

# FUNCTIONAL ANALYSIS FOR KERNEL-BASED LEARNING... AND STOCHASTIC REPRESENTATION THEOREMS

ALEXEY IZMAILOV

**ABSTRACT.** This paper metaphorically serves as a time tunnel from the level of an intermediate undergraduate math major to a sufficient level of functional analysis to understand the mathematical reason behind the inefficiencies of kernel-based learning methods in machine learning. To accomplish this, it contains a self-contained run-through of important results from Hilbert spaces and mathematical underpinnings of the important kernel trick. To illustrate these concepts, a plethora of worked problems are included. Many of these examples are either solved exercises or primary examples from textbooks and are cited accordingly. As an interesting application of these concepts, a small section regarding the Karhunen-Loève expansion theorem is included, although this requires a level of probability theory that is not developed in the main thrust of the paper.

The exposition makes no pretensions on novelty of content and if there is anything original within, it lies exclusively in the connective tissue between theorems. Solutions to the problems are mainly my own, although a few are adapted from other sources for results that I was unable to prove independently but were crucial for the flow of the exposition. These are attributed correspondingly. Give credit where credit is due.

In [a recent paper in the Notices](#), Princeton professor Weinan E briefly describes the distinction between "Courant-style" and "British-style" applied mathematics in the '80s. The former championed numerics and theorems: "its philosophy was that as long as the underlying PDEs and numerical algorithms are solid, much can be learned through computing". Some of its leaders were Von Neumann, Courant, Friedrichs, and Lax, who were of course also great *pure mathematicians* in addition to their distinguished careers as applied mathematicians. This is in stark contrast with the latter camp, which "championed physical insight and asymptotics". The present paper leans strongly in the direction of the Courant-style.

## CONTENTS

1. General Preliminaries	3
1.1. Inner Product Spaces	3
1.2. Linear Algebra Meets Fourier	4
1.3. Topological Vector Spaces	6
1.4. Metric Spaces	8
1.5. Hilbert Spaces	11
1.6. Isometries	13
1.7. Dual Spaces, Riesz Representation Theorem	15
2. Reproducing Kernel Hilbert Space	17
2.1. Definitions	17
2.2. Characterizing RKHS's: Continuous Evaluation Functionals	20
2.3. Characterizing RKHS's: Positive Type Functions (Kernels)	21
2.4. RKHS and Positive Type Functions	22
2.5. The Representer Theorem	22
3. Stochastics, or Something	23
3.1. Hilbert Space Generated by a Process:	24
3.2. Representation Theorems:	24
3.3. Karhunen-Loève Expansion Theorem	26
4. Acknowledgments	26
5. Bibliography	27
References	27

## 1. GENERAL PRELIMINARIES

This section focuses on developing and illustrating theory of Hilbert spaces and some related functional analysis from a background in proof-based linear algebra (think Axler) and basic analysis. After covering the important Riesz Representation Theorem, which provides a representation for linear functionals on an inner product space through the inner product, Reproducing Kernel Hilbert spaces are introduced and some basic, yet deep, characterizations are provided. The main aim is to give a sufficient background for understanding more advanced topics to be covered in Chapter 2, which discusses the kernel trick and the mathematical motivation for kernel-based learning methods, and Chapter 3, which discusses the Karhunen-Lo  ve theorem, a way to represent a stochastic process through orthogonal functions, particularly integral operators.

Something something about the one Courant applied math school thing from Princeton's Weinan E

**1.1. Inner Product Spaces.** Real and complex vector spaces can have an additional function defined on them, called the inner product, that satisfies the following defns.

**Definition 1.1.** Given a vector space  $V$  over  $F = \mathbb{R}$  or  $F = \mathbb{C}$ , an inner product on  $V$  is the function  $\langle \cdot, \cdot \rangle : V \times V \rightarrow F$  satisfying:

- 1) **(positive definiteness)** For all  $v \in V$ ,

$$\langle v, v \rangle \geq 0 \text{ and } \langle v, v \rangle = 0 \iff v = 0,$$

- 2) If the vector space is over the field  $\mathbb{C}$ , **(conjugate symmetry)**

$$\langle u, v \rangle = \overline{\langle v, u \rangle}$$

If the vector space is over the field  $\mathbb{R}$ , **(symmetry)**

$$\langle u, v \rangle = \langle v, u \rangle.$$

- 3) **(linearity in the first coordinate)** For all  $u, v \in V$  and  $r, s \in F$

$$\langle ru + sv, w \rangle = r\langle u, w \rangle + s\langle v, w \rangle.$$

An inner product induces a norm on the underlying vector space and is defined by

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

When  $F = \mathbb{R}$ , properties 2) and 3) imply that the inner product is linear in both coordinates, meaning that it is bilinear. But if  $F = \mathbb{C}$ , then the conjugate symmetry of the inner product results in

$$\langle w, ru + sv \rangle = \bar{r}\langle w, u \rangle + \bar{s}\langle w, v \rangle.$$

**Example 1.2.** The vector space  $\ell^2$  of all complex sequences  $(s_n)$  with

$$\sum |s_n|^2 < \infty \text{ with inner product } \langle (s_n), (t_n) \rangle = \sum_{n=0}^{\infty} s_n \bar{t}_n$$

of so called square summable sequences. The reason that the inner product converges follows from the trivial inequality:

$$0 \leq (|s_n| - |t_n|)^2 = |s_n|^2 - 2|s_n||t_n| + |t_n|^2,$$

so

$$2|s_n t_n| \leq |s_n|^2 + |t_n|^2,$$

implying that  $(s_n t_n) \in \ell^2$ .

**Example 1.3** (Roman Chap. 9, Exer 3). In the triangle inequality,

$$\|u + v\| \leq \|u\| + \|v\|,$$

equality holds if and only if one of  $u$  and  $v$  is a non-negative scalar multiple of the other.

Without loss of generality, suppose  $u = av$  for some  $a \geq 0$ . Then

$$\|u + v\| = (a + 1)\|v\| = \|u\| + \|v\|.$$

To show the converse, observe that

$$\|u\|^2 + \langle u, v \rangle + \langle v, u \rangle + \|v\|^2 = \|u\|^2 + 2\|u\|\|v\| + \|v\|^2$$

while

$$|\langle u, v \rangle| \leq \|u\|\|v\| = \mathcal{R}(\langle u, v \rangle) \leq |u, v|.$$

Equality in Cauchy-Schwarz holds only if one of  $u$  or  $v$  is a scalar multiple of the other. Without loss of generality, let  $u = av$  for some  $a \in F$ . Then

$$\|u\|\|v\| = \langle u, v \rangle = \langle av, u \rangle = a\|v\|^2,$$

meaning that  $a$  must be non-negative.

**1.2. Linear Algebra Meets Fourier.** An important problem in linear algebra is the determination of the basis for a particular space. As was covered in a first linear algebra class, orthonormal bases have a great practical advantage over arbitrary basis. Fourier expansions provide a mechanism by which an explicit expression can be found for any vector. A nonempty set  $\mathcal{O} = \{u_i | i \in K\}$  of vectors in an inner product space is said to be an orthogonal set if  $\langle u_i, u_j \rangle = \delta_{ij}$  for all  $i, j \in K$ . Given that  $\mathcal{O}$  is an orthonormal subset of an inner product space  $V$  and  $S = \langle \mathcal{O} \rangle$ , the Fourier expansion with respect to  $\mathcal{O}$  of  $v \in V$  is

$$\hat{v} = \langle v, u_1 \rangle u_1 + \cdots + \langle v, u_k \rangle u_k.$$

This vector  $\hat{v}$  can be characterized as

- 1) the unique vector  $s \in S$  for which  $(V - s) \perp S$ .
- 2) the best approximation to  $v$  from within  $S$ , that is,  $\hat{v}$  is the unique vector  $s \in S$  that is closest to  $v$  in the sense that

$$\|v - \hat{v}\| < \|v - s\|$$

for all  $s \in S \setminus \{\hat{v}\}$ .

- 3) (Bessel's Inequality) for all  $v \in V$ ,

$$\|\hat{v}\| \leq \|v\|.$$

**Example 1.4.** (Roman Chap. 9, Exer. 12) Prove Bessel's inequality for inner product spaces.

Bessel's inequality is really a statement about the coefficients of an element  $v \in \mathcal{H}$  with respect to an orthonormal sequence (though not a basis! that is the topic of Parseval's

identity, that will be covered shortly). The proof is rather direct. By the trivial inequality,

$$\begin{aligned}
0 &\leq \left\| v - \sum_{i=1}^k \langle x, u_i \rangle u_i \right\|^2 \\
&= \left\langle v - \sum_{i=1}^k \langle x, u_i \rangle u_i, v - \sum_{i=1}^k \langle x, u_i \rangle u_i \right\rangle \\
&= \langle v, v \rangle - \left\langle v, \sum_{i=1}^k \langle x, u_i \rangle \right\rangle - \left\langle \sum_{i=1}^k \langle x, u_i \rangle, v \right\rangle + \left\langle \sum_{i=1}^k \langle x, u_i \rangle, \sum_{i=1}^k \langle x, u_i \rangle \right\rangle \\
&= \|x\|^2 - 2 \sum_{i=1}^k |\langle v, u_i \rangle|^2 + \sum_{i=1}^k |\langle v, u_i \rangle|^2 \\
&= \|x\|^2 - \sum_{i=1}^k |\langle v, u_i \rangle|^2.
\end{aligned}$$

From this, it follows trivially that

$$\sum_{i=1}^k |\langle v, u_i \rangle|^2 < \|x\|^2,$$

showing the desired result.

The name "Fourier expansion" is not a misnomer: this is really a generalization of the usual Fourier expansions that are the backbone of any PDE course. This is a nice illustration of the beautiful possibilities lying on the interfaces of seemingly disparate fields of study.

