume of quicksilver, knowing the change in temperature ΔT.,

2. Describing the turbulences of the gases in a furnace (where the temperature is high) from the knowledge of the furnace temperature,

3. Taking a photograph, and blurring it, which, mathematically, translates to convolving a function with a smoothing kernel.

4. Generating data $y(t_j) = x(t_j) + \varepsilon_j$, for a mesh $t_j \in [0, 2\pi]$ for $j = 1, \ldots, n$, and where x is a function with Fourier expansion $(c_0, c_1, \ldots, c_{2N})$ with respect to the Fourier basis $1, \sin(t), \cos(t), \ldots, \sin(Nt), \cos(Nt)$, and where ε_j is iid Gaussian noise.

The following are the corresponding inverse problems:

1. Finding the change in temperature ΔT from the knowledge of the change in volume of quicksilver,
2. Finding the furnace temperature from the knowledge of the gas turbulences in it,
3. Taking a blurred photograph and deblurring it,
4. Estimating the Fourier coefficients $(c_0, \ldots, c_{2N})$ given the perturbed observations y_j.

Remark 4.0.5. Often, inverse problems lead to ill-posed problems, and we shall simply say that a problem is an inverse problem instead of saying it is an *ill-posed inverse problem*.

Here is a detailed example of an inverse problem:

Example 4.0.6. Imagine a rod of unit length and unit thermal conductivity, with its ends at $x = 0$ and $x = 1$ that are set to a fixed temperature 0. Physical theory tell us that the temperature $u(x,t)$ of the point x on the rod at time t must satisfy the heat equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial u}{\partial t} = 0, \quad 0 < x < 1,\ t > 0, \tag{4.1}$$

and where the boundary conditions are

$$u(0,t) = u(1,t) = 0,\ u(x,0) = u_0(x), \qquad t \geq 0, x \in [0,1].$$

The inverse problem here is to estimate the initial condition $u_0(x)$ given the temperature of the rod at a time $T > 0$.

Writing the solution of (4.1) in terms of a Fourier expansion yields

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-(n\pi)^2 t} \sin n\pi x,$$

where the coefficients c_n are the coefficients of the Fourier sine series of the initial condition:

$$u_0(x) = \sum_{n=1}^{\infty} c_n \sin n\pi x.$$

Thus we need to find the Fourier coefficients of the final data. But this problem is ill posed, because assume that we have two initial states $u_0^{(j)}(x)$, for $j = 1, 2$ differing only by a high frequency, namely

$$u_0^{(1)}(x) - u_0^{(2)}(x) = c_N \sin N\pi x,$$

then the difference of temperatures at time T will be exponentially small:

$$u^{(1)}(x,T) - u^{(2)}(x,T) = c_N e^{-(N\pi)T} \sin N\pi x.$$

Thus, the solution $u_0(x)$ of our problem is not continuous in the data $u(x,T), x \in [0,1]$: the problem is *ill-posed.* ◀

This example shows clearly that in presence of measurement errors, finding a solution to an ill posed problem by the "usual" method will yield uninteresting results. Indeed, noise in the measurements is usually assumed to have the same variance over all frequencies, and thus we are almost guaranteed to have small perturbations in high frequencies. We therefore need some other technique – via a regularization method – to try and solve inverse problems.

4.1 Fredholm Equations of the First Kind

The two regularization methods we are going to present are easier to understand when they are applied to *linear* inverse problems called *Fredholm equation of the first kind.* These sorts of inverse problems are similar to those we will encounter with functional data.

Definition 4.1.1. Let H_1, H_2 be Hilbert spaces, and $A : H_1 \to H_2$ be a compact linear operator. Given $y \in H_2$, the problem of finding $x \in H_1$ such that

$$Ax = y \tag{4.2}$$

is called a *Fredholm equation of the first kind.* ※

Because of the linearity of A, we know that a unique solution to (4.2) exists if and only if $y \in \mathrm{Im}(A)$ and $\ker(A) = \{0\}$. In practice, the vector y is measured and contains some error. We have thus an approximate equation

$$Ax \approx y,$$

and because the inverse of A, when defined, is rarely continuous (unless both H_j's are finite dimensional), we *cannot* take $x \approx A^{-1}y$ as a solution. A way of avoiding such a problem is through spectral truncation.

4.2 Spectral Truncation

Let us denote by $\langle x, y\rangle_j$ the scalar product in H_j. In order to understand spectral truncation, we need to use the singular value decomposition of a compact operator (Theorem 1.7.1), which we reformulate here for completeness:

Proposition 4.2.1. *Let H_1, H_2 be Hilbert spaces, $A : H_1 \to H_2$ be a linear compact operator and A^* be the adjoint operator of A. Then*

1. *The spaces H_j allow the following* orthogonal *decompositions:*

$$\begin{aligned} H_1 &= \ker(A) \oplus \ker(A)^\perp &&= \ker(A) \oplus \overline{\mathrm{Im}(A^*)}, \\ H_2 &= \overline{\mathrm{Im}(A)} \oplus \mathrm{Im}(A)^\perp &&= \overline{\mathrm{Im}(A)} \oplus \ker(A^*). \end{aligned}$$

2. *There exists orthonormal sequences $(v_n) \subset H_1$, $(u_n) \subset H_2$ and a* decreasing *sequence of* positive *numbers (λ_n) with $\lim_{n\to\infty} \lambda_n = 0$ such that*

$$\overline{\mathrm{Im}(A)} = \overline{\mathrm{span}}\{(u_n)\},\ (\ker(A))^\perp = \overline{\mathrm{span}}\{(v_n)\},$$

and the operator A has the representation

$$Ax = \sum_n \lambda_n \langle x, v_n\rangle_1 u_n. \tag{4.3}$$

The equation (4.3) is called the singular value decomposition *(SVD) of A, and the system (v_n, u_n, λ_n) is called the* singular system *of the operator A. We will often omit the subindex 1 for the scalar product in (4.3) for notation simplicity.*

3. *The equation $Ax = y$ has a solution if, and only if*

$$y = \sum_n \langle y, u_n\rangle u_n, \quad \sum_n \frac{1}{\lambda_n^2} |\langle y, u_n\rangle|^2 < \infty,$$

in which case the solutions of the equation are given by

$$x = x_0 + \sum_n \frac{1}{\lambda_n} \langle y, u_n\rangle v_n, \quad x_0 \in \ker(A). \tag{4.4}$$

This proposition tells us that the orthogonal projection $P : H_2 \to \overline{\mathrm{Im}(A)}$ is defined by

$$Py = \sum_n \langle y, u_n\rangle u_n,$$

and hence we get the above bound for the error $\|Ax - y\|$ of our estimate x by the equation

$$\begin{aligned} \|Ax - y\|^2 &= \|Ax - Py - (Id - P)y\|^2 \\ &= \|Ax - Py\|^2 + \|(Id - P)y\|^2 \geq \|(Id - P)y\|^2. \end{aligned}$$

Here Id is the identity operator on H_2. Thus if the error ε on the measured y has a non-zero component in the space orthogonal to $\overline{\mathrm{Im}(A)}$, that is, $(Id - P)\varepsilon \neq 0$, then the equation $Ax = y$ cannot be solved exactly. The best we can do is to solve the projected equation

$$Ax = Py.$$

But doing this doesn't guarantee that we can find a solution of the form (4.4), because the convergence of the series is not guaranteed in the presence of noise. Even if the series converges, the estimate might be very bad. Indeed, assume that the true y is 0, and suppose we observe $\hat{y} = K\lambda_N u_N$ with $K > 0$ large. The error is $\|y - \hat{y}\| = K\lambda_N$, which can be made arbitrarily small for a large N. However, the estimate will be $\hat{x} = Kv_n$, and our error with respect to the true $x = 0$ will be $\|\hat{x} - x\| = K$, which is large. Hence we see that the problem of estimating the coefficients of $x \in \ker(A)^\perp$ from the projected data Py is an inverse problem by itself!

The problem comes from the fact that the sequence (λ_n) converges to 0: if we think of coefficients of $y \in \ker(A^*)^\perp$ in the (u_n) basis as frequencies, with the higher frequencies being the coefficients of u_n with larger n, then we see that a small error in the higher frequencies of y yields a big error in the estimated x, explicitly, an error of $\epsilon > 0$ in the N^{th} frequency of y gives an error or ϵ/λ_N in the estimated x. Thus most of the error comes from the high frequencies of y, and a natural way of overcoming this is to cut out these high frequencies, in other words to *truncate the spectrum* of y, and hence the term *spectral truncation.*

More rigorously, we are going to solve the problem

$$Ax = P_k y, \quad k \in \mathbb{N}, \tag{4.5}$$

where $P_k : H_2 \to \mathrm{span}\{u_1, \ldots, u_k\}$ is the orthogonal projection defined by $P_k y = \sum_{n=1}^{k} \langle y, u_n \rangle u_n$. This yields the solutions

$$x_k = x_0 + \sum_{n=1}^{k} \frac{1}{\lambda_n} \langle y, u_n \rangle v_n, \quad x_0 \in \ker(A), \tag{4.6}$$

to equation (4.5). Notice that the distance between Ax_k and Py is decreasing in k, namely

$$\|Ax_k - Py\|^2 = \|(P_k - P)y\|^2 = \sum_{n=k+1}^{\infty} |\langle y, u_n \rangle|^2 \to 0.$$

The last problem we have to tackle is the possible non-injectivity of A. By orthogonality, notice that

$$\|x_k\|^2 = \|x_0\|^2 + \sum_{n=1}^{k} \frac{1}{\lambda_n^2} |\langle y, u_n \rangle v_n|^2 \leq \sum_{n=1}^{k} \frac{1}{\lambda_n^2} |\langle y, u_n \rangle v_n|^2.$$

Hence a consistent way of choosing x_k is to pick the unique one that is in $\ker(A)^\perp$, or equivalently, the one with minimal norm. This leads to the following definition:

Definition 4.2.2. Let $A : H_1 \to H_2$ be a compact operator with singular system (v_n, u_n, λ_n). By the *truncated SVD approximation* (TSVD) of the problem $Ax = y$, we mean the problem of finding $x \in H_1$ such that

$$Ax = P_k y, x \perp \ker(A),$$

for some $k \leq 1$. ※

We have already seen that the solution to the TSVD problem is given by

$$x_k = \sum_{n=1}^{k} \frac{1}{\lambda_n} \langle y, u_n \rangle v_n.$$

The choice of the truncation parameter k is crucial in the TSVD. If the noise is assumed to be random, then techniques such as cross-validation or a version of the Akaike Information Criterion (Yao et al. 2005, Section 2.5) can be used. However, if an exact noise level is known, the *discrepancy principle* can be used. It states that we should not attempt to be more precise than the noise level. That is, if we now that the noise is roughly ϵ, namely

$$\|y - \hat{y}\| \approx \epsilon,$$

where $\hat{y}$ is the observation of the true y, then we should pick k to be the smallest integer such that

$$\|Py - P_k y\| \leq \epsilon.$$

4.3 Tikhonov Regularization

As discussed in page 88, we have seen that the problem of ill-posedness comes from the λ_n tending to zero when $n \to \infty$, and this makes the norm of our estimate grow to infinity as $k \to \infty$. The idea of Tikhonov regularization is to try to minimize the norm of $r = y - Ax$ by putting a constraint on the

norm of x. This is similar to what is done in Stein estimation, when the MLE is shrunken towards the origin.