**Example 1.5** (Fourier Series). Given an arbitrary function  $f(\theta)$  defined on  $[-\pi, \pi]$ , we want to represent it as a sum of sines and cosines:

$$f(\theta) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\theta + \sum_{n=1}^{\infty} a_n \sin n\theta.$$

In other words, we are analyzing the periodic extension for a function. This is a familiar phenomenon. In calculus, we consider the expansion of a function in terms of polynomials, with the only compromise being that an infinite number of them is necessary. These classical approaches are known as decomposition problems. Another intuitive way to think about this particular exmp is that we are trying to span the space of functions, but of course this space is not finite, so it is impossible to express every function in it as a linear combination of finitely many basis elements. Consider the inner product on the space of these functions

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(\theta)g(\theta)d\theta.$$

Suppose we want to find an approximation of the function in an inner product subspace, that is,

$$f(\theta) \approx a_0 + \sum_{n=1}^N a_n \cos n\theta + \sum_{n=1}^N a_n \sin n\theta.$$

As can be confirmed, the terms in the above expansion are orthogonal, which is a classical Fourier series exercise. That is, computing  $\langle 1, \cos n\theta \rangle$  and  $\langle \cos n\theta, \cos m\theta \rangle$  reveals that they are orthogonal. Moreover, one of the most important parts of Fourier series is the determination of the coefficients of the above approximation. Observe that

$$a_0 = \frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)d\theta,$$

$$a_n = \frac{\langle f, \cos n\theta \rangle}{\langle \cos n\theta, \cos n\theta \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \cos n\theta d\theta$$

$$b_n = \frac{\langle f, \sin n\theta \rangle}{\langle \sin n\theta, \sin n\theta \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \sin n\theta d\theta.$$

This is exactly the same as what is covered in a first Fourier series lecture, but the derivations of these formulas are much more direct and motivated!

The previous discussion concerned itself primarily with *approximations* of a function through a subset of the orthonormal vectors. However, suppose we actually took an subset  $O$  of orthonormal vectors from an inner product space that is an orthonormal basis for  $V$ . Then we would have that every vector is equal to its Fourier expansion: for every  $v \in V$ ,  $\hat{v} = v$ . Moreover, Bessel's inequality becomes an equality as  $\|\hat{v}\| = \|v\|$ , and we also obtain Parseval's equality: for all  $v, w \in V$ ,

$$\langle v, w \rangle = [\hat{v}]_O \cdot [\hat{w}]_O$$

where

$$[\hat{v}]_O \cdot [\hat{w}]_O = \langle v, u_1 \rangle \overline{\langle w, u_1 \rangle} + \cdots + \langle v, u_k \rangle \overline{\langle w, u_k \rangle}.$$

**1.3. Topological Vector Spaces.** The study of topological vector spaces dates back to Alexander Grothendieck who was engaged in a completely general approach to the study of these spaces and collected some of the deepest results in his thesis. After his dissertation, he exclaimed that "there is nothing more to do, the subject is dead!" Although Grothendieck's insight into mathematics was second only to a handful of the most illustrious mathematicians, the theory of topological vector spaces is far from dead. Many aspects are still unknown and the theory lively interacts with several open problems, most notably highlighted by research done in recent years by the Russian school of topology led by Archangel'skii into the theory of metrizable spaces and topological function spaces. For a high-level survey, see [this paper](#) by Shakhmatov.

To provide the intuitive underpinnings of Hilbert spaces, it is first necessary to introduce some basic notions from point-set topology. As with other structures in mathematics, topology arises from the definition of a series of properties of a set and the relations among its subsets. A topology on a set  $X$  is usually defined by specifying its open subsets, but when dealing with topological vector spaces, it is more convenient to specify what the neighborhoods of each point are.

**Definition 1.6.** A topology  $\tau$  on  $X$  is a family of subsets of  $X$  which satisfies all of the following:

- (i). Both  $\emptyset$  and all of  $X$  are in  $\tau$ ,
- (ii).  $\tau$  is closed under finite intersections,
- (iii).  $\tau$  is closed under arbitrary unions.

The pair  $(X, \tau)$  is called a topological space.

Sets  $O \in \tau$  are called **open sets** while their complements  $C = X \setminus O$  are called the **closed sets**. A given subset of  $X$  may be neither closed nor open, either closed or open, or both. As in other areas of math, it is of great importance to consider a set that in some sense *generates* the structure of interest.

**Definition 1.7.** Given that  $(X, \tau)$  is a topological space,

- (i). A subfamily  $\beta$  of  $\tau$  is called a **basis** if every open set can be written as a possibly empty union of sets in  $\beta$ ,

- (ii). A subfamily  $\mathcal{X}$  of  $\tau$  is a subbasis if the finite intersections of its sets form a basis. In other words, every open set can be written as a union of finite intersections of sets in  $\mathcal{X}$ .

With this in mind, a topology  $\tau$  on  $X$  can be completely determined by a basis or a subbasis.

**Example 1.8** (Standard Topology). The family

$$\beta = \{(a, b) : a, b \in \mathbb{Q}, a < b\}$$

is a basis of the Euclidean/standard topology on  $\mathbb{R}$ .

**Example 1.9** (Semi-infinite intervals). Sometimes a concept is better illustrated by a non-example and the reasons that it fails the criteria to be an example properly. Consider the collection  $\mathcal{S}$  of all semi-infinite intervals of the real line in the forms  $(-\infty, a)$  and  $(a, \infty)$ , where  $a \in \mathbb{R}$ .  $\mathcal{S}$  is not a basis for any topology on  $\mathbb{R}$ . Suppose to the contrary that  $\mathcal{S}$  is a basis for a topology  $\tau$ . Then by definition, the intervals  $(-\infty, 1)$  and  $(0, \infty)$  would be in  $\tau$ , so their intersection, which is  $(0, 1)$  would also be in  $\tau$  by axiom (ii). But evidently  $(0, 1)$  cannot be written as a union of elements in  $\mathcal{S}$ , meaning it is not a basis. On the other hand,  $\mathcal{S}$  is a subbasis of the Euclidean topology on  $\mathbb{R}$  because every interval in that basis can be written as a finite intersection of elements of  $\mathcal{S}$ .

But these are topological *spaces*. The *vector space* part comes into play when we consider the set  $X$  as a vector space. Thus, the pair  $(V, \tau)$  where  $V$  is a real vector space and  $\tau$  is a topology on  $V$  forms a **topological vector space** where the operations are defined as

$$\mathcal{A} : V \times V \rightarrow V, \mathcal{A}(v, w) = v + w,$$

$$\mathcal{M} : \mathbb{R} \times V \rightarrow V, \mathcal{M}(r, w) = rw.$$

**Example 1.10** (Standard Topology on  $\mathbb{R}^n$ ). The vector space  $\mathbb{R}^n$  is a topological vector space under the standard topology, which is the topology for which the set of open rectangles

$$\beta = \{I_1 \times \cdots \times I_n \mid I_i \text{ is open in } \mathbb{R}\}.$$

is a basis. That is, a subset of  $\mathbb{R}^n$  is open if and only if it is the union of open rectangles. The standard topology on  $\mathbb{R}^n$  has the addition function

$$\mathcal{A} : \mathbb{R}^n \times \mathbb{R}^n; \mathcal{A}(v, w) \rightarrow v + w,$$

and scalar multiplication function

$$\mathcal{M} : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n; \mathcal{M}(r, v) \rightarrow rv.$$

Both of these functions are continuous on  $\mathbb{R}^n$ . To show this, consider that if

$$(u_1, \dots, u_n) + (v_1, \dots, v_n) \subset (a_1, b_1) \times \cdots \times (a_n, b_n) \in \beta,$$

then for all  $i$ ,  $u_i + v_i \in (a_i, b_i)$ , so there is a  $\varepsilon > 0$  for which

$$(u_i - \varepsilon, u_i + \varepsilon) + (v_i - \varepsilon, v_i + \varepsilon) \subset (a_i, b_i).$$

Therefore,

$$(u_1, \dots, u_n) \in I = (u_1 - \varepsilon, u_1 + \varepsilon) \times \cdots \times (u_n - \varepsilon, u_n + \varepsilon) \subset \beta$$

and analogously,

$$(v_1, \dots, v_n) \in J = (v_1 - \varepsilon, v_1 + \varepsilon) \times \cdots \times (v_n - \varepsilon, v_n + \varepsilon) \subset \beta.$$

Consequently,

$$(u_1, \dots, u_n) + (v_1, \dots, v_n) \in \mathcal{A}(I, J) \subset (a_1, b_1) \times \cdots \times (a_n, b_n).$$

To show that  $\mathcal{M}$  is continuous, if  $v \in \mathbb{R}^n$  and  $r \in \mathbb{R}$  and  $R$  is a neighborhood of  $rv$ , then for some  $\varepsilon > 0$  and some neighborhood  $W$  of  $v$ ,  $\beta W \subset V$  whenever  $|\beta - r| < \varepsilon$ .

**Example 1.11** (Roman Chap., 2, Exer. 25). Any linear functional  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a continuous map.

Let  $M = \max(|f(e_1)|, \dots, |f(e_n)|)$  (where  $e_i$ 's form the canonical basis),  $x \in \mathbb{R}$  and  $\varepsilon > 0$ . For all  $y \in \mathbb{R}$  such that

$$\|x - y\| < \frac{\varepsilon}{Mn},$$

it is true that

$$|f(x) - f(y)| = |f(x - y)| = \|x - y\| \left\| f \left( \frac{x - y}{\|x - y\|} \right) \right\|.$$

Then write

$$\frac{x - y}{\|x - y\|} = a_1 e_1 + \dots + a_n e_n,$$

where  $a_1^2 + \dots + a_n^2 = 1$  because it is a unit vector. Then

$$|f(x) - f(y)| = \|x - y\| |a_1 f(e_1) + \dots + a_n f(e_n)| = M \|x - y\| (|a_1| + \dots + |a_n|).$$

But by the condition that  $a_1^2 + \dots + a_n^2 = 1$ , this is equivalent to

$$MN \|x - y\| \max(|a_1|, \dots, |a_n|) < \varepsilon$$

because each  $|a_i| \leq 1$  for all  $i \in [n]$ .

The continuity of addition and multiplication extends to all topological vector spaces. But this is for a more abstract reason. Namely that if  $V$  is a real vector space of dimension  $n$  and we fix an ordered basis  $\beta = \{v_1, \dots, v_n\}$  for  $V$ , then there is precisely one topology  $\tau$  on  $V$  for which  $(V, \tau)$  is a topological vector space and all linear functionals are continuous. This topology is called the **natural topology**.

**1.4. Metric Spaces.** While much of linear algebra focuses itself on the basic properties of real and complex inner product spaces, few of those properties rely on whether the space under consideration is finite or infinite dimensional. However, the metric induced by an inner product raises a host of convergence issues. To study the convergence properties of real and complex inner product spaces, the notion of a metric space is discussed. It is important to remember that a metric space is **not an algebraic structure** and is simply designed as a model of abstract properties of distance.

**Definition 1.12.** A metric space is a pair  $(M, d)$  where  $M$  is a nonempty set and  $d : M \times M \rightarrow \mathbb{R}$  is a metric on  $M$  with the following properties:

- 1) (positive definiteness) for all  $x, y \in M$ ,

$$d(x, y) \geq 0$$

and  $d(x, y) = 0$  if and only if  $x = y$ .

- 2) (symmetry) for all  $x, y \in M$ ,

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

- 3) (triangle inequality) for all  $x, y, z \in M$ ,

$$d(x, y) \leq d(x, z) + d(z, y).$$

**Example 1.13** (Euclidean metric on  $\mathbb{R}^n$ ). The most obvious example of a metric space is the set  $\mathbb{R}^n$  under the metric defined for  $x = (x_1, x_2, \dots, x_n)$  and  $y = (y_1, \dots, y_n)$  by

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}.$$



**Example 1.14** (Manhattan metric on  $\mathbb{R}^n$ ).  $\mathbb{R}^n$  is also a metric space under

$$d(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n|.$$

Of course the space under this metric is different from the space under the Euclidean metric.

**Example 1.15** (Roman Chap. 12, Exer. 1, 2). Here are some basic properties of the metric.

- (1) (generalized triangle inequality) If  $x_1, x_2, \dots, x_n \in M$ , show that

$$d(x_1, x_n) \leq d(x_1, x_2) + d(x_2, x_3) + \cdots + d(x_{n-1}, x_n).$$

By induction on  $n$ , the base case is  $n = 2$  from the usual triangle inequality. Then for the inductive step, observe that

$$d(x_1, x_n) = d(x_1, x_{n-1}) + d(x_{n-1}, x_n) \leq d(x_1, x_2) + \cdots + d(x_{n-1}, x_n)$$

by an application of the inductive hypothesis.

- (2) If  $x, y, a, b \in M$ , then

$$|d(x, y) - d(a, b)| \leq d(x, a) + d(y, b).$$

By two applications of the triangle inequality,

$$d(x, y) \leq d(x, a) + d(a, b) + d(b, y)$$

$$d(a, b) \leq d(a, x) + d(x, y) + d(y, b),$$

so by the symmetry property of the metric, the desired inequality follows.

- (3) If  $x, y, z \in M$ , then

$$|d(x, z) - d(y, z)| \leq d(x, y).$$

Similarly as to part (2), the triangle inequality gives that

$$d(x, z) \leq d(x, y) + d(y, z)$$

$$d(y, z) \leq d(x, y) + d(x, z).$$

**Example 1.16** (Roman Chap. 12, Exer. 4 - Hamming Distance Function). An important function from the theory of error-correcting codes is the Hamming distance function. Let  $M = \{0, 1\}^n$  be the set of all binary  $n$ -tuples. Define the function  $h : S \times S \rightarrow \mathbb{R}$  by letting  $h(x, y)$  be the number of positions in which  $x$  and  $y$  differ (in other words, counting the number of 1's in the XOR of the two strings). This is a metric on the space of binary strings of length  $n$ .

Trivially  $h(x, y)$  is positive definite and satisfies the symmetry criteria by facts from counting. The interesting part is to show that  $h(x, y)$  satisfies the triangle inequality. Note that we can identify  $d(x, y) = |D(x, y)|$  where

$$D(x, y) = \{k \mid x_k \neq y_k\}$$

is the set of indices  $k$  for which  $x, y \in M$  differ in the  $k$ th position. Recall that for sets  $A, B$ , the symmetric difference  $A \oplus B$  consists of the elements in the union of  $A$  and  $B$  that are not in the intersection of  $A$  and  $B$ . By the binary nature of these bit strings (pun(s) intended?), it must be that if  $x$  and  $z$  differ at some index and  $y$  and  $z$  differ at the same index, that  $x$  and  $y$  cannot differ at that index. Therefore, we have

$$D(x, y) \subset D(x, z) \oplus D(y, z) \subset D(x, y) \cup D(y, z),$$

from which

$$|D(x, y)| \leq |D(x, z) \oplus D(y, z)| \leq |D(x, y) \cup D(y, z)| = |D(x, z)| + |D(y, z)|,$$

and the conclusion follows since  $d(x, y) = |D(x, y)|$  and likewise for the other terms.

**Example 1.17** (Sequence spaces, Hölder's and Minkowski's Inequalities). Many important sequences spaces are also metric spaces. For the sake of brevity, let bold-face letters denote sequences, such as  $\mathbf{x} = (x_n)$  and  $\mathbf{y} = (y_n)$ .

- 1) The set  $\ell_{\mathbb{R}}^{\infty}$  of all bounded sequences of real numbers is a metric space under the metric defined by

$$d(\mathbf{x}, \mathbf{y}) = \sup_n |x_n - y_n|.$$

The set  $\ell_{\mathbb{C}}^{\infty}$  of all bounded complex sequences under the same metric is also a metric space. This is the reason that both of these spaces are identified by  $\ell^{\infty}$ .

- 2) For  $p \geq 1$ , let  $\ell^p$  be the set of all sequence  $\mathbf{x} = (x_n)$  of real or complex numbers for which

$$\sum_{n=1}^{\infty} |x_n|^p < \infty.$$

Then the **p-norm** on this space is defined as

$$\|\mathbf{x}\|_p = \left( \sum_{n=1}^{\infty} |x_n|^p \right)^{1/p}.$$

Then  $\ell^p$  is a metric space under the metric

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p = \left( \sum_{n=1}^{\infty} |x_n - y_n|^p \right)^{1/p}.$$

$\ell^p$  space is governed by the following two inequalities that give meaning to the metric on products and sums of sequences: Hölder's and Minkowski's inequalities, respectively.

**Hölder's Inequality:** Let  $p, q \geq 1$  and  $p + q = pq$ . If  $\mathbf{x} \in \ell^p$  and  $\mathbf{y} \in \ell^q$ , then the product sequence  $\mathbf{xy} = (x_n y_n) \in \ell^1$  and

$$\|\mathbf{xy}\|_1 \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q.$$

In other words, we have the general inequality

$$\sum_{n=1}^{\infty} |x_n y_n| \leq \left( \sum_{n=1}^{\infty} |x_n|^p \right)^{1/p} \left( \sum_{n=1}^{\infty} |y_n|^q \right)^{1/q}$$

for which Cauchy-Schwarz is the  $p = q = 2$  special case.

**Minkowski's inequality:** For  $p \geq 1$ , if  $\mathbf{x}, \mathbf{y} \in \ell^p$ , then the sum  $\mathbf{x} + \mathbf{y} = (x_n + y_n)$  is in  $\ell^p$  and

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p,$$

or equivalently,

$$\left( \sum_{n=1}^{\infty} |x_n + y_n|^p \right)^{1/p} \leq \left( \sum_{n=1}^{\infty} |x_n|^p \right)^{1/p} + \left( \sum_{n=1}^{\infty} |y_n|^p \right)^{1/p}.$$

One of the central objects in any discussion regarding metric spaces is the *complete metric space*, which provides further insight into concepts from regular analysis. Recall that a sequence  $(x_n)$  in a metric space  $M$  is a Cauchy sequence if for any  $\varepsilon > 0$ , there exists an  $N > 0$  for which

$$n, m > N \implies d(x_n, x_m) < \varepsilon.$$

Any convergent sequence is a Cauchy sequence. When the converse holds, the space is said to be **complete**.

**Definition 1.18.** Given a metric space  $M$ ,

- 1)  $M$  is complete if every Cauchy sequence in  $M$  converges in  $M$ ,
- 2) A subspace  $S$  of  $M$  is complete if it is complete as a metric space. Thus,  $S$  is complete if every Cauchy sequence  $(s_n)$  in  $S$  converges to an element in  $S$ .

**Theorem 1.19.** Given a metric space  $M$ ,

- 1) Any complete subspace of  $M$  is closed,
- 2) If  $M$  is complete, then a subspace  $S$  of  $M$  is complete if and only if it is closed.

*Proof.* Suppose  $S$  is a complete subspace of  $M$  and  $(x_n)$  a sequence in  $S$  which converges  $(x_n) \rightarrow x \in M$ . So  $(x_n)$  is a Cauchy sequence and by the completeness of  $S$ , the sequence converges to an element of  $S$ . And since limits of sequences are unique,  $x \in S$  and hence  $S$  is closed.

To show the other part, suppose that  $S$  is complete. Then by the previous discussion, it is closed. Conversely, suppose that  $S$  is closed and let  $(x_n)$  be a Cauchy sequence in  $S$ . Because  $(x_n)$  is also a Cauchy sequence in  $M$ , which is complete, it converges to some  $x \in M$ , but since  $S$  is closed,  $(x_n) \rightarrow x \in S$ , making the subspace complete.  $\square$

Any standard analysis book has a plethora of examples of complete metric spaces. Some of the more basic ones include: the Euclidean space  $\mathbb{R}^n$ , the unitary space  $\mathbb{C}^n$ , and  $(C[a, b], d)$ , the space of all real-valued continuous functions on  $[a, b]$  under the sup metric

$$d(f, g) = \sup_{x \in [a, b]} |f(x) - g(x)|.$$

A metric space  $(X, d)$  is a topological space since the space can be equipped with the topology induced by the metric  $d$ . Namely, let the topology have a basis consisting of the collection of neighborhoods  $\{B_\varepsilon(x) : \varepsilon \in \mathbb{R}^+\}$  where  $B_\varepsilon = \{y \in X : d(y, x) < \varepsilon\}$  for any  $x \in X$ . The additional structure that a topology obtains from being *over* some algebraic structure endows it with improved structural properties of its own.

**1.5. Hilbert Spaces.** Hilbert spaces are an important example of topological vector spaces and are the foundation of the discussion in this DRP paper. Given a non-empty set  $S$ , let  $\mathcal{H}$  be a vector space of functions defined on  $S$  in which their values are taken in the field  $\mathbb{C}$ .  $\mathcal{H}$  is endowed with the structure of a Hilbert space when defined by an inner product space  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ :

$$\mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C},$$

$$(\varphi, \psi) \mapsto \langle \varphi, \psi \rangle_{\mathcal{H}}.$$

When the space is understood to be  $\mathcal{H}$  in context, we omit the subscript. The final part of what makes a space a *Hilbert space* is that the metric resulting from the inner product endows  $\mathcal{H}$  with properties of a complete metric space – every partial sum of norms of any sequence of vectors converges to an element of  $\mathcal{H}$ .