Hence we look at the problem

$$\min_x \|y - Ax\|^2 \text{ such that } \|x\|^2 \leq M, \tag{4.7}$$

for $M \leq 0$, provided a minimizer exists. The Lagrange form of this problem is

$$\min_x \left\{ \|y - Ax\|^2 + \delta \|x\|^2 \right\}, \tag{4.8}$$

for a $\delta \geq 0$. This motivates the following definition:

Definition 4.3.1. Let $\delta > 0$. The *Tikhonov regularized solution* $x_\delta \in H_1$ is the minimizer of the functional

$$F_\delta(x) = \|y - Ax\|^2 + \delta \|x\|^2,$$

provided such a minimizer exists. The parameter δ is called the *regularization parameter*. ※

Notice that Tikhonov regularization is the same as Ridge regression when A is a matrix and x, y are vectors. The good news about the Tikhonov regularized solution is that there is a closed-form formula for computing it, which is given by the following Theorem.

Theorem 4.3.2. *Let $A : H_1 \to H_2$ be a compact operator with singular system (v_n, u_n, λ_n). Then the Tikhonov regularized solution exists, is unique, and is given by the formula*

$$x_\delta = (A^*A + \delta I)^{-1} A^* y = \sum_n \frac{\lambda_n}{\lambda_n^2 + \delta} \langle y, u_n \rangle v_n. \tag{4.9}$$

Notice the effect of Tikhonov regularization on the frequency $\langle y, u_n \rangle$ of y in comparison with direct inversion $A^{-1}y$: the frequency is multiplied by a factor $\frac{1}{\lambda_n}$ in the latter, which tends to infinity as $n \to \infty$, whereas in Tikhonov regularization, the frequency is multiplied by $\lambda_n/(\lambda_n^2 + \delta)$, which tends to 0 as $n \to \infty$, because $\delta > 0$. Also, notice that if $\lambda_n \gg \delta$, which is possible for small values of n, then

$$\frac{\lambda_n}{\lambda_n^2 + \delta} = \frac{1}{\lambda_n} \left(\frac{1}{1 + \frac{\delta}{\lambda_n}} \right) \approx \frac{1}{\lambda_n}.$$

Hence Tikhonov regularization can be seen as a kind of smoothed spectral truncation: it is almost like the TSVD for the low frequencies, and then it uses less and less of the higher frequencies of y for estimating x, where the TSVD just doesn't use the high frequencies from one point on.

The essential question in Tikhonov regularization is about the choice of the regularization parameter $\delta > 0$. As with the choice of the smoothing parameter in Section 3.2, a cross-validation method can be used for choosing δ. However, if the exact noise level in the data is known, one method of choosing δ it is given by the *Morozov discrepancy principle*, which is similar to the discrepancy principle for the TSVD.

Let x_δ be defined by (4.9), and define the discrepancy function $f : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$f(\delta) = \|Ax_\delta - y\|. \tag{4.10}$$

The Morozov discrepancy principle states that δ should be chosen such that

$$f(\delta) = \epsilon, \tag{4.11}$$

where ϵ is the noise level. In other words, it states that we shouldn't attempt to be more precise than the noise level. The following Theorem tells us when this principle can be applied.

Theorem 4.3.3. *The discrepancy function* (4.10) *is continuous, strictly increasing and*

$$\|Py\| \leq f(\delta) \leq \|y\|, \quad \delta \geq 0,$$

where P is the orthogonal projection $H_2 \to \mathrm{Im}(A)^\perp$. Hence the equation (4.11) *is satisfied if, and only if we have $\|Py\| \leq \epsilon \leq \|y\|$.*

Remark 4.3.4. The first inequality $\|Py\| \leq \epsilon$ should be naturally satisfied, for the component Py orthogonal to $\mathrm{Im}(A)$ should be due to noise. The second inequality $\|y\| \leq \epsilon$ means that the signal level is greater than the noise. Indeed, we would expect this, because otherwise, if $\|y\| < \epsilon$, we can take $x = 0$ in view of our discrepancy principle.

4.4 A Generalization of Tikhonov Regularization

Up to now, we have considered the operator $A : H_1 \to H_2$ to be linear. Tikhonov regularization is also applicable when this is not the case. We need to introduce a weaker condition on A, which is *Fréchet differentiability*.

Definition 4.4.1. An operator $A : H_1 \to H_2$ is *Fréchet differentiable* at $x_0 \in H_1$ if it allows an expansion

$$A(x_0 + z) = A(x_0) + R_{x_0} z + W_{x_0}(z),$$

where $R_{x_0} : H_1 \to H_2$ is a continuous linear operator and

$$\|W_{x_0}(z)\| \leq \|z\| \varepsilon_{x_0}(z),$$

where the functional $z \mapsto \varepsilon_{x_0}(z) \in H_2$ tends to zero as $z \to 0 \in H_1$. ※

Think of Fréchet differentiability as an order 1 Taylor expansion. As with Taylor expansions, we are going to use R_{x_0} to linearize A around x_0, and thus we have the following approximation to our functional F_δ defined in definition 4.3.1.

$$F_\delta(x) \approx \check{F}_\delta(x; x_0) = \|R_{x_0} x - g(y, x_0)\|^2 + \delta \|x\|^2,$$

where

$$g(y, x_0) = y - A(x_0) + R_{x_0} x_0.$$

We have seen in the previous section that the minimizer of $\check{F}_\delta(x; x_0)$ is

$$x = (R_{x_0}^* R_{x_0} + \delta I)^{-1} R_{x_0}^* g(y, x_0). \tag{4.12}$$

This is the basis to an iterative method for minimizing $F_\delta(x)$. However, it appears that it is better to take $x_0 + s(x - x_0)$ as our new x_0, with some step-size s, instead of just taking x for our next iteration. The complete iterative method is

1. Pick an initial x_0 and set $k = 0$,
2. Calculate R_{x_k},
3. Compute x from (4.12) with x_k instead of x_0, and define $\Delta x = x - x_k$.
4. Find $s > 0$ minimizing the function

$$f(s) = \|A(x_k + s\Delta x) - y\|^2 + \|x_k + s\Delta x\|^2,$$

5. Set $x_{k+1} = x_k + s\Delta x$ and increase $k \leftarrow k + 1$,
6. Repeat steps 2.-5 until the method converges.

Conclusions

We have seen in this section some examples of inverse problems. In particular, we observed that the estimation of high frequencies is ill-posed. More generally, we have seen the easiest canonical form of inverse problems: Fredholm equations of the first kind. Then we presented two methods to solve them: Spectral truncation – based on the SVD decomposition of compact operators – and Tikhonov regularization – which is essentially an infinite dimensional Ridge regression. We have seen that an essential part in these regularization methods was the choice of the parameter k, and that they were some techniques for choosing it: the discrepancy principles, if we assumed the observation noise to be of deterministic norm, or the cross-validation techniques if it was considered random. Eventually, we saw a generalization of Tikhonov regularization to non-linear equations, which essentially boils down to considering local linear approximations of the equation and proceeding by iterations.

5

Testing for Functional Data

Now that we have some knowledge of Hilbert spaces, probability in abstract spaces, functional data analysis, and that we have seen some basic techniques for dealing with inverse problems, we can turn our attention to statistical inference for functional data.

We will first start by giving a review of Mas (2007). The article presents a test statistic for the mean of sample curves. We will see that an inverse problem occurs when we try to "free" the centered sample mean $\overline{X} - m_0$ from dependencies on the unknown distribution of the data, and that Tikhonov regularization is necessary. Then, we will present a paper of Cardot et al. (2003), which gives a test for the functional linear model with scalar responses. Again, an inverse problem will naturally arise, and Spectral truncation will be used.

5.1 Testing for the Mean of Random Curves

Let us review the article of Mas (2007), which proposes an asymptotic test for the mean of random curves. An interesting feature of this article is that it considers $X_1, \dots, X_n$ being an i.i.d. sample of $L^2([0,1])$-valued random variable X, without the assumption that X is Gaussian. It also supposes a mild condition on the rate of decrease of the eigenvalues of the operator. Letting m be the mean of the random variable X, the paper gives a test for

$$\begin{cases} H_0: & m = m_0, \text{ against} \\ H_a: & m \neq m_0, \end{cases}$$

where m_0 is a given curve in $L^2([0,1])$. Let us denote by $\langle f, g \rangle$ ($\|f\|$) the scalar product (respectively the norm) in $L^2([0,1])$. The first assumption about X is

A1: $\mathbb{E}\|X\|^4 < \infty$.