**Example 1.20** ( $\ell^2$  Sequence Space). Let  $S = \{1, 2, \dots\} = \mathbb{N}^*$  and  $\mathcal{H} = \ell^2(\mathbb{C})$  be the set of complex sequences  $(x_i)_{i \in \mathbb{N}^*}$  such that  $\sum_{i \in \mathbb{N}^*} |x_i|^2 < \infty$ .  $\mathcal{H}$  is a Hilbert space with the inner product

$$\langle (x_i), (y_i) \rangle_{\ell^2(\mathbb{C})} = \sum_{i \in \mathbb{N}^*} x_i \bar{y}_i$$

**Example 1.21** (Berlinet Chap. 1, Exer. 1 - Probability Distribution and Polynomials). Let  $\mathbb{P}_r$  be the space of polynomials of degree at most  $r$  where  $r$  is a nonnegative integer. Let

$K_0$  be a probability distribution on  $\mathbb{R}$ , a nonnegative integrable function satisfying

$$\int_{\mathbb{R}} K_0(x) d\lambda(x) = 1.$$

Suppose that  $K_0$  has finite moments up to order  $2r$ . Then the space  $\mathbb{P}_r$  is a Hilbert space under the inner product

$$\langle P, Q \rangle_{K_0} = \int_{\mathbb{R}} P(x)Q(x)K_0(x)d\lambda(x).$$

Since every function in this space can be uniquely represented as a linear combination of elements of the canonical basis  $\{1, x, x^2, \dots, x^r\}$  for  $\mathbb{P}_r$ , the inner product is entirely defined by

$$g_{ij} = \langle x^{i-1}, x^{j-1} \rangle = \int_{\mathbb{R}} x^{i+j-2} K_0(x) d\lambda(x), 1 \leq i, j \leq r$$

where  $G = (g_{ij})$  is the Gram matrix for the basis. Moreover, it is the Hankel matrix for the moments of  $K_0$ . To show that this is indeed a Hilbert space, it's enough to see that

$$\langle x^i, x^i \rangle = \int_{\mathbb{R}} x^{2i} K_0(x) d\lambda(x) \geq 0$$

with equality occurring if and only if  $x^i = 0$ , that is, the zero function in  $\mathbb{P}_r$ . Linearity of the proposed inner product easily follows from the linearity of Lebesgue integration. Finally, symmetry of the inner product follows trivially. As a last note, it's clear that from properties of the inner product, we ascertain that  $G$  is Hermitian ( $G = G^*$ ) from conjugate symmetry and positive definite ( $v^* G v > 0$  whenever  $v \neq 0$ ) from the positive definiteness of the inner product without knowing anything else about this mathematical object.

The maximal orthonormal set of vectors in an inner product space is known as a Hilbert basis. Note that this is distinct from the usual notion of a maximal set of linearly independent vectors, which is known as a Hamel basis to avoid confusion. As is the case with showing the existence of the Hamel basis, it's possible to show the existence of a Hilbert basis using Zorn's lemma. A salient observation is that for finite-dimensional inner product spaces, the two types of bases are one and the same.

**Example 1.22** (Roman Ch. 9, Exer. 9). Show that any nontrivial inner product space has a Hilbert basis.

Let  $V$  be an inner product space with more than one element. Let  $\mathcal{A}$  be a poset of all orthonormal sets in  $V$  with ordering by set inclusion.  $\mathcal{A}$  is not empty since a nonzero  $v \in V$  can be chosen such that  $\{v/||v||\} \in \mathcal{A}$ . Let  $C = \{O_i | i \in I\} \subset \mathcal{A}$  be a nonempty chain in  $\mathcal{A}$ , then the union

$$\bigcup_{i \in I} O_i \in \mathcal{A},$$

and by Zorn's lemma, there exists a maximal element  $O$ , the Hilbert basis.

Since the previous exmp established that every Hilbert space admits an orthonormal basis, it is meaningful to consider the cardinality of bases of a Hilbert space. Two bases of the same Hilbert space are of the same cardinality, which is sometimes called the Hilbert dimension of the space. We say that a Hilbert space is known as separable if and only if it admits a countable orthonormal basis. Separable Hilbert spaces play an important role in physics, in particular, attempts at rigorous mathematical formulations of quantum field theory, such as the Wightman axioms. As shall be seen later in the exposition, there is only one infinite-dimensional separable Hilbert space that is referred to as "the Hilbert space".

So far, the field over which the Hilbert space is considered has been assumed to be  $\mathbb{C}$ . This is for the reason of **complexification**: if the Hilbert space is over  $\mathbb{R}$ , it is isomorphic to its complexification, therefore, it is sufficient to consider the complex Hilbert space.

**1.6. Isometries.** Here is a review of terminology and basic properties of isometries for inner product spaces.

**Definition 1.23** (Isometry). Given inner product spaces  $V, W$  and  $\tau \in \mathcal{L}(V, W)$ ,

- 1)  $\tau$  is an isometry if it preserves the inner product:

$$\langle \tau u, \tau v \rangle = \langle u, v \rangle$$

for all  $u, v \in V$ .

- 2) A bijective isometry is called an isometric isomorphism. When  $\tau : V \rightarrow W$  is an isometric isomorphism, the inner product spaces  $V$  and  $W$  are called isometrically isomorphic.

**Example 1.24** (Roman Ch. 9, Exer. 8). Show that any isometry is injective.

Let  $\tau \in \mathcal{L}(V, W)$ . Suppose  $\tau v = 0$ , then

$$\langle v, v \rangle = \langle \tau v, \tau v \rangle = 0,$$

implying that  $v = 0$ .

Finally note that a linear transformation is an isometry if and only if it preserves the norm:  $\|\tau v\| = \|v\|$ . By definition, an isometry preserves the norm. The converse can be shown by an application of the **polarization identities**.

While the above definition concerns itself primarily with isometries in vector spaces, the notion is also relevant for metric spaces (and thus Hilbert spaces).

**Definition 1.25.** Let  $(M, d)$  and  $(M', d')$  be metric spaces. Then a function  $f : M \rightarrow M'$  is called an isometry if

$$d'(f(x), f(y)) = d(x, y)$$

for all  $x, y \in M$ . If  $f : M \rightarrow M'$  is a bijective isometry, then the two spaces are isometric.

Additional characterizations of isometries between metric spaces are in order.

**Theorem 1.26.** Let  $f : (M, d) \rightarrow (M', d')$  be an isometry, then

- 1)  $f$  is injective,
- 2)  $f$  is continuous,
- 3)  $f^{-1} : f(M) \rightarrow M$  is also an isometry and hence is continuous.

*Proof.* Observe that if

$$f(x) = f(y), d'(f(x), f(y)) = 0 \implies d(x, y) = 0,$$

which only occurs when  $x = y$ . Next, let  $(x_n) \rightarrow x \in M$ . Then

$$d'(f(x_n), f(x)) = d(x_n, x) \rightarrow 0$$

in the limit for  $n \rightarrow \infty$ . Thus,  $(f(x_n)) \rightarrow f(x)$  in the limit, meaning that  $f$  is continuous. Finally,

$$d(f^{-1}(f(x)), f^{-1}(f(y))) = d(x, y) = d'(f(x), f(y)),$$

so  $f^{-1} : f(M) \rightarrow M$  is an isometry. □

Isometries are of great importance to metric spaces since they give a mechanism by which any metric space that is not complete can be embedded in a complete metric space. The following theorem is stated without (lengthy) proof, which can be found on pages 316-320 of Roman's textbook.

**Theorem 1.27.** *Let  $(M, d)$  be any metric space. Then there is a complete metric space  $(M', d')$  and an isometry  $\tau : M \rightarrow \tau M \subset M'$  for which  $\tau M$  is dense in  $M'$ .  $(M', d')$  is called a completion of  $(M, d)$ . Moreover,  $(M', d')$  is unique up to isometric isomorphism.*

With the existence of unique completions for metric spaces in mind, we can return to inner product spaces. Just as any metric space has a completion, so do inner product spaces.

**Theorem 1.28.** *Let  $V$  be an inner product space, then there exists a Hilbert space  $\mathcal{H}$  and an isometry  $\tau : V \rightarrow \mathcal{H}$  for which  $\tau V$  is dense in  $\mathcal{H}$ . Moreover,  $\mathcal{H}$  is unique up to isometric isomorphism.*

*Proof.* By theorem 1.23, we know that any metric space  $(V, d)$ , where  $d$  is induced by the inner product, has a unique completion  $(V', d')$  which consists of equivalence classes of Cauchy sequences in  $V$ . If  $(x_n) \in \overline{(x_n)} \in V'$  and  $(y_n) \in \overline{(y_n)} \in V'$  are Cauchy sequences,

$$\overline{(x_n)} + \overline{(y_n)} = \overline{(x_n + y_n)}, \quad r\overline{(x_n)} = \overline{(rx_n)}$$

and

$$\langle \overline{(x_n)}, \overline{(y_n)} \rangle = \lim_{n \rightarrow \infty} \langle x_n, y_n \rangle.$$

These definitions are well-defined – are independent of the choice of representative from each equivalence class. To see this, if  $(\hat{x}_n) \in \overline{(x_n)}$  then

$$\lim_{n \rightarrow \infty} \|x_n - \hat{x}_n\| = 0$$

and so

$$|\langle x_n, y_n \rangle - \langle \hat{x}_n, y_n \rangle| = |\langle x_n - \hat{x}_n, y_n \rangle| \leq \|x_n - \hat{x}_n\| \|y_n\| \rightarrow 0.$$

Therefore

$$\langle \overline{(x_n)}, \overline{(y_n)} \rangle = \lim_{n \rightarrow \infty} \langle x_n, y_n \rangle = \lim_{n \rightarrow \infty} \langle \hat{x}_n, y_n \rangle = \langle \overline{(\hat{x}_n)}, \overline{(y_n)} \rangle.$$

To see that  $V'$  is an inner product under these operations, observe that positive definiteness follows from sequence norms, symmetry of the inner product from its symmetry in  $V$ , and linearity in the first coordinate also from this property being held over  $V$ . The inner product on  $V'$  induces a metric  $d'$  since

$$\langle \overline{(x_n - y_n)}, \overline{(x_n - y_n)} \rangle = \lim_{n \rightarrow \infty} \langle \overline{x_n - y_n}, \overline{x_n - y_n} \rangle = \lim_{n \rightarrow \infty} d(x, y)^2 = d'((x_n), (y_n))^2.$$

Finally, the metric space isometry  $\tau : V \rightarrow V'$  is an isometry of inner product spaces since

$$\langle \tau x, \tau y \rangle = d'(\tau x, \tau y)^2 = d(x, y)^2 = \langle x, y \rangle.$$

Therefore  $V'$  is a complete inner product space and  $\tau V$  is a dense subspace of  $V'$  that is isometrically isomorphic to  $V$ . □

From Theorem 1.15, Theorem 1.22, and the above discussion of Fourier expansions, the following important results concerning subspaces of inner product spaces follow.