This allows us to define the covariance operator $\Gamma : L^2([0,1]) \to L^2([0,1])$ of X under H_0:

$$\Gamma f(t) = \mathbb{E}\left[\langle X - m_0, f\rangle (X - m_0)(t)\right] = \int_0^1 r(t,s) f(s) \mathrm{d}s, \tag{5.1}$$

where $r(t,s) = \mathbb{E}\left[(X(t) - m_0(t))(X(s) - m_0(s))\right]$ and the second equality comes from Fubini's Theorem. Let us also define the empirical covariance operator under H_0:

$$\Gamma_n f(t) = \frac{1}{n} \sum_{k=1}^{n} \left[\langle X_k - m_0, f\rangle (X_k - m_0)(t)\right] \tag{5.2}$$

$$= \int_0^1 \left[\frac{1}{n} \sum_{k=1}^{n} (X_k - m_0)(t)(X_k - m_0)(s)\right] f(s) \mathrm{d}s. \tag{5.3}$$

The assumption **A1** implies that the operator Γ is nuclear (see Proposition 2.5.9), and self-adjoint. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ denote its eigenvalues associated with the eigenelements $e_1, e_2, \ldots \in L^2([0,1])$. The spectral Theorem for compact self-adjoint operators (Theorem 1.6.19) tells us that Γ can be written as

$$\Gamma f = \sum_{k=1}^{\infty} \lambda_k \langle e_k, f\rangle e_k. \tag{5.4}$$

For $u, v \in L^2([0,1])$, let $u \otimes v$ denote the operator on $L^2([0,1])$ defined by $u \otimes v(f) = \langle u, f\rangle v$. Then Γ can be rewritten

$$\Gamma f = \mathbb{E}\left[(X - m) \otimes (X - m)\right] = \sum_{k=1}^{\infty} \lambda_k (e_k \otimes e_k). \tag{5.5}$$

The Hilbert-Schmidt operator $\Gamma^{1/2}$, the square root of Γ, will play a key role in this section, and is defined by

$$\Gamma^{1/2} = \sum_{k=1}^{\infty} \sqrt{\lambda_k} (e_k \otimes e_k). \tag{5.6}$$

The second assumption about X is:

A2: The covariance operator Γ of X is injective, that is, $\lambda_k > 0$ for all $k \geq 1$.

When **A2** holds, the inverse $\Gamma^{-1/2}$ of $\Gamma^{1/2}$ is an unbounded operator, defined on the dense subspace of $L^2([0,1])$:

$$\mathcal{D} = \left\{ \sum_{k=1}^{\infty} \alpha_k e_k \in L^2([0,1]) : \sum_{k=1}^{\infty} \frac{\alpha_k^2}{\lambda_k} < \infty \right\}.$$

The Finite Dimensional Setting

In order to understand who to construct such a test for the mean of random curves, let us first consider the finite dimensional setting, in which $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an i.i.d. sample of vectors of $\mathbb{R}^p$ with mean $\mathbf{m} \in \mathbb{R}^p$. In order to test for

$$\begin{cases} H_0: & \mathbf{m} = \mathbf{m}_0, \text{ against} \\ H_a: & \mathbf{m} \neq \mathbf{m}_0, \end{cases}$$

we define the empirical covariance matrix $M_n(p)$ of our sample by

$$M_n(p) = \frac{1}{n} \sum_{k=1}^{n} (\mathbf{X}_k - \mathbf{m_0})(\mathbf{X}_k - \mathbf{m_0})^{\mathsf{T}}, \tag{5.7}$$

under the assumption H_0. This matrix is often supposed invertible for n large enough, and then the test statistic is derived from

$$M_n(p)^{-1/2} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (\mathbf{X}_k - \mathbf{m}_0) \right), \tag{5.8}$$

which converges under H_0, as $n \to \infty$, to a Gaussian random vector with covariance matrix I, the identity matrix.

The Infinite Dimensional Setting

In the infinite dimensional setting, we still have that

$$S_n = n^{-1/2} \sum_{k=1}^{n} (X_k - m_0)$$

converges in distribution to a $L^2([0,1])$-valued Gaussian random variable G by a version of the Central Limit Theorem (see Theorem 2.6.7). However, the distribution of the random variable G depends on the covariance struture Γ of X, and hence on the unknown λ_j's. In order to free S_n from the dependence on these unknown eigenvalues, one might be tempted to compute "$\Gamma_n^{-1/2} S_n$", where Γ_n is the empirical covariance operator is defined by

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - m_0) \otimes (X_i - m_0), \tag{5.9}$$

under the null hypothesis $\mathbb{E}X_i = m_0$.

However, $\Gamma_n^{1/2}$ is a finite rank operator, and is therefore not invertible. Even if we knew the true covariance operator Γ, we couldn't use "$\Gamma^{-1/2}G$" as a test statistic, for $\Gamma^{-1/2}$ is an unbounded operator.

Indeed, the problem we are looking at is the following: if Y is a $L^2([0,1])$-valued random variable with the identity as covariance operator, then

$$G \sim \Gamma^{1/2} Y,$$

where $X \sim Z$ means that the random variables X and Z have the same distribution. Ideally, we would take $\Gamma^{-1/2}G$ as a test statistic. But notice that even if $G = \Gamma^{1/2}Y$, the problem of recovering Y from the knowledge of G would be ill-posed, and some regularization would be needed.

Our situation is a bit more complicated here: we know that S_n converges in distribution to G, some Gaussian random variable with covariance operator Γ, of which we only have an estimate Γ_n. Thus for a large n, we would assume that

$$S_n \dot{\sim} \Gamma_n^{1/2} Y,$$

where $\dot{\sim}$ means "approximatively equal in distribution". Then again, even if $\Gamma_n^{1/2}Y = S_n$, recovering Y from the knowledge of S_n would be an ill-posed problem, because Γ_n is of finite rank.

Our goal is consequently double: we have to approximate $\Gamma^{-1/2}$ by a pseudo-inverse L_n, built from Γ_n, and then study the convergence of $L_n S_n$. Before getting into more details, let us make some remarks:

1. The pseudo-inverse L_n is built from Γ_n, and is thus *random*,
2. At the limit $n \to \infty$, we would like the operator norm $\|L_n\|$ of L_n to be a non-decreasing sequence tending to infinity, to approximate "perfectly" $\Gamma^{-1/2}$ at the limit $n \to \infty$.

The Choice of Regularization Method

In order to regularize $\Gamma^{-1/2}$, we can either use spectral truncation (section 4.2) or Tikhonov regularization. With spectral truncation, we would use the operator

$$\Gamma_{(p)}^{-1/2} = \sum_{k=1}^{p} \lambda_k^{-1/2} (e_k \otimes e_k),$$

which is a bounded operator with norm $\lambda_p^{-1/2}$, instead of the operator $\Gamma^{-1/2}$. This method has the disadvantage that we need to estimate $\Gamma_{(p)}^{-1/2}$ through functional principal component analysis of Γ_n, where Γ_n is defined in (5.9).

However, this is a very unstable procedure. Denote by e_{kn} the k^{th} eigenvector of Γ_n such that $\|e_{kn}\| = 1$, and define $e'_{kn} = sgn(\langle e_{kn}, e_k\rangle)e_k$, which is just choosing between $\pm e_k$ to get $\langle e_{kn}, e'_{kn}\rangle \geq 0$. Then if the subspace $\mathcal{V}_k = \{v \in L^2([0,1]) : \Gamma v = \lambda_k v\}$ has dimension 1, Lemma 4.3 of Bosq (2000) gives us the bound

$$\|e_{kn} - e'_{kn}\| \leq a_k \|\Gamma_n - \Gamma\|_{op}, \tag{5.10}$$

where $a_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}$,

$$a_k = 2\sqrt{2}\max[(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}], \quad k \geq 2,$$

and $\|\Gamma\|_{op}$ is the operator norm of Γ (see Proposition 1.3.5). Notice that (5.10) is an inequality between random elements (e_{kn} and Γ_n are random), and

1. The term a_k grows to infinity as $k \to \infty$, because the sequence (λ_k) is a Cauchy sequence tending to 0,

2. We have the following asymptotic bound on $\|\Gamma_n - \Gamma\|$ from Bosq (2000, Corollary 4.1):

   $$\|\Gamma_n - \Gamma\| = O\left(\left(\frac{\log n}{n}\right)^{1/2}\right) \quad a.s., \tag{5.11}$$

   where $X_n = O(g(n))$ a.s. if $\exists C > 0$ such that

   $$\mathbb{P}\left[\lim_{n\to\infty} \frac{X_n}{g(n)} \leq C\right] = 1.$$

3. Furthermore, if we consider the $L^2(\mathbb{P})$-norm of (5.10), then

   $$\sqrt{\|e_{kn} - e'_{kn}\|} \leq a_k \frac{1}{\sqrt{n}} \sqrt{\mathbb{E}\|X\|^4},$$

   by Example 2.6.8.

   Thus we have no guarantee that $\|e_{kn} - e'_{kn}\|$ converges fast to 0.

All these reasons motivate Mas (2007) to use a variant of Tikhonov regularization. Instead of using the "classical" Tikhonov regularization $(\Gamma + kI)^{-1}\Gamma^{1/2}$ to approximate $\Gamma^{-1/2}$, the operator $(\Gamma + \alpha_n I)^{-1/2}$ is chosen, with operator norm $1/\sqrt{\alpha_n}$, where (α_n) a decreasing sequence of positive real numbers tending to 0. It is important to note that, conversely to $\Gamma_{(p)}^{-1/2}$, the norm

of $(\Gamma + \alpha_n I)^{-1/2}$ and $(\Gamma + \alpha_n I)^{-1/2}$ are non-random and do not depend on the rate of decay of the eigenvalues of Γ. Indeed, notice that

$$(\Gamma + \alpha_n I)^{-1/2} = \sum_{i=1}^{\infty} (\lambda_i + \alpha_n)^{-1/2} e_i \otimes e_i,$$

where λ_i's are allowed to be equal to zero. Thus

$$\|(\Gamma + \alpha_n I)^{-1/2}\| = \sup_{i=1,2,\ldots} \frac{1}{\sqrt{\lambda_i + \alpha_n}} = \frac{1}{\sqrt{\alpha_n}},$$

and the same result holds for the empirical counterpart because it has finite rank. Thus, the empirical version $(\Gamma_n + \alpha_n I)^{-1/2}$ is chosen for the test statistic.

Main results

Before we state the main result of Mas (2007), let us recall the assumptions for completeness purposes:

A1: $\mathbb{E}\|X\|^4 < \infty$,

A2: The operator Γ is injective,

The hypotheses we wish to test for are

$$\begin{cases} H_0 : & m = m_0, \text{ against} \\ H_a : & m \neq m_0. \end{cases}$$

Letting

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - m_0),$$

and

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - m_0) \otimes (X_i - m_0),$$

we know by the central limit Theorem that S_n converges weakly to a Gaussian random variable G with mean zero and covariance operator Γ. Under the further assumption

A3: The random process X is continuous in the mean (see Definition 2.7.3),

the Karhunen-Loève expansion (Theorem 2.4.10) tells us that G admits the representation

$$G = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \eta_k e_k,$$

where the η_k's are independent Gaussian random variables with unit variance, and the convergence is uniform in $L^2(\Omega, \mathcal{O}, \mathbb{P})$.

Let $(k_n) \subset \mathbb{N}$ be a sequence increasing to infinity, and $(\alpha_n) \subset \mathbb{R}_+$ be a sequence decreasing to zero, and define

$$\hat{c}_n = \sum_{p=1}^{k_n} \frac{\hat{\lambda}_p}{\hat{\lambda}_p + \alpha_n} \quad \text{and} \quad \hat{d}_n = \sqrt{\sum_{p=1}^{k_n} \left(\frac{\hat{\lambda}_p}{\hat{\lambda}_p + \alpha_n} \right)^2}, \tag{5.12}$$

where $\hat{\lambda}_p$ is the empirical p^{th} eigenvalue of the empirical covariance operator Γ_n (in fact they actually depend on n).

The following Theorem gives the test statistic:

Theorem 5.1.1 (Mas (2007))**.** *Under assumptions **A1**, **A2** and **A3**, there exists two conjugated sequences (k_n) and (α_n) such that, under H_0,*

$$\hat{T}_n \xrightarrow{w} N(0, 3),$$

where $\hat{T}_n$ is the statistic

$$\hat{T}_n = \frac{1}{\hat{d}_n} \left(\|(\Gamma_n + \alpha_n I)^{-1/2} S_n\|^2 - \hat{c}_n \right), \tag{5.13}$$

and $\hat{c}_n, \hat{d}_n$ are defined in (5.12).

Notice that this Theorem doesn't suppose any particular distribution of the data, except that the sample is iid and assumptions **A1**, **A2**, and **A3** hold. However, it does not tell us how to choose k_n and α_n. Under a further assumption, the next Theorem answers this question:

Theorem 5.1.2. *Let $0 < \nu < 1/2$, and assume **A1**, **A2**, **A3** and*

***A4**: for some $M > 0$ and $\varepsilon > 0$, we have $\lambda_p \leq \frac{M}{p(\log p)^{1+\varepsilon}}, \quad \forall p > 1$.*

Then with $k_n = n^{1/2-\nu}$, and $\alpha_n = (\log n)^{-\varepsilon/(1+\nu\varepsilon)}$, if H_0 holds, we have

$$\hat{T}_n \xrightarrow{w} N(0, 3).$$

5.2 Testing in the Functional Linear Model with Scalar Responses

In the previous section, we have seen a test for the mean of functional data. It used the so-called Tikhonov regularization to build a pseudo-inverse of the empirical covariance matrix, which was then used in the test statistic.

In this section, we are going to present a test for the Functional Linear Model with Scalar Responses (FLMSR), based on Cardot et al. (2003). First, let us introduce some notation: we shall denote by $\langle f, g\rangle = \int_0^1 f(t)g(t)\mathrm{d}t$ the scalar product on the Hilbert space $L^2([0,1])$ of *real*-valued square-integrable function on $[0,1]$, and by $\|f\|_2$ the norm it defines.

The FLMSR is defined by

$$Y = \Psi(X) + \varepsilon, \tag{5.14}$$

where the assumptions are

A1: X is a centered random variable with values in $L^2([0,1])$, and $\mathbb{E}\|X\|_4 < \infty$,

A2: $\Psi : L^2([0,1]) \to \mathbb{R}$ is a continuous linear operator,

A3: Y is a real-valued random variable,

A4: ε, which represents the noise, is a real-valued random variable with mean zero and variance σ^2, and is independent from X.

By the Riesz-Fréchet representation Theorem, we know that there exists a $\psi \in L^2([0,1])$ such that

$$\Psi(f) = \langle \psi, f\rangle, \quad \forall f \in L^2([0,1]),$$

and hence we can rewrite our model as

$$Y = \int_0^1 \psi(t)X(t)\mathrm{d}t + \varepsilon = \langle \psi, X\rangle + \varepsilon, \tag{5.15}$$

with $\psi \in L^2([0,1])$. The problem we are going to consider is that of giving a confidence set for the function ψ, which is the infinite dimensional equivalent of the parameter $\boldsymbol{\beta}$ in a regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5.16}$$

where $\mathbf{y} \in \mathbb{R}^p, \boldsymbol{\beta} \in \mathbb{R}^n$, and $\mathbf{X}$ is a $p \times n$ matrix.

Because the notion of confidence set is dual to that of testing, our problem is actually of finding a test for the hypothesis:

$$\begin{cases} H_0: & \psi = \psi_0, \text{ against} \\ H_a: & \psi \neq \psi_0, \end{cases}$$

where $\psi_0 \in L^2([0,1])$. Notice that assumption 1 tell us that the covariance operator $\Gamma = \mathbb{E}[X \otimes X] : L^2([0,1]) \to L^2([0,1])$ of X, defined explicitly by

$$\Gamma f = \mathbb{E}[\langle X, f\rangle X] = \int_0^1 \mathbb{E}[X(\cdot)X(s)]\, f(s)\mathrm{d}s, \quad f \in L^2([0,1]),$$

is nuclear, self-adjoint and non-negative. Let us also define the cross-covariance operator $\Delta = \mathbb{E}[X \otimes Y] : L^2([0,1]) \to \mathbb{R}$, which is explicitly defined by

$$\Delta f = \int_0^1 \mathbb{E}[X(t)Y]\, f(t)\mathrm{d}t, \quad f \in L^2([0,1]).$$

Notice that the three operators Ψ, Γ, and Δ are linked by the relation

$$\Delta = \Psi\Gamma$$

because

$$\begin{aligned} \Psi\Gamma &= \langle \psi, \mathbb{E}[X \otimes X]\rangle \\ &= \mathbb{E}[\langle \psi, X \otimes X\rangle] \\ &= \mathbb{E}[X \otimes \langle \psi, X\rangle] \\ &= \mathbb{E}[X \otimes Y] = \Delta. \end{aligned}$$

Before describing the test procedure, notice that Proposition 4.2.1 gives us the orthogonal decomposition

$$L^2([0,1]) = \ker(\Gamma) \oplus \overline{\mathrm{Im}(\Gamma^*)} = \ker(\Gamma) \oplus \overline{\mathrm{Im}(\Gamma)},$$

where the second equality holds because Γ is self-adjoint. Using this decomposition to get $\psi = \psi_1 + \psi_2$, we notice that

$$\begin{aligned} \mathbb{E}Y^2 &= \mathbb{E}\left[(\langle X, \psi_1\rangle + \langle X, \psi_2\rangle)^2\right] \\ &= \mathbb{E}\left[\langle X, \psi_1\rangle^2\right] + 2\mathbb{E}\left[\langle X, \psi_1\rangle\langle X, \psi_2\rangle\right] + \mathbb{E}\left[\langle X, \psi_2\rangle^2\right] \\ &= \mathbb{E}\left[\langle X, \psi_1\rangle^2\right] + \mathbb{E}\left[\langle X, \psi_2\rangle^2\right], \end{aligned}$$

where the second equality holds because

$$\mathbb{E}\left[\langle X, \psi_1\rangle\langle X, \psi_2\rangle\right] = \mathbb{E}\left[\int X_s\psi_1(s)\mathrm{d}s \int X_t\psi_2(t)\mathrm{d}t\right] = \int (\Gamma\psi_1)(t)\psi_2(t)\mathrm{d}t = 0.$$

But the random variable $\langle X, \psi_1 \rangle$ is equal to zero with probability 1 because

$$\mathbb{E}\left[\langle X, \psi_1 \rangle^2\right] = \langle \Gamma\psi_1, \psi_1 \rangle = 0.$$

Therefore, if we do not assume that $\psi_1 = 0$, which is equivalent to assuming $\psi \in \overline{\text{Im}(\Gamma)}$, the parameter ψ would be unidentifiable in our model. Observe that in the finite dimensional setting (5.16), this would happen if, and only if, the columns of the design matrix $\mathbf{X}$ are not linearly independent.

So from now on, we will assume that $\psi \in \overline{\text{Im}(\Gamma)}$, and for simplicity, we shall denote $H = \overline{\text{Im}(\Gamma)}$, which is a separable Hilbert space. If H is a finite dimensional, then our problem simplifies to fitting a regression model, which have been extensively studied (e.g. Seber (1977)). We are thus going to consider that H is infinite dimensional.

If we restrict the operators Γ, Δ to H, then we still have $\Delta = \Psi\Gamma$, and the operator Γ admits the decomposition

$$\Gamma = \sum_{i=1}^{\infty} \lambda_i e_i \otimes e_i,$$

where (λ_i, e_i) are the pairs of eigenvalue/eigenvectors of Γ, (e_i) is a complete orthonormal sequence in H and $(\lambda_i) \subset \mathbb{R}_+$ is a decreasing sequence tending to zero, $\lambda_i > 0$ for all $i \in \mathbb{N}$.

Our testing problem further simplifies to

$$\begin{cases} H_0 : & \psi = 0, \text{ against} \\ H_a : & \psi \neq 0, \end{cases}$$

because we can replace Y by $Y - \langle \psi_0, X \rangle = \langle 0, X \rangle + \varepsilon = \varepsilon$.

The next proposition gives the idea of our test. Recall that $\Gamma : H \to H$ and $\Delta : H \to H$.

Proposition 5.2.1. *Let $\psi \in H = \overline{\text{span}(e_i)}$. Then $\psi = 0$ if, and only if $\Delta = 0$.*

Proof. First let $\psi = 0$. Then $\Psi = 0$ and $\Delta = \Psi\Gamma = 0$. Conversely, if $\Delta = 0$, then taking the adjoints yields $\Gamma\Psi^* = 0$, because Γ is self-adjoint. But Γ is injective, thus $\Psi^* = 0$ and $\Psi = 0$. This latter is equivalent to $\psi = 0$, and the proof is complete. □

This proposition tells us that testing for $\psi = 0$ is the same as testing for $\Delta = 0$. If we have an i.i.d sample $(X_i, Y_i), i = 1, \ldots, n$, then by the strong law of large numbers for Hilbert spaces, we know that

$$\Delta_n = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes Y_i$$

converges almost surely to Δ. Thus from the continuous mapping Theorem, the operator norm $\|\Delta_n\|$ converges almost surely to $\|\Delta\| = 0$. So we are going to construct a test procedure based on the norm of the empirical cross-covariance operator Δ_n.