**Theorem 1.29.** *Holding in all generality, we have:*

- 1) Any complete subspace of an inner product space is closed,
- 2) A subspace of a Hilbert space is a Hilbert space if and only if it is closed,

- 3) Any finite-dimensional subspace of an inner product space is closed and complete.

*Proof.* Suppose that  $(x_n)$  is a sequence in  $S$  with  $(x_n) \rightarrow x$  and  $x \notin S$ . Let  $\beta = \{b_1, \dots, b_m\}$  be an orthonormal Hamel basis for  $S$ . The Fourier expansion

$$s = \sum_{i=1}^m \langle x, b_i \rangle b_i$$

in  $S$  has the property that  $x - s \neq 0$  but

$$\langle x - s, b_j \rangle = \langle x, b_j \rangle - \langle s, b_j \rangle = 0.$$

Writing  $y = x - s$  and thus  $y_n = x_n - s \in S$ , the sequence  $(y_n)$ , which converges in  $S$ , converges to a vector  $y$  that is orthogonal to the entire subspace  $S$ . But this is impossible since  $y_n \perp y$  necessarily implies that

$$\|y_n - y\|^2 = \|y_n\|^2 + \|y\|^2 \geq \|y\|^2 \not\rightarrow 0.$$

Therefore,  $S$  is **closed**. □

**1.7. Dual Spaces, Riesz Representation Theorem.** Recall that for a (topological) vector space  $V$ , its dual space  $V^*$  is defined as the space of all continuous linear functionals  $\varphi : V \rightarrow \mathbb{F}$ . In a Hilbert space, it is easy to write down continuous linear functionals: for any  $f \in \mathcal{H}$ , the map  $u \mapsto \langle f, u \rangle$  is a continuous linear functional on  $\mathcal{H}$ . What's really remarkable about Hilbert spaces is that this way of obtaining a continuous linear functional on  $\mathcal{H}$  works the same for *every* linear functional – *all* continuous linear functionals on  $\mathcal{H}$  are obtained in this fashion:

**Theorem 1.30 (Riesz Representation Theorem).** *For any bounded linear functional  $f$  on Hilbert space  $\mathcal{H}$  (that is,  $f \in \mathcal{H}^*$ ), there is a unique  $z_0 \in \mathcal{H}$  such that*

$$f(x) = \langle x, z_0 \rangle$$

*for all  $x \in \mathcal{H}$ . Moreover,  $\|z_0\| = \|f\|$ . An equivalent formulation goes as follows: given any  $\varphi \in \mathcal{H}^*$ , there exists a unique  $f \in \mathcal{H}$  such that*

$$\langle \varphi, u \rangle = \langle f, u \rangle$$

*for all  $u \in \mathcal{H}$ . Moreover,  $\|f\| = \|\varphi\|_{\mathcal{H}^*}$*

*Proof.* We'll prove the above two formulations with two different proofs, the first one being due to Riesz. Note that if  $f = 0$ , then  $z_0 = 0$  so we can assume that  $f \neq 0$ . Let  $K = \ker(f) \neq \mathcal{H}$  and since  $f$  is continuous, we have that  $K$  is closed. Therefore,

$$H = K \oplus K^\perp.$$

By the first isomorphism theorem, when applied to the linear functional  $f : \mathcal{H} \rightarrow \mathbb{F}$ , we have that  $H/K \cong \mathbb{F}$  as vector spaces. But by another application of the first isomorphism theorem, we have that  $H/K \cong K^\perp$  and so it follows that  $K^\perp \cong \mathbb{F}$ . Thus,  $\dim(K^\perp) = 1$ . For any  $z \in K^\perp$ ,

$$x \in K \implies f(x) = 0 = \langle x, z \rangle,$$

and since  $\dim(K^\perp) = 1$ , it remains to find a non-zero  $z \in K^\perp$  such that  $f(z) = \langle z, z \rangle$ . But  $f(rz) = rf(z) = r\langle z, z \rangle = \langle rz, z \rangle$  for all  $r \in \mathbb{F}$ , so  $f(x) = \langle x, z \rangle$  for  $x \in K$ . So if  $z \neq 0$ , we can choose

$$z_0 = \frac{\overline{f(z)}}{\langle z, z \rangle} z.$$

An alternative proof relies on isometries. Consider the map  $T : \mathcal{H} \rightarrow \mathcal{H}^*$  such that for  $f \in \mathcal{H}$ ,  $u \mapsto \langle f, u \rangle$  is a continuous linear functional on  $\mathcal{H}$  (this is the observation prior to this theorem). It defines an element  $Tf \in \mathcal{H}^*$  such that

$$\langle Tf, u \rangle = \langle f, u \rangle$$

for all  $u \in \mathcal{H}$ . Therefore,  $\|Tf\|_{\mathcal{H}^*} = \|f\|$  and so  $T$  is an isometry from  $\mathcal{H}$  to  $T(\mathcal{H})$ , which is a closed subspace of  $\mathcal{H}^*$ . Choose a continuous linear functional  $h$  on  $\mathcal{H}^*$  that vanishes on  $T(\mathcal{H})$ . Since  $\mathcal{H}$  is reflexive,  $h \in \mathcal{H}$  and satisfies  $\langle Tf, h \rangle = 0$  for all  $f \in \mathcal{H}$ . Thus  $\langle f, h \rangle = 0$  for all  $f \in \mathcal{H}$ , so  $h = 0$ .  $\square$

**Example 1.31** (Bra-Ket Notation from Quantum Mechanics). Dirac's Bra and Ket notation for has been immensely useful for physicists, especially in quantum mechanics. Essentially, its utility derives from the notation's ability to emphasize the connection between vectors in Euclidean space and functions in Hilbert space (after all, Hilbert spaces can be considered to be the infinite-dimensional analogs of  $\mathbb{R}^n$  or  $\mathbb{C}^n$ ). In the following discussion, it is important to remember that the physical entities that Dirac's notation intends to represent – state vectors – have an extremely idealized character and are oftentimes not physically realizable. In other words, there exist mathematically rigorous frameworks in which the following discussion is meaningful, though not with the operational ease which is frequently desired in practical applications.

Let a ket  $|\psi\rangle \in \mathcal{H}$  represent an abstract vector in the Hilbert space  $\mathcal{H}$ . It has size and dimension, but without specifying a basis for the space, we only know about its *existence*. Let  $\langle\psi_2|\psi_1\rangle$  be the scalar product of two such ket vectors. By convention, this is linear in the right entry and antilinear in the left-most one. Symbolically, this means that

$$\langle\psi_2|\alpha\psi + \beta\psi'\rangle = \alpha\langle\psi_2|\psi\rangle + \beta\langle\psi_2|\psi'\rangle$$

where  $\alpha, \beta \in \mathbb{C}$  while antilinearity results from

$$\langle\psi_2|\psi_1\rangle = \langle\psi_1|\psi_2\rangle^*.$$

Note that this is exactly the reverse of the definition of the inner product in Definition 1.1! Once the components of  $|\psi\rangle$  are written down (in terms of basis elements), we can see the projection that it has on each of the basis vectors, etc... In particular, if we choose the canonical basis  $\{e_1, e_2, e_3\}$  for  $\mathbb{R}^3$ , then we can write

$$|\psi\rangle = \langle e_1|\psi\rangle|e_1\rangle + \langle e_2|\psi\rangle|e_2\rangle + \langle e_3|\psi\rangle|e_3\rangle.$$

$|\psi\rangle$  can be expressed in terms of whatever basis we like. Whether it's fixing the vector based on any given point (this is  $\langle x|\psi\rangle = \psi(x)$ ) or through a Fourier transform of the vector at some point, it's just an abstract vector that's waiting for some basis.

The *bra* vector  $\langle\psi|$  is associated with the ket  $|\psi\rangle$  and represents a continuous linear functional of the kets:

$$\langle\psi| : |\psi_1\rangle \rightarrow \langle\psi|\psi_1\rangle.$$

That every such continuous functional can be associated with a ket is a consequence of the Riesz representation theorem, which states that the map of  $|\psi\rangle$  onto  $\langle\psi|$  is bijective and antilinear.

**Example 1.32** (Roman Chap. 13, Exer. 14). If  $f \in \mathcal{H}^*$ , then  $\ker(f)$  is a closed subspace of  $\mathcal{H}$ . ( $\mathcal{H}^*$  denotes the set of all bounded linear functionals on  $\mathcal{H}$ )

In a finite-dimensional Hilbert space, all subspaces are closed, unlike in the infinite-dimensional case. So, if  $f$  is bounded, it is continuous and since  $\ker(f) = f^{-1}(\{0\})$ , it is closed since  $\{0\}$  is closed.



**Example 1.33** (Roman Chap 13., Exer. 19). If  $\mathcal{H}$  is a real Hilbert space, then  $\mathcal{H} \cong \mathcal{H}^*$ .

Let  $R : \mathcal{H}^* \rightarrow \mathcal{H}$  be the Riesz map, which is injective. For any  $z_0 \in \mathcal{H}$ , the linear functional  $x \mapsto \langle x, z_0 \rangle$  is bounded, so  $R$  is a bijection. This is a conjugate isomorphism if the Hilbert space is over  $\mathbb{C}$  and an isomorphism if  $\mathcal{H}$  is a real Hilbert space.

## 2. REPRODUCING KERNEL HILBERT SPACE

### 2.1. Definitions.

**Definition 2.1** (Reproducing Kernel). Given a non-empty abstract set  $S$ , a function

$$K : S \times S \rightarrow \mathbb{C},$$

$$(s, t) \mapsto K(s, t)$$

is a reproducing kernel of the Hilbert space  $\mathcal{H}$  if and only if for all  $t \in S$ ,

- a)  $K(., t) \in \mathcal{H}$
- b) for all  $\varphi \in \mathcal{H}$ ,  $\langle \varphi, K(., t) \rangle = \varphi(t)$ .