We are going to use the central limit Theorem for Hilbert spaces to have an asymptotic distribution for Δ_n. But first, let us notice that for any $T \in H^* = \mathcal{L}(H, \mathbb{R})$, the Riesz-Fréchet Theorem tells us that $T = \langle f, \cdot \rangle$ for a unique $f \in H$, and $\|T\| = \|f\|$. Thus $T = v \otimes 1$ in our tensor product notation, and this gives us a simple way of identifying a functional on H with an element of H, and vice-versa, via

$$\begin{array}{rcl} H^* & \longleftrightarrow & H \\ f \otimes t & \longleftrightarrow & t \cdot f, \quad f \in H, t \in \mathbb{R}. \end{array}$$

Notice that this identification is norm-preserving, in the sense that

$$\|f \otimes t\| = \|t \cdot f\|,$$

where the $\|\cdot\|$ denotes the operator norm in H^* on the left-hand side, and the norm in H on the right-hand side. In the latter, we are going to use this identification extensively and abuse notation by writing $f \otimes t = t \cdot f$.

Notice that we can rewrite the empirical cross-covariance operator as $\Delta_n = \frac{1}{n}\sum_{i=1}^n Y_i \cdot X_i \in H$, where each $Y_i \cdot X_i$ are i.i.d. with mean $\Delta = 0$. Thus the Central Limit Theorem for Hilbert spaces yields

Theorem 5.2.2. *Under the null hypothesis $\psi = 0$, the random variable $\sqrt{n}\Delta_n$ converges in distribution to a Gaussian random variable G_Δ with mean zero and covariance operator $C = \sigma^2\Gamma$.*

Proof. We only need to show that $C = \sigma^2\Gamma$. The Central Limit Theorem tells us that C is the covariance operator of the random variable $X \otimes Y = Y \cdot X$. Direct computation yields

$$\begin{aligned} C &= \mathbb{E}\left[(Y \cdot X) \otimes (Y \cdot X)\right] \\ &= \mathbb{E}\left[Y^2 \cdot (X \otimes X)\right] \\ &= \mathbb{E}\left[Y^2\right] \mathbb{E}\left[X \otimes X\right] \\ &= \sigma^2\Gamma, \end{aligned}$$

where the third equality is justified because $Y = \langle 0, X \rangle + \varepsilon = \varepsilon$ is independent from X. □

In order to understand the distribution of $\sqrt{n}\Delta_n$ for a large n, we need a Karhunen-Loève expansion of the random variable G_Δ. We therefore make another assumption that allows us to use it:

A5: The random process X is continuous in the mean (see Definition 2.7.3),

We have the following decomposition from the Karhunen-Loève expansion.

Theorem 5.2.3. *If **A5** holds, then the random variable G_Δ admits the decomposition*

$$G_\Delta = \sigma \sum_{j=1}^{\infty} \sqrt{\lambda_j} \eta_j e_j, \tag{5.17}$$

the convergence is uniform in $L^2(\Omega, \mathcal{O}, \mathbb{P})$, and (η_j) are independent Gaussian random variables with mean zero and variance 1.

We can thus aproximate the distribution $\|\sqrt{n}\Delta_n\|^2$ by that of $\|G_\Delta\|^2$, for which we have the explicit formula

$$\|G_\Delta\|^2 = \sum_j \sigma^2 \lambda_j \eta_j^2, \tag{5.18}$$

in view of Theorems 5.2.2 and 5.2.3. However, the distribution of (5.18) depends on the unknowns σ and (λ_j). To "free" our random variable of these dependences, let us first write

$$G_\Delta = \sigma \Gamma^{1/2} Z = \sigma \Gamma^{1/2} Z = \sigma Z \Gamma^{1/2},$$

where $Z = \sum_j \eta_j e_j$, has a known distribution, and the last equality makes sense with our identification of H^* and H. Indeed,

$$Z\Gamma^{1/2} = \left(\sum_j \eta_j e_j \otimes 1 \right) \left(\sum_i \sqrt{\lambda_i} (e_i \otimes e_i) \right) = \sum_j \sqrt{\lambda_j} \eta_j \otimes 1,$$

which in turn equals to $\Gamma^{1/2} Z$ after the identification $\eta_j \otimes 1 \equiv \eta_j$.

Because neither G_Δ, σ nor Γ are known, we approximate G_Δ by $\sqrt{n}\Delta_n$, σ by some estimator $\hat{\sigma}$ and the covariance operator by its empirical counterpart

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes X_i = \sum_{i=1}^{n} \hat{\lambda}_i \hat{e}_i \otimes \hat{e}_i, \tag{5.19}$$

where the $\hat{\lambda}_i, \hat{e}_i$ are the eigenvalues and eigenfunctions of Γ_n, respectively. We now have to find Z that solves

$$\hat{\sigma} Z \Gamma_n^{1/2} = \sqrt{n} \Delta_n.$$

But as seen in Section 5.1, this leads to an inverse problem. One way of bypassing this problem is through spectral truncation. Let us define a truncated version of $\Gamma^{-1/2}$,

$$A_n = \sum_{j=1}^{p_n} = \lambda_j^{-1/2} e_j \otimes e_j,$$

where (p_n) is an increasing sequence satisfying $p_n < n$. The empirical counterpart of A_n is naturally defined by

$$\hat{A}_n = \sum_{j=1}^{p_n} = \hat{\lambda}_j^{-1/2} \hat{e}_j \otimes \hat{e}_j, \tag{5.20}$$

with $\hat{\lambda}_j$ and $\hat{e}_j$ defined in (5.19). Our test statistic will be based on

$$D_n = \frac{n}{\hat{\sigma}^2} \|\Delta_n \hat{A}_n\|^2. \tag{5.21}$$

We would assume that D_n follows approximatively a χ^2 distribution with p_n degrees of freedom if n is large. Furthermore, if we consider the test statistic

$$T_n = \frac{1}{\sqrt{p_n}} \left(\frac{n}{\hat{\sigma}^2} \|\Delta_n \hat{A}_n\|^2 - p_n \right) = \frac{1}{\sqrt{p_n}} (D_n - p_n), \tag{5.22}$$

then we have the following convergence Theorem:

Theorem 5.2.4 (Cardot et al. (2003)). *Let Δ_n, $\hat{A}_n$ be defined as in (5.21) and (5.20), and let*

$$T_n = \frac{1}{\sqrt{p_n}} (D_n - p_n).$$

Then under the null hypothesis $\psi = 0$, if the following assumptions hold:

1. $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_{p_n} > 0$, *a.s, (see (5.19))*
2. $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ *is bounded in probability,*
3. $\mathbb{E}\|X\|^4 < \infty$,
4. *The random process X is continuous in the mean,*

there exists a sequence (p_n) satisfying

$$\frac{n\lambda_{p_n}^2}{\left(\sum_{j=1}^{p_n} a_j\right)^2} \to \infty \text{ as } n \to \infty,$$

where $a_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}$ and

$$a_k = 2\sqrt{2} \max\left[(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}\right], \quad k \geq 2,$$

such that $T_n \xrightarrow{d} \mathcal{N}(0, 2)$, with T_n defined in (5.22).

Conclusions

In this section, we have presented two tests for functional data. First, we have given an asymptotic test for the mean of random curves, under the assumption of finite fourth moment, continuity in the mean, and injectivity of the covariance operator. We saw that the infinite dimensional character of the curves leads to an inverse problem when we try to make the test independent of the particular distribution of the data. Tikhonov regularization was therefore used as a solution to this. Then we presented an asympotitic test for the functional linear model with scalar responses. We saw how the linearity of the problem allowed us to simplify it to test whether or not the linear operator was trivial. The test was based on the empirical cross-covariance operator, and used Spectral truncation to deal with the inverse problem involved. The assumptions involved where stronger than for the previous test, in that the empirical eigenvalues need to be assumed almost surely different. Also, no technique for choosing the truncation level p_n was presented in the second paper.

6

An Application to DNA Shape Analysis

Now that we have seen some tests for functional data, let us give an application to testing for the mean of a set of DNA Minicircles. We would like to test if the mean of the DNA minicircles is indeed a circle.

A DNA Minicircle is a short strand of DNA (with around 150 base pairs) with its ends bound together to form a loop, by a reaction called *cyclization*. The data set we are going to test is a set of 94 DNA minicircles of 158 base-pair length. Among these, 63 curves are contain a so-called *TATA sequence*, and the remaining 31 curves contain the so-called *CAP sequence* instead. Our goal will be to test if the mean of TATA curves or CAP curves is circle.

These curves were reconstructed (Jacob et al. 2006) from electron micrographs obtained by Jan Bednar at the Laboratory of Ultrastructural Analysis of the University of Lausanne, Switzerland. They were then centered and scaled, in order to have center of mass zero and unit length, before being individually aligned using the coordinate system induced by their *moments of inertia tensor* (see Panaretos et al. (to appear) for details).

The data we are going to consider are the orthogonal projections of these curves onto the principal plane of their second and third principal axes of inertia, which we will call *PAI2*, *PAI3* (following the notation of Panaretos et al. (to appear)), and are shown in Figures 6.1 and 6.2.

Each curve was described by 401 points, and were fitted using a Fourier basis of length 151, with basis elements

$$1, \sqrt{2}\cos(2\pi t), \sqrt{2}\sin(2\pi t), \ldots, \sqrt{2}\cos(150\pi t), \sqrt{2}\sin(150\pi t), \quad t \in [0,1],$$

for each coordinate *PAI2, PAI3*.

Smoothing by penalization of the second derivative was considered a priori, but happened to be unnecessary. Indeed, an investigation of the Generalized Cross-Validation (GCV) criterion (Ramsay & Silverman 2005, Section 5.4.3) for choosing the smoothing parameter λ suggested a value of about $\lambda = 10^{-5.4}$, corresponding to a value of $4.78 \cdot 10^{-4}$ for the GCV criterion.

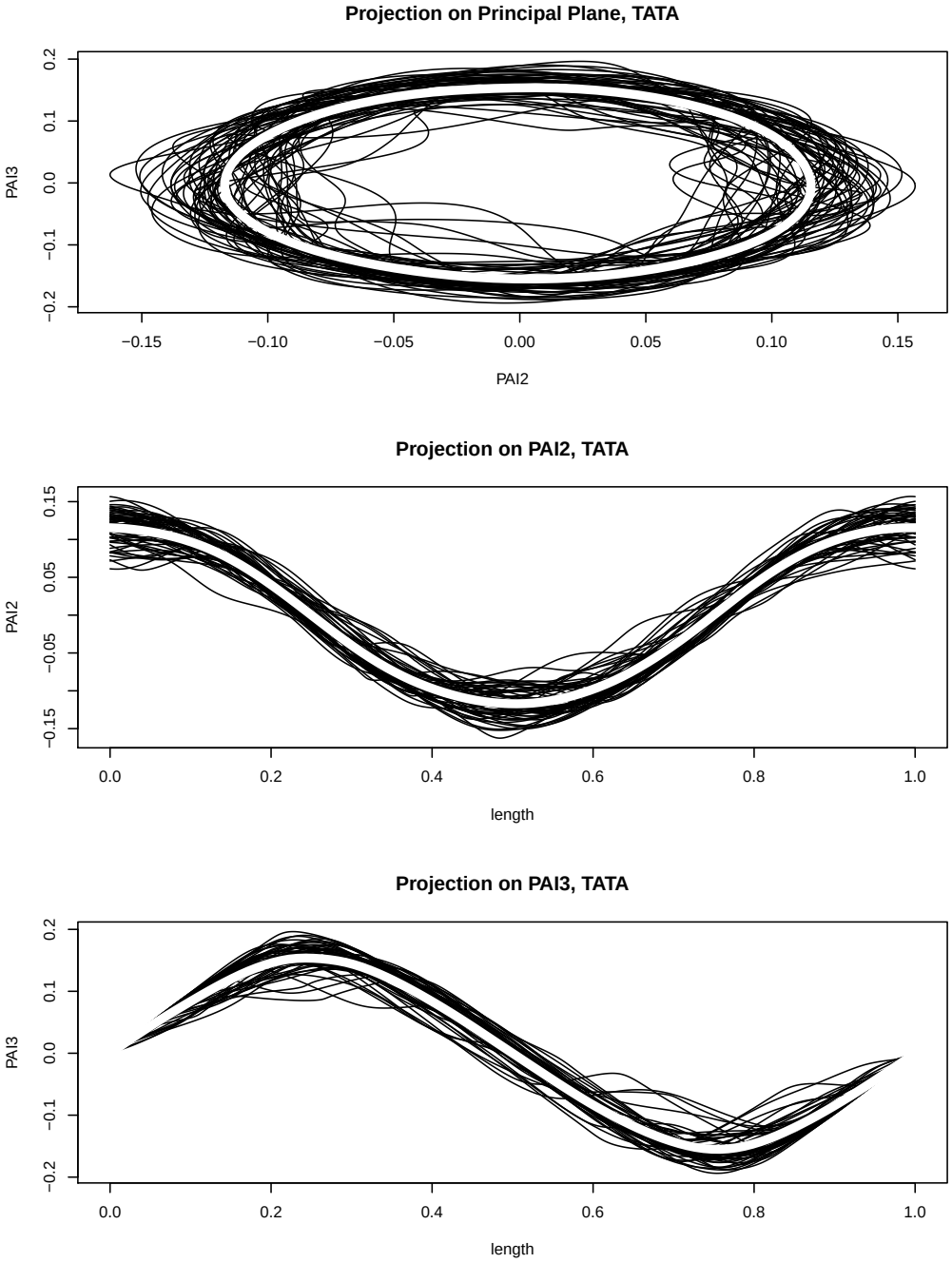

Figure 6.1: The TATA curves projected on the principal plane (top), and the coordinates on the PAI2 (middle) and PAI3 (bottom). The mean curve is represented in white.

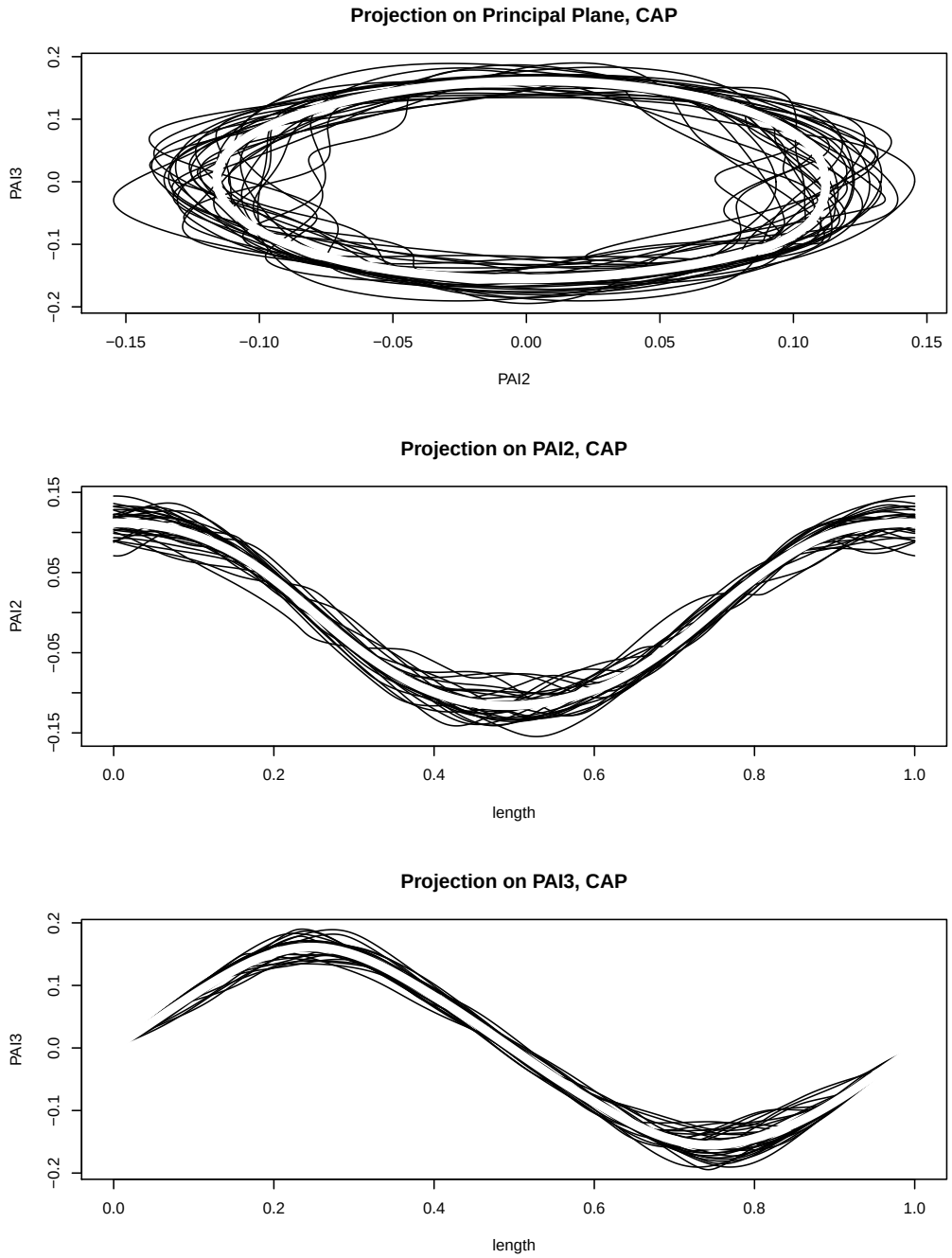

Figure 6.2: The CAP curves projected on the principal plane (top), and the coordinates on the PAI2 (middle) and PAI3 (bottom). The mean curve is represented in white.

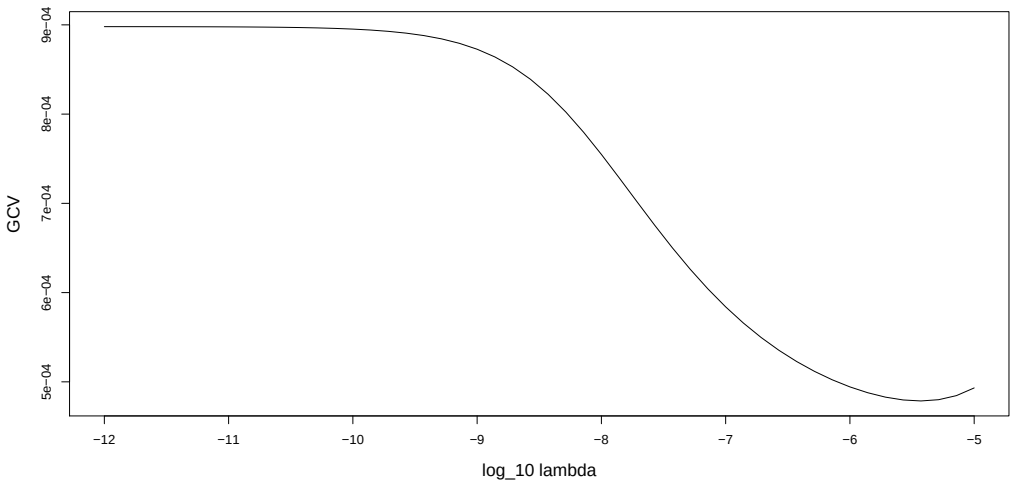

Figure 6.3: The values of the Generalized Cross-Validation criterion for choosing the smoothing parameter λ for fitting the DNA minicircle curves. The λ's are in a logarithmic scale.

However, the value of the GCV criterion seemed to be around $9 \cdot 10^{-4}$ for $\lambda \to 0$ (see Figure 6.3), which led to believe that smoothing was not necessary. Notice that this is coherent with the fact that the data we used was actually resampled from a B-spline representation of the DNA minicircles (Jacob et al. 2006), which apparently smoothed the data.

In order to test whether or not the mean curve is a circle, we first compute numerically the length of the sample mean curve, and choose the radius of the circle to have the that length. Notice that even though each curve has unit length, the length of the sample mean curve is less or equal to one. Numerical integration yields an approximate length of 0.854 for the TATA sample mean, and of 0.867 for the CAP sample mean. Figure 6.4 shows the plot of the sample mean and the circle of same length for TATA and CAP curves.