Condition b) is known as the *reproducing property*. Intuitively, the value of the function  $\varphi$  at the point  $t$  is reproduced by the inner product of  $\varphi$  with  $K(., t)$ . A straightforward consequence of these definitions is that for all  $(s, t) \in S \times S$ ,

$$K(s, t) = \langle K(., t), K(., s) \rangle.$$

A Reproducing Kernel Hilbert Space is associated with a kernel that reproduces every function in the space in the sense that for any  $t \in S$  on which the functions are defined, "evaluation at  $t$ " can be performed by taking an inner product with a function determined by the kernel. As will be proved in the sequel, the reproducing kernel is also positive definite.

**Example 2.2** (Vector Space of Functions). Let  $\mathcal{H}$  be a finite dimensional complex vector space of functions with basis  $(f_1, f_2, \dots, f_n)$ . Therefore an inner product  $\langle ., . \rangle$  on  $\mathcal{H}$  is entirely defined by

$$g_{ij} = \langle f_i, f_j \rangle, \quad 1 \leq i, j \leq n$$

so that for two functions

$$v = \sum_{i=1}^n v_i f_i \text{ and } w = \sum_{j=1}^n w_j f_j,$$

their inner product is

$$\langle v, w \rangle = \left\langle \sum_{i=1}^n v_i f_i, \sum_{j=1}^n w_j f_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n v_i \bar{w}_j g_{ij}.$$

Choose an orthonormal basis  $(e_1, e_2, \dots, e_n)$  for  $\mathcal{H}$  and define the kernel

$$K(x, y) = \sum_{i=1}^n e_i(x) \bar{e}_i(y).$$

Then for any  $y \in S$ ,

$$K(., y) = \sum_{i=1}^n \bar{e}_i(y) e_i(.)$$

belongs to  $\mathcal{H}$  and for any function

$$\varphi(.) = \sum_{i=1}^n \lambda_i e_i(.) \in \mathcal{H},$$

for all  $y \in S$ ,

$$\langle \varphi, K(., y) \rangle = \left\langle \sum_{i=1}^n \lambda_i e_i(., y), \sum_{j=1}^n \bar{e}_j(y) e_j(.) \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \bar{e}_j(y) \langle e_i, e_j \rangle.$$

By definitions, since  $\langle e_i, e_j \rangle = \delta_{ij}$  and complex conjugation is an involution, it follows that

$$\langle \varphi, K(., y) \rangle = \sum_{i=1}^n \lambda_i e_i(y) = \varphi(y).$$

The conclusion is that any finite dimensional Hilbert space of functions has a reproducing kernel.

**Example 2.3.** Let  $E = \mathbb{N}^*$  be the set of positive integers and  $\mathcal{H} = \ell^2(\mathbb{C})$  be the set of complex sequences  $(x_i)_{i \in \mathbb{N}^*}$  such that

$$\sum_{i \in \mathbb{N}^*} |x_i|^2 < \infty.$$

Then  $\mathcal{H}$  is a Hilbert space with inner product

$$\langle x, y \rangle = \sum_{i \in \mathbb{N}} x_i \bar{y}_i, \quad x = (x_i), y = (y_i).$$

The kernel is  $K(i, j) = \delta_{ij}$ , then for all  $j \in \mathbb{N}$ ,

$$K(., j) = (0, 0, \dots, 0, 1, 0, \dots) \in \mathcal{H} \text{ (1 at the } j\text{-th term)}$$

$$\forall x = (x_i)_{i \in \mathbb{N}} \in \mathcal{H}, \quad \langle x, K(., j) \rangle = \sum_{i \in \mathbb{N}} x_i \bar{\delta}_{ij} = x_j.$$

Despite the fact that all infinite dimensional separable Hilbert spaces are isometrically isomorphic to  $\ell^2(\mathbb{C})$  (as can be found [here](#)), which as we saw in the previous exmp, has a reproducing kernel, the property is not sufficient to imply that a given separable Hilbert space of functions has a reproducing kernel. Specifically, while  $\ell^2(\mathbb{N}) \cong L^2[0, 1]$  as Hilbert spaces, the former is an RKHS while the latter is not. In order to understand why this is the case, it is necessary to provide a characterization of the conditions that must be satisfied in order for a reproducing kernel to exist.

**Example 2.4** (Space of Square-Integrable Functions is not a RKHS). A standard analysis fact is that  $L^2[a, b]$  is the only  $L^p$  space that is a Hilbert space, so we take it for granted that this is a Hilbert space. Recall that if  $f : [0, 1] \rightarrow \mathbb{C} \in L^2[0, 1]$  then

$$\int_0^1 |f(x)|^2 dx < \infty.$$

This space is a Hilbert space under the inner product given by

$$\langle f, g \rangle = \int_0^1 f(x) \overline{g(x)} dx.$$

In fact, square integrability is equivalent to saying that  $\langle f, f \rangle < \infty$ . When thought of as an inner product space, a square-integrable function is actually an equivalence class of functions that are equal almost everywhere. The reason that this space doesn't satisfy the characterization derived in Theorem 1 is that it is not meaningful to define the "evaluation functional" on a representative of an equivalence class. Indeed, for any  $x$  and  $y \in \mathbb{R}$ , every equivalence class has some function for which  $f(x) = y$ . Note, however, that it is still meaningful to write down integrals of equivalence classes as integrals of arbitrary representatives since every function in an equivalence class has the same integral as they differ on sets with measure zero by construction.

**Example 2.5** (Berlinet Chap. 1, Exer. 5 - Sobolev Space  $H^1(\mathbb{R})$ ). [This video](#) provides the fundamental intuition behind Sobolev spaces, which are simple exmps of Hilbert spaces that arise in differential equations. The space is defined as

$$\mathcal{H} = \{f : [0, 1] \rightarrow \mathbb{R} : f \text{ is completely continuous, } f(0) = f(1) = 0, f, f' \in L^2[0, 1]\}.$$

As an analysis review aside, if a function is absolutely continuous, it is differentiable almost everywhere and is equal to the integral of its derivative.  $\mathcal{H}$  is a vector space of functions on  $[0, 1]$  (this is a frequently used tool – finding a space in which functions can be thought of as vectors, since we know how to operate with the latter better than with the former). Suppose we define the inner product on this space as

$$\langle f, g \rangle = \int_0^1 f'(t)g'(t)dt.$$

Then

$$f(x) = \int_0^1 f'(t)dt = \int_0^1 f'(t)\chi_{[0,x]}(t)dt.$$

By Cauchy-Schwarz, it follows that

$$|f(x)| \leq \left( \int_0^1 f'(t)^2 dt \right)^{1/2} \left( \int_0^1 \chi_{[0,x]}(t) dt \right)^{1/2} = \|f'\| \sqrt{x}.$$

Moreover, it's not difficult to confirm that the sequences  $\{f_n\}$  and  $\{f'_n\}$  are Cauchy and from there to show that this space is a RKHS. To find the kernel function, consider the following:

$$f(y) = \langle f, K(t, y) \rangle = \int_0^1 f'(t)K'(t, y)dt = - \int_0^1 f(t)K''(t, y)dt,$$

from an integration by parts. Next, because

$$f(y) = \int_0^1 f(t)\delta_y(t)dt$$

where  $\delta_y(t)$  is the usual Dirac delta function/distribution, we have that the kernel satisfies the differential equation

$$-K''(t, y) = \delta_y(t)$$

which has boundary conditions  $K(0, y) = K(1, y) = 0$ . This is a standard problem for a Green's function. One integration yields

$$K'(t, y) = -\chi_y(t) + C_1$$

where  $\chi_y(t)$  is the Heaviside function. Another integration yields

$$K(t, y) = -\rho_y(t) + C_1x + C_2$$

where  $\rho_y(t)$  is the ramp function. Combining this with the initial conditions yields the kernel function

$$K(t, y) = \begin{cases} (1-y)x, & t < y \\ (1-x)y, & t \geq y \end{cases}.$$

Interestingly, the space remains a RKHS under the different inner product

$$\langle f, g \rangle = \int_0^1 (f(t)g(t) - f'(t)g'(t))dt.$$

In this space, the kernel is

$$K(x, y) = \frac{1}{2} \exp(-|x - y|).$$

What is far more interesting is the fact that these notions generalize to arbitrary Sobolev spaces, which are some of the most highly studied spaces in applied mathematics. The general kernel computation for these spaces was completed in [this publication](#) and applications of the Sobolev space RKHS to polynomial integration tractability are discussed.

Hopefully these examples have been illustrative of the idea that RKHS's are ubiquitous in applied mathematics. But it's still not clear exactly why they are useful, nor whether or not there are conditions that we can check in an arbitrary scenario to determine whether a space is a RKHS or not. Ideally, we would like necessary and sufficient conditions for a space to be a RKHS.

From the above, we have the alternative characterization of an RKHS. A Hilbert space  $\mathcal{H} \subset \mathbb{R}^X$  is an RKHS if and only if for all  $x \in X$ , the linear functional  $\delta_x : \mathcal{H} \rightarrow \mathbb{R} : f \mapsto f(x)$  is continuous. As a corollary, we have that convergence in an RKHS implies pointwise convergence. We also have a smoothness criteria.

**Theorem 2.6.** *For any function  $f$  in an RKHS  $\mathcal{H}$ , for any  $x, y \in X$ ,*

$$|f(x) - f(y)| \leq \|f\|_{\mathcal{H}} \|K_x - K_y\|_{\mathcal{H}}.$$

*Proof.* Roughly, note that

$$|f(x) - f(y)| = |\langle f, K(x, t) \rangle - \langle f, K(y, t) \rangle| = |\langle f, K_x - K_y \rangle| \leq \|f\|_{\mathcal{H}} \|K_x - K_y\|_{\mathcal{H}}$$

where the latter is  $d(x, y)$  and defines the geometry in  $\mathcal{H}$ .  $\square$

## 2.2. Characterizing RKHS's: Continuous Evaluation Functionals.

**Theorem 2.7.** *A reproducing kernel exists if and only if every evaluation functional  $e_t, t \in S$  is continuous on  $\mathcal{H}$ .*

*Proof.* If  $\mathcal{H}$  has a reproducing kernel  $K$ , then for any  $t \in S$ , it's true that for all  $\varphi \in \mathcal{H}$ ,

$$e_t(\varphi) = \langle \varphi, K(., t) \rangle.$$

This is just the reproducing property of an RKHS. So we know that the evaluation functional is linear (the inner product is linear in its first argument), so by Cauchy-Schwarz, it's continuous:

$$|e_t(\varphi)| = |\langle \varphi, K(., t) \rangle| \leq \|\varphi\| \|K(., t)\| = \|\varphi\| [K(t, t)]^{1/2}.$$

If  $\varphi = K(., t)$  is the kernel, then by Cauchy-Schwarz's second condition, the upper bound can be obtained. Thus,

$$\|e_t\| = \sup_{\|\varphi\| \neq 0} \frac{|e_t(\varphi)|}{\|\varphi\|} = [K(t, t)]^{1/2}.$$

Conversely, from the Riesz representation theorem, if the linear mapping  $\varphi \mapsto e_t(\varphi) = \varphi(t)$  is continuous, there exists a function  $N_t(.) \in \mathcal{H}$  such that for all  $\varphi \in \mathcal{H}$ ,

$$\langle \varphi, N_t(.) \rangle = \varphi(t).$$

If this is true for any  $t \in S$ , then  $K(s, t) = N_t(s)$  is the reproducing kernel of  $\mathcal{H}$ .  $\square$

**2.3. Characterizing RKHS's: Positive Type Functions (Kernels).** The aim of this next subsection is to show that the set of positive type functions and the set of reproducing kernels on  $S \times S$  are identical.