Let us summarize the test we will conduct. For each of the curve set $X_1, \ldots, X_n \overset{iid}{\sim} X$, (TATA or CAP), we would like to test the hypothesis

$$\begin{cases} H_0: & \mathbb{E}X = m_0, \text{ against} \\ H_a: & \mathbb{E}X \neq m_0, \end{cases}$$

where m_0 is a circle of given length, centered at the origin. In order to use the test statistic presented in Theorem 5.1.1, we need to make the following

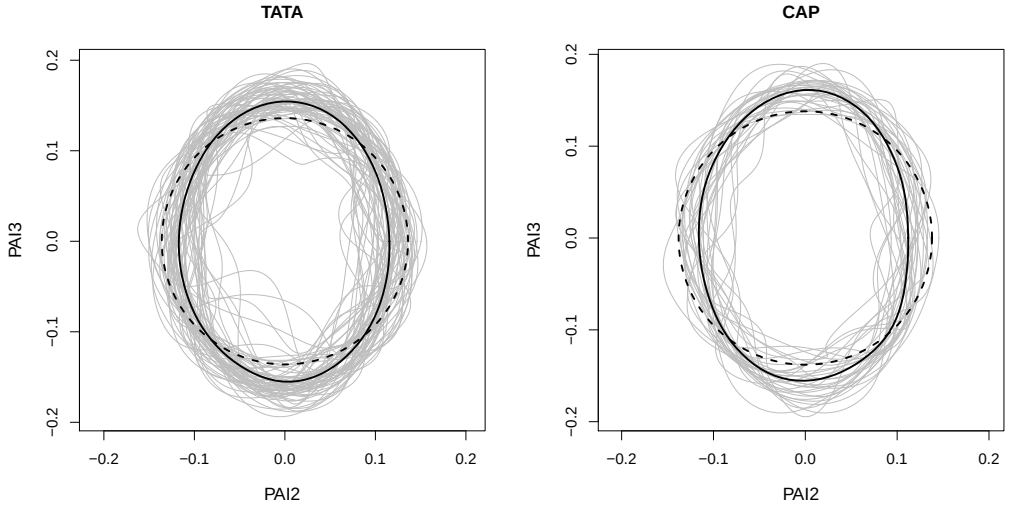

Figure 6.4: In black, the sample mean curves for TATA curves (left) and CAP curves (right), and the circles of corresponding length in dashed.

assumptions:

A1: $\mathbb{E}\|X\|^4 < \infty$,

A2: The covariance operator Γ of X is injective, that is, $\lambda_k > 0$ for all $k \geq 1$,

A3: The random process X is continuous in the mean (see Definition 2.7.3).

We can then apply Theorem 5.1.1 that tells us that for a large n and some α_n, k_n, the test statistic

$$\hat{T}_n = \frac{1}{\hat{d}_n} \left(\|(\Gamma_n + \alpha_n I)^{-1/2} S_n\|^2 - \hat{c}_n \right),$$

has approximatively a normal distribution with variance 3. The quantities $\hat{c}_n, \hat{d}_n$ are defined in (5.12), and depend on the eigenvalues of the empirical covariance operator, but also on α_n, k_n. These can be obtained via Theorem 5.1.2, which relies on some knowledge on the behaviour of the eigenvalues of the true covariance operator.

Hence we need to perform an eigenanalysis of the empirical covariance operator. For curves $X_1, \ldots, X_n$, the empirical covariance operator, under

the null hypothesis $\mathbb{E}X = m_0$, is

$$\Gamma_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m_0) \otimes (X_i - m_0).$$

Recall that we have sample curves $X_i : [0,1] \to \mathbb{R}^2$, and that we have used a Fourier basis for each of the PAI2 and PAI3 coordinates. Because PAI2 and PAI3 define orthogonal axes, and the Fourier basis is orthonormal, the joint basis

$$(\varphi_1, 0), (\varphi_2, 0), \ldots, (\varphi_{151}, 0), (0, \varphi_1), (0, \varphi_2), \ldots, (0, \varphi_{151}) \tag{6.1}$$

is orthonormal in $\mathcal{L}^2([0,1], \mathbb{R}^2)$, where

$$\varphi_1 = 1, \varphi_2 = \sqrt{2}\cos(2\pi t), \varphi_3 = \sqrt{2}\sin(2\pi t), \ldots, \varphi_{151} = \sqrt{2}\sin(150\pi t).$$

Therefore, if we denote by $\mathbf{c}_i, \mathbf{m}_0 \in \mathbb{R}^{302}$ the coefficient of the ith curve (respectively the mean under H_0) in the basis (6.1), the eigenvalues/eigenvectors of the empirical covariance operator satisfy the following equation (see Ramsay & Silverman (2005, p.163)):

$$\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbf{c}_i - \mathbf{m}_0)(\mathbf{c}_i - \mathbf{m}_0)^\top\right)\mathbf{v} = \rho\mathbf{v}, \quad \mathbf{v} \in \mathbb{R}^{302}, \rho \geq 0. \tag{6.2}$$

We know that this equation will yield at most n linearly independent eigenvectors with eigenvalue different from zero. This is verified if we look at a plot of the logarithm of the numerical eigenvalues of the empirical covariance operator for our two curve sets. The computations where carried out with R, and are shown in Figures 6.5 and 6.6. We thus retain only the 63 largest eigenvalues for TATA curves, and the 31 largest for CAP curves.

Now that we have the eigenvalues of the covariance operator, let us use Theorem 5.1.2 to choose the parameters α_n, k_n. Recall that the assumption of this Theorem was the bound

$$\exists M > 0, \varepsilon > 0 \text{ such that } \lambda_p \leq \frac{M}{p(\log p)^{1+\varepsilon}} \text{ for all } p > 1$$

on the true eigenvalues (λ_p) of the covariance operator. Notice that this expression is equivalent to

$$\exists M > 0, \varepsilon > 0\text{: } \log\lambda_p + \log p + (1+\varepsilon)\log(\log p) \leq \log(M) \text{ for all } p > 1.$$

A way of verifying if this bound holds for our data is thus to plot the function

$$\Xi(p) = \log\hat{\lambda}_p + \log p + (1+\varepsilon)\log(\log p), \quad p > 1 \tag{6.3}$$

	TATA	CAP
α_n	$8.19 \cdot 10^{-4}$	$2.09 \cdot 10^{-3}$
k_n	5	4
t_{obs}	38.62	23.65
p-value	$3.78 \cdot 10^{-110}$	$1.93 \cdot 10^{-42}$

Table 6.1: The values of α_n, k_n for the TATA and CAP curves, and the p-value of the testing if the mean curves are circles.

for different values of ε, and try to find an ε for which $\Xi(p)$ looks bounded. Trials with different values of ε yields the value $\varepsilon = 10$ for both TATA and CAP curves, as can be seen in Figure 6.7.

We can plug in the value of $\varepsilon = 10$ in Theorem 5.1.2, with the (arbitrary) choice of $\nu = 0.1$ to get the values of α_n, k_n. These values are shown in Table 6.1, together with the value of the test statistic, and their corresponding p-value:

$$\text{p-value}(t_{obs}) = \mathbb{P}\left(|Z| \geq t_{obs}\right), \quad Z \sim \mathcal{N}(0, 3).$$

We see that the p-values obtained are numerically zero, thus for both TATA and CAP curves, we reject the hypothesis that the mean curve is a circle.

Conclusions

In this last section, we gave an application of Section 5 to DNA shape analysis, to test if the mean shape of a DNA minicircle is a circle. We fitted the two samples of DNA minicircles to a Fourier basis of length 151, and it turned out that roughness penalization was not necessary, because the data was smooth enough already. Then we computed the length of the sample means numerically, and used it to find the radii of the corresponding circles. At this point, we had to perform an eigenanalysis of the empirical covariance operator, and use Theorem 5.1.2 to find the regularization parameters. In the end, the p-values of the test statistic (under the hypothesis that the mean DNA curve is a circle) where numerically zero for both DNA minicircle sets, and thus we rejected the hypothesis that the DNA minicircles are circles.

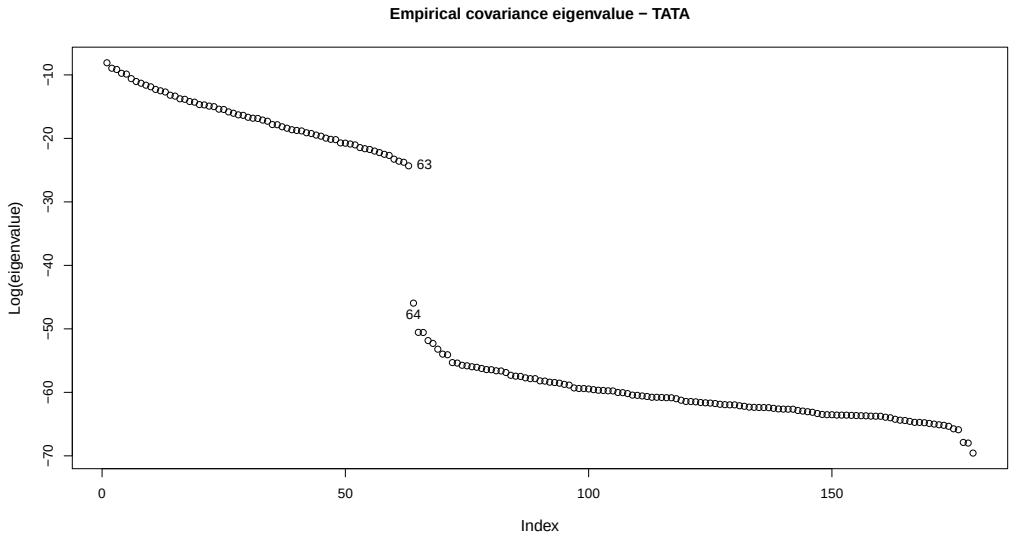

Figure 6.5: The logarithm of the eigenvalues of the empirical covariance operator for TATA curves. We see that starting from the 64^{th} value, the eigenvalues are numerically zero.

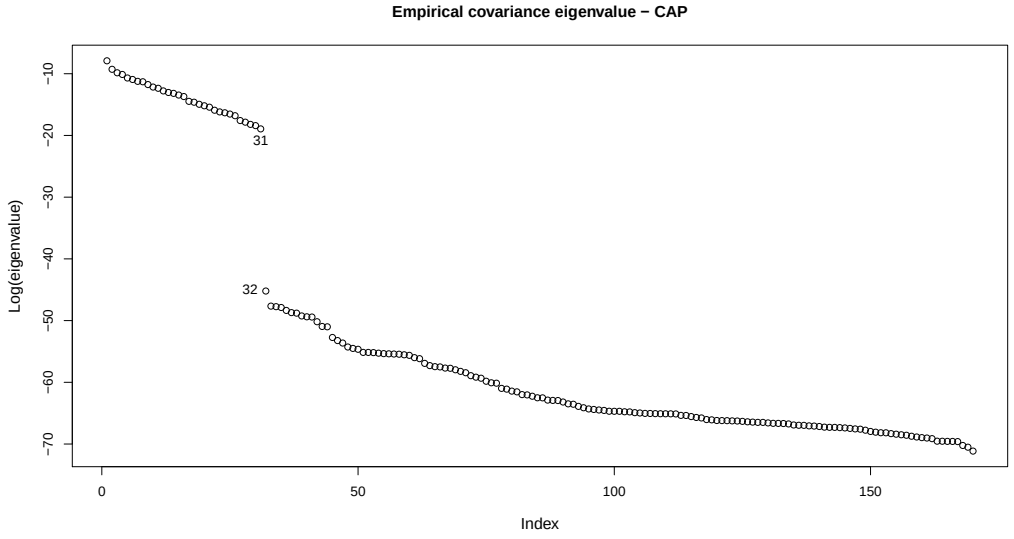

Figure 6.6: The logarithm of the eigenvalues of the empirical covariance operator for CAP curves. We see that starting from the 32^{nd} value, the eigenvalues are numerically zero.

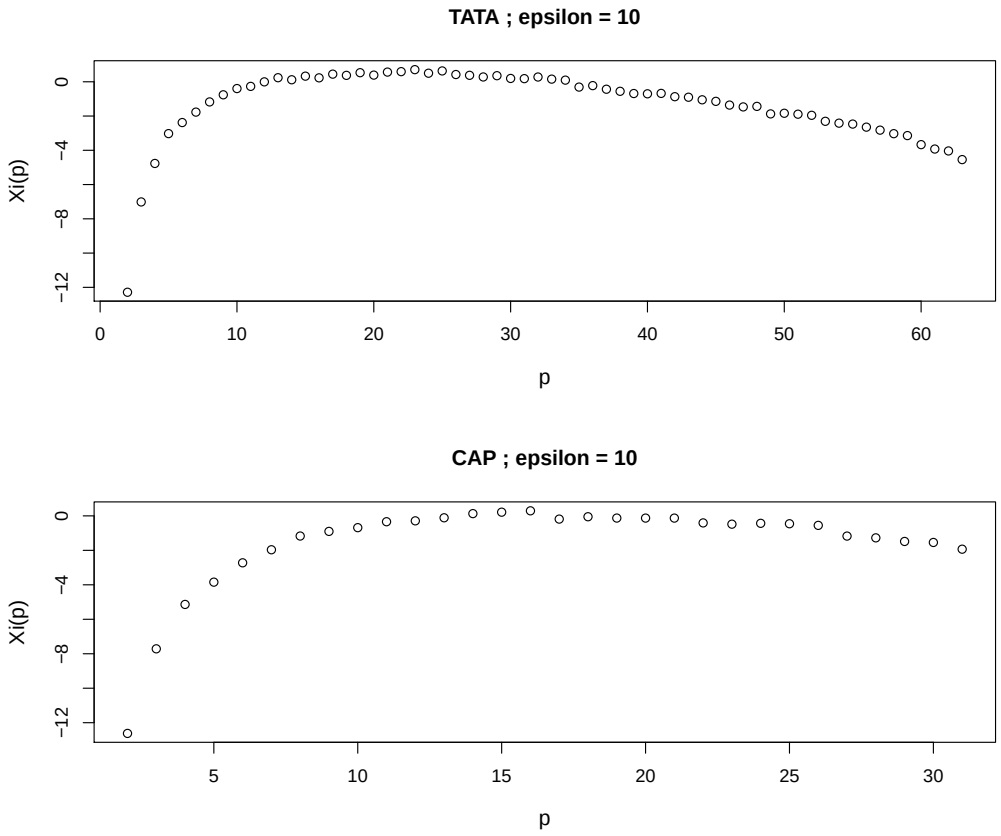

Figure 6.7: Plot of the function Ξ, defined in (6.3), for TATA (top) and CAP curves (bottom), with $\varepsilon = 10$.

Conclusion

I found the realization of this Master Project very interesting. I learned a lot about Hilbert spaces, the different types of operators, and the use of the tensor product. Then with probability in abstract spaces, I learned how to define formally random processes, and got some intuition about random variables in abstract spaces. The functional data analysis part was perhaps one of the most challenging for me, because I had to run through the subject quite quickly, and my main reference (Ramsay & Silverman 2005) lacked sometimes of details, in my view. It was also my first encounter with some basic notions such as PCA, and I am satisfied now that I understand them. I was also amazed how inverse problems occur often in various fields (I found out about this when talking with friends in engineering), and I'm glad that I've learned some basics about them.

I found the review of Mas (2007), Cardot et al. (2003) interesting, but I had the impression that some details where hidden and not properly dealt with, especially with the use of Karhunen-Loève expansion. Indeed, I had the impression that the necessary hypotheses (based on Grenander (1981)) where not clearly specified for its use. Eventually, the DNA shape analysis was a good exercise, because it obliged me to understand how to put into application some of the concepts I had theoretically exposed in the previous sections.

Now that I am at the end of this project, I have the funny impression that all the things I've learned were not so complicated, in contrast with the first impression I had. I guess this is how the learning process works!

Acknowledgements

I would like to thank Professor Victor M. Panaretos for all the time he took for transmitting his intuition of the subject, and his patience with reading my intermediary work. I would also like to thank my office collegue David Kraus, who helped me with all the day-to-day questions that I had. Eventually, many thanks to Anne-Fanny, who reminded me at the last minute (at the lake, the evening before the handout) to put mention "Mathematics" on the cover page!

References

Axler, S. (1997), *Linear Algebra Done Right*, 2 edn.

Bosq, D. (2000), *Linear Processes in Function Spaces*, Springer.

Cardot, H., Ferraty, F., Mas, A. & Sarda, P. (2003), 'Testing hypotheses in the functional linear model', *Scandinavian Journal of Statistics* **30**(1), 241–255.

De Boor, C. (2001), *A practical guide to splines. Rev. ed.*, Springer.

Debnath, L. & Mikusiński, P. (2005), *Introduction to Hilbert Spaces*, Elsevier Academic Press.

Dunford, N. & Schwartz, J. T. (1988), *Linear operators. Part I: General theory.* , John Wiley & Sons.

Gihman, I. & Skorohod, A. (1974), *The theory of stochastic processes. I. Translated from the Russian by S. Kotz.*, Springer-Verlag.

Grenander, U. (1981), *Abstract Inference*, John Wiley & Sons.

Hadamard, J. (1952), *Lectures on Cauchy's problem in linear partial differential equations.*, Dover Publications V.

Halmos, P. R. (1957), *Introduction to Hilbert space and the theory of spectral multiplicity. Reprint of the 2nd ed. 1957.*, AMS Chelsea Publishing.

Halmos, P. R. (1974*a*), *Finite-dimensional vector spaces. Reprint of the 2nd ed.*, Springer-Verlag.

Halmos, P. R. (1974*b*), *Measure theory. 2nd printing.*, Springer-Verlag.

Hastie, T. & Tibshirani, R. (1990), *Generalized additive models.*, Chapman and Hall.

Hoffmann, K. (2000), 'Stein estimation – a review', *Statistical Papers* **1**(41), 127–158.

Jacob, M., Blu, T., Vaillant, C., Maddocks, J. H. & Unser, M. (2006), '3d shape estimation of dna molecules from stereo cryo-electron micro-graphs using a projection-steerable snake', *IEEE Trans. Image Process* **15**, 214–227.

Kaipio, J. & Somersalo, E. (2005), *Statistical and Computational Inverse Problems*, Springer.

Kallenberg, O. (1997), *Foundations of modern probability.*, Springer.

Ledoux, M. & Talagrand, M. (1991), *Probability in Banach spaces. Isoperimetry and processes.*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge, 23. Berlin etc.: Springer-Verlag. xii, 480 p. .

Mas, A. (2007), 'Testing for the mean of random curves: A penalization approach', *Statistical Inference for Stochastic Processes* **10**(2), 147–163.

Munkres, J. R. (2000), *Topology. 2nd ed.*, Prentice Hall.

Panaretos, V. M., Kraus, D. & Maddocks, J. H. (to appear), 'Second-order comparison of gaussian random functions and the geometry of dna minicircles', *JASA Theory & Methods* .

Ramsay, J. & Silverman, B. (2005), *Functional Data Analysis*, 2 edn, Springer.

Reed, M. & Simon, B. (1972), *Methods of modern mathematical physics. I: Functional analysis.*, Academic Press.

Roussas, G. G. (1997), *A course in mathematical statistics. 2nd ed.*, Academic Press.

Rudin, W. (1991), *Functional Analysis*, 2 edn, Mc Graw Hill.

Seber, G. (1977), *Linear regression analysis.*, John Wiley & Sons.

Stein, C. (1956), 'Inadmissibility the usual estimator for the mean of a multivariate normal distribution', *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability* **1**, 197–206.

Tikhonov, A. N. & Arsenin, V. Y. (1977), *Solutions of Ill-posed Problems*, V. H. Winston & Sons (Scripta Technica).

Weidmann, J. (1980), *Linear operators in Hilbert spaces. Transl. by Joseph Szücs.*, Graduate Texts in Mathematics, Springer-Verlag.

Yao, F., Müller, H.-G. & Wang, J.-L. (2005), 'Functional data analysis for sparse longitudinal data.', *J. Am. Stat. Assoc.* **100**(470), 577–590.

Young, N. (1988), *An Introduction to Hilbert Space*, Cambridge University Press.

Zaanen, A. (1953), *Linear analysis.*, North-Holland Publishing Company.