**Definition 2.8** (Positive Type Function). A function  $K : S \times S \rightarrow \mathbb{C}$  is a positive type function (alternatively called positive definite function or a kernel in other contexts) if for all  $n \geq 1$  and for all complex  $n$ -tuples  $(a_1, \dots, a_n) \in \mathbb{C}^n$  and tuples from  $S$ ,  $(x_1, \dots, x_n) \in S^n$ ,

$$\sum_{i=1}^n \sum_{j=1}^n a_i \bar{a}_j K(x_i, x_j) \in \mathbb{R}^+.$$

Note that this is equivalent to the positive definiteness of the matrix

$$(K(x_i, x_j))_{1 \leq i, j \leq n}$$

for any choice of  $n \in \mathbb{N}^*$ .

**Example 2.9.** The delta function

$$(x, y) \mapsto \delta_{xy} = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$$

is positive definite.

Let  $n \geq 1$  and choose vectors  $(a_1, \dots, a_n) \in \mathbb{C}^n$ ,  $(x_1, \dots, x_n) \in S^n$  and  $\{\alpha_1, \dots, \alpha_p\}$  the set of different values among the coordinates of the latter vector. Then

$$\sum_{i=1}^n \sum_{j=1}^n a_i \bar{a}_j \delta_{x_i x_j} = \sum_{i=1}^n \sum_{x_j = x_i} a_i \bar{a}_j = \sum_{k=1}^p \sum_{x_i = x_j = \alpha_k} a_i \bar{a}_j = \sum_{k=1}^p \left| \sum_{x_i = \alpha_k} a_i \right|^2 \in \mathbb{R}^+.$$

Evidently the product  $\alpha K$  of a positive type function  $K$  with a non-negative constant  $\alpha$  is also a positive type function.

Showing that a particular function is a positive type function is frequently an intractable problem. So instead, we show a lemma that states that if we can write  $K(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}$  in some space  $\mathcal{H}$ , then that is sufficient to show that it is positive definite.

**Theorem 2.10.** Let  $\mathcal{H}$  be some Hilbert space with inner product  $\langle \cdot, \cdot \rangle$  and let  $\varphi : S \rightarrow \mathcal{H}$ . Then the function  $K$

$$S \times S \rightarrow \mathbb{C}$$

$$(x, y) \mapsto K(x, y) = \langle \varphi(x), \varphi(y) \rangle$$

is a positive type function.

*Proof.* Observe that

$$\sum_{i=1}^n \sum_{j=1}^n a_i \bar{a}_j K(x_i, x_j) = \sum_{i=1}^n \sum_{j=1}^n \langle a_i \varphi(x_i), a_j \varphi(x_j) \rangle = \left\| \sum_{i=1}^n a_i \varphi(x_i) \right\|^2.$$

□

#### 2.4. RKHS and Positive Type Functions.

**Theorem 2.11.** *Any reproducing kernel is a positive type function.*

*Proof.* If  $K$  is a reproducing kernel of  $\mathcal{H}$ , then

$$\sum_{i=1}^n \sum_{j=1}^n a_i \bar{a}_j K(x_i, x_j) = \left\| \sum_{i=1}^n \bar{a}_i K(., x_i) \right\|^2 \in \mathbb{R}^+.$$

□

Thus we have shown that an RKHS defines a kernel that is both symmetric and positive definite (a function of positive type). The Moore-Aronszajn theorem goes in the opposite direction by stating that every symmetric, positive definite kernel, defines a unique RKHS.

**Theorem 2.12** (Moore-Aronszajn). *Suppose  $K$  is a symmetric, positive definite kernel on a set  $X$ . Then there is a unique Hilbert space of functions on  $X$  for which  $K$  is a reproducing kernel.*

The original proof for this theorem can be found in Aronszajn's [Theory of Reproducing Kernels](#), although he attributes it to Moore. The proof is a nice synthesis of the concepts covered in this chapter of the exposition - Hilbert spaces, their completions, and orthogonality among subspaces. This, alongside Riesz, is one of the main results here.

We can also combine positive definite kernels. Given two positive definite kernels  $K_1, K_2$  and  $\alpha \geq 0$ , all of  $K_1 + K_2$ ,  $K_1 \times K_2$  and  $\alpha K_1$  are positive definite kernels. Moreover, if  $\{K_i\}_{i \geq 1}$  is a sequence of positive definite kernels that converges pointwise to  $K$ , then  $K$  is a positive definite kernel.

**2.5. The Representer Theorem.** Having characterized RKHS's, we now move to the *why*. Although we consider a particular example of a more general family of theorems, called *Representer theorems*, we give a broad overview for why they are useful. In most applications, they are useful because they simplify empirical risk minimization problems, which aim at bounding the performance of an algorithm. Intuitively, before applying an algorithm, we don't know about the structure of the data that it will be operating on, but we can measure its performance on a known set of training data, the "empirical" risk.

The search domain for the minimization is usually an infinite-dimensional space and hence it is impossible to perform a search on a finite-memory computer. The representation that Representer theorems grant us reduces the original infinite-dimensional minimization problem to essentially a search for an optimal  $n$ -dimensional vector. This is already a solved problem and any standard function minimization algorithm can be applied. Thus, Representer theorems establish a theoretical basis for the reduction of a general machine learning problem to algorithms that can be practically executed on computers. Below, we give a simple statement, due to Schölkopf and Smola.

**Theorem 2.13** (Representer Theorem). *Let  $X$  be a set endowed with a positive definite kernel  $K$ . Let  $\mathcal{H}$  be the associated RKHS. Let  $\mathcal{S} := \{x_i\}_{i=1}^n$  be a finite set of points in  $X$ . Let  $\Psi : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  be strictly increasing with respect to its right-most (last) variable. The solution of the problem*

$$\min_{f \in \mathcal{H}} \{\Psi(f(x_1), \dots, f(x_n), \|f\|_{\mathcal{H}})\}$$

*for all  $x \in X$  is of the form*

$$f(x) = \alpha^T K(x) = \sum_{i=1}^n \alpha_K(x_i, x)$$

where  $\alpha = [\alpha_1 \cdots \alpha_n]^T$  and  $K(x) = [K(x_1, x) \cdots K(x_n, x)]^T$ . The solution lies in a subspace of dimension  $n$ . Practically, if we let  $K = [K(x_i, x_j)]_{i,j \in [n]}$  be the Gram matrix of  $\mathcal{S}$ , we can solve

$$\min_{\alpha} \{\Psi([K\alpha]_1, \dots, [K\alpha]_n, \alpha^T K \alpha)\}$$

where  $[K\alpha]_j := K(x_j)$ . This can be solved analytically or by numerical methods.

A proof can be found in Schölkopf and Smola's book [Learning with kernels](#).

### 3. STOCHASTICS, OR SOMETHING

Of particular use to probability theory is the space  $L^2(\Omega, \mathcal{A}, P)$  of square integrable random variables on some probability space  $(\Omega, \mathcal{A}, P)$ . In this context, showing that a function  $K$  is of positive type shows that it is the covariance function (also called the kernel) of some complex-valued zero mean stochastic process  $\{X_t\}_{t \in S}$ . Since

$$0 \leq \text{Var} \left( \sum_{i=1}^n a_i X_{t_i} \right) = \sum_{i=1}^n \sum_{j=1}^n a_i \bar{a}_j \mathbb{E}(X_{t_i}, \bar{X}_{t_j}),$$

the covariance function

$$\begin{aligned} S \times S &\rightarrow \mathbb{C} \\ \mathbb{E}(X_t \bar{X}_s) &= \langle X_t, X_s \rangle \end{aligned}$$

is a positive type function. To elaborate further, recall that  $L^2(\Omega, \mathcal{A}, P)$  is a Hilbert space equipped with the inner product

$$\langle X, Y \rangle = \mathbb{E}(X \bar{Y}).$$

Let  $X_t$  for  $t$  ranging through a set  $T$  (often a subset of  $\mathbb{R}^p$ , if  $p = 1$ , then the term "process" is used for  $X_t$  and otherwise the term "random field" is employed) be a second order stochastic process defined on the probability space  $(\Omega, \mathcal{A}, P)$  with values in  $\mathbb{R}$  or  $\mathbb{C}$ . Then denote  $m$  as the mean function of the process

$$m(t) = \mathbb{E}(X_t)$$

the second moment function by  $R$ ,

$$R(t, s) = \mathbb{E}(X_t \bar{X}_s)$$

and the covariance function by  $K$

$$K(t, s) = \mathbb{E}[(X_t - \mathbb{E}(X_t))(\bar{X}_s - \mathbb{E}(\bar{X}_s))] = \text{Cov}(X_t, X_s).$$

Then

$$R(s, t) = m(t) \overline{m(s)} + K(s, t).$$

The following theorem establishes the exact equivalence between the class of positive type functions and class of covariance functions.

**Theorem 3.1** (Loéve's Theorem).  *$R$  is a second moment function of a second order stochastic process indexed by  $T$  if and only if  $R$  is a function of positive type on  $T \times T$ .*

*Proof.* By the definition given above of the second moment function, it's clear that it is of positive type. Conversely, suppose  $R$  is a positive type function on  $T \times T$ . Then for any finite subset  $T_n = \{t_1, \dots, t_n\} \in T$ , the form

$$Q(u) = \sum_{i,j=1}^n R(t_i, t_j) u_i u_j$$

is positive definite (it's a quadratic form) and there exist jointly normal random variables  $X_{t_i}$  such that  $R(t_i, t_j) = \mathbb{E}(X_{t_i} X_{t_j})$ . It's enough to check that these laws defined for any finite subset  $T_n$  are consistent.  $\square$

**Example 3.2** (Stationary Process). A stationary process is a stochastic process whose unconditional joint probability distribution does not change when shifted in time. Consequently, neither the mean nor the variance change over time. Oftentimes, as is the case with time series analysis, non-stationary data are transformed to become stationary to ease analyses. More rigorously, a second order process is said to be strongly stationary when its distribution is translation invariant. For any increment vector  $h \in T \subset \mathbb{R}^d$ , the joint distribution of

$$X_{t_1+h}, \dots, X_{t_n+h}$$

is the same as the distribution of

$$X_{t_1}, \dots, X_{t_n}.$$

When the discussion is about random fields, the term "homogeneous" is oftentimes employed. The proper covariance of a stationary process is translation invariant, as  $K(s, t) = k(s - t)$ . The positive definiteness of  $K$  corresponds to Bochner's theorem, which allows us to define the spectral measure of a stationary process as the Fourier transform of the positive type function  $k$ . This holds for a larger class of processes that retain the statistical advantages of strongly stationary processes: those of weakly stationary processes. These processes have a constant mean and their proper covariance is translation invariant.

**3.1. Hilbert Space Generated by a Process:** Let  $\mathcal{H}$  be a Hilbert space and  $\{\phi(t), t \in T\}$  be a family of vectors of  $\mathcal{H}$ . Denote  $\overline{\mathcal{L}}(\phi(t), t \in T)$  as the closure of  $\mathcal{H}$  of the linear span  $\mathcal{L}(\phi(t), t \in T)$  generated by  $\{\phi(t), t \in T\}$ .  $\mathcal{L}(\phi(t), t \in T)$ , a subspace of  $\mathcal{H}$ , is called the Hilbert subspace generated by the set of vectors. Analogously, we can define a Hilbert space that is generated by a process to be  $\overline{\mathcal{L}}(X)$ . Let the linear space spanned by the process  $\mathcal{L}(X)$  be the set of finite linear combinations of random variables of the form  $X_{t_i}$  for  $t_i \in T$ .  $\mathcal{L}(X)$  can be equipped with an inner product but does not necessarily possess the completeness property necessary for the space to actually be a Hilbert one. To remedy this, define  $\overline{\mathcal{L}}(X)$  as the closure in  $L^2(\Omega, \mathcal{A}, P)$  of  $\mathcal{L}(X)$ . The set is equipped with the inner product induced by that of  $L^2(\Omega, \mathcal{A}, P)$ :

$$\langle U, V \rangle_{\overline{\mathcal{L}}(X)} = \mathbb{E}(UV).$$

This set contains all the random variable that can be obtained by linear operations, including limits, and the measurements of the process. Also, since

$$\mathbb{E}(UV) = \text{Cov}(U, V) + \mathbb{E}_m(U)\mathbb{E}_m(V)$$

depends on the mean function for any  $U, V \in \overline{\mathcal{L}}(X)$ , the inner product does not depend on the mean and the random variables may not be the same for all values of  $m$ , with the exception of the case when  $T$  is a finite set.

**3.2. Representation Theorems:** Let  $\{X_t, t \in T\}$  be a second order stochastic with covariance function  $R$ . Representation theorems find various concrete function spaces that are isometrically isomorphic to the Hilbert space generated by the process  $\overline{\mathcal{L}}(X)$ .

**Definition 3.3.** The family of vectors  $\{\phi(t), t \in T\}$  is said to be a Hilbert space  $\mathcal{H}$  with inner product  $\langle \cdot, \cdot \rangle$  is a representation of the process  $X_t$  if for every  $s, t \in T$ ,

$$\langle \phi(t), \phi(s) \rangle = R(t, s).$$



Intuitively, the above definition states that the family of vectors is a representation of the process  $X_t$  if the Hilbert space  $\overline{\mathcal{L}}(\phi(t), t \in T)$  is isomorphic to the Hilbert space generated by the process. The following congruence theorem is fundamental to understanding the representation theorems that follow.

**Theorem 3.4** (Basic Congruence Theorem). *Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be Hilbert spaces equipped with respective inner products  $\langle \cdot, \cdot \rangle_1$  and  $\langle \cdot, \cdot \rangle_2$ . Let  $\{u(t), t \in T\}$  be a family of vectors which spans  $\mathcal{H}_1$  and  $\{v(t), t \in T\}$  be a family of vectors spanning  $\mathcal{H}_2$ . If for every  $s, t \in T$ , we have that*

$$\langle u(s), u(t) \rangle_1 = \langle v(s), v(t) \rangle_2,$$

then  $\mathcal{H}_1 \cong \mathcal{H}_2$ .

As is usually the case with isomorphic objects in mathematics, there exists a "canonical congruence". If  $\{\phi(t), t \in T\}$  is a family of vectors in  $\mathcal{H}$  and  $K$  is defined by

$$K(s, t) = \langle \phi(s), \phi(t) \rangle,$$

then  $\mathcal{H}_K \cong \overline{\mathcal{L}}(\phi(t), t \in T)$  by the map

$$J(K(\cdot, t)) = \phi(t).$$

Perhaps a more concrete way to think about this (bypassing the potentially confusing kernel notation) is that there exists a congruence  $\psi : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  such that

$$\psi(u(t)) = v(t), \quad \forall t \in T,$$

which we can define as follows. For each vector in the family  $\{u(t), t \in T\}$ , define  $\psi(u(t)) = v(t)$ . Then for each vector  $u$  that is in  $\mathcal{L}(u(t), t \in T)$  (the linear manifold, actually),

$$\psi(u) = \sum c_i v(t_i), \quad \text{if } u = \sum c_i u(t_i).$$

It remains to show well-definedness and then the convergence of sequences.

From the Gram-Schmidt orthogonalization process, it's easy to see that a Hilbert space possesses an orthonormal basis. Using that in combination with the Basic Congruence Theorem, it follows that any two Hilbert spaces of the same dimension are congruent.

**3.2.1. Loève Representation Theorem:** The Hilbert space  $\overline{\mathcal{L}}(X)$  generated by the process  $\{X_t, t \in T\}$  with covariance function  $R$  is congruent to the RKHS  $\mathcal{H}_R$ .

*Proof.* Let  $\psi : \overline{\mathcal{L}}(X) \rightarrow \mathcal{H}_R$  defined on  $\mathcal{L}(X)$  by the following: for all  $a_1, a_2, \dots, a_n \in \mathbb{R}$ ,

$$\psi \left( \sum_{i=1}^n a_i X_{t_i} \right) = \sum_{i=1}^n a_i R(t_i, \cdot).$$

By continuity,  $\psi$  can be extended to  $\overline{\mathcal{L}}(X)$ . Next, since

$$\langle X_t, X_s \rangle_{\overline{\mathcal{L}}(X)} = R(t, s) = \langle R(t, \cdot), R(s, \cdot) \rangle_{\mathcal{H}_R}.$$

□

**3.3. Karhunen-Loève Expansion Theorem.** A common thread through mathematical analysis is the expansion of entities in terms of simple ones. For instance, one of the most famous analytic methods, the Fourier expansion, expresses functions in terms of the orthonormal basis of trigonometric functions. With this in mind, the Karhunen-Loève theorem can be thought of as the stochastic parallel of Fourier expansions. Historically, this theorem created a bond between stochastic analysis and information theory.

From a high level, Karhunen-Loève operates the same as a Fourier expansion, except on continuous-parameter random functions in the set  $\{X_t, t \in [a, b]\}$  in  $L^2(\Omega, \mathcal{F}, P)$ . The basis that a particular random function is projected onto is spanned by the eigenfunctions of Hilbert-Schmidt integral operator. Thus, the coefficients of the infinite linear combination are expected to be random variables since for every  $t \in [a, b]$ , they are the projection of a random variable onto a deterministic orthogonal basis. Consequently, these random coefficients are orthogonal in  $L^2(\Omega)$ , or equivalently, they are uncorrelated.

**Theorem 3.5** (Karhunen-Loève). *Let  $X = \{X_t\}, t \in [a, b] \subset \mathbb{R}, a, b < \infty$  be a continuous-parameter real-valued second-order random process having zero mean and a continuous covariance function  $K_X(t, \tau)$ . Then for every  $t \in [a, b]$ , we may decompose*

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$$

with

$$Z_k = \int_a^b X_t e_k(t) dt$$

where the  $\{e_k\}_{k=1}^{\infty}$  are the eigenfunctions of the Hilbert-Schmidt integral operator on  $L^2[a, b]$ . The series

$$\sum_{k=1}^{\infty} Z_k e_k(t)$$

converges to  $X_t$  in mean square, uniformly for  $t \in [a, b]$ .

A number of expositions on this topic have served as thesis projects for math majors, and I will simply cite [Giordano Giambartolomei's thesis](#) for a proof and an incredible discussion of the theory underlying this remarkable theorem.

As with other expansion theorems, after performing such an expansion, it is meaningful to consider different types of approximations that it gives. Expectedly, this is a useful way to go about determining whether there is any valuable information in a signal originating in a noisy channel. It is also of great importance for SVD, which is widely known to have a myriad of applications to image processing and the like. See [this survey paper](#) that connects SVD, the above transform, and principal component analysis.

#### 4. ACKNOWLEDGMENTS

I am tremendously grateful for John Ahn's patience with me throughout the program. His insight provided the core of my intuitive understanding of most of this content and his maturity has considerably helped ease me out of my adolescent need for instant gratification, both as a math student and a regular bro. Check him out on [his personal webpage](#)!

Thank you also to [Brown University DRP](#) for organizing this fantastic opportunity for mathematical enrichment. As a first semester freshman, this has undoubtedly been the highlight of my math "career".

Finally, credit goes to the [University of Chicago REU](#) for the  $\text{\LaTeX}$  template!

## 5. BIBLIOGRAPHY

The bibliography lists all sources that I have used and referenced. As stated in the abstract, I make no claim on the originality of the work beyond the order in which material has been presented and my solutions to the problems. I have elucidated on points in theorems where I have been able to and/or thought useful for ease of exposition if I were learning this material anew. But as is often the predicament of an undergraduate freshman reflecting on the work of professional mathematicians who have been producing world-class educational material for decades, it is difficult to restate concepts in a novel way when my understanding has already been determined by – is indeed almost wholly predicated upon – their way of thinking and notation.

## REFERENCES

- [1] <https://terrytao.wordpress.com/2010/03/11/a-type-diagram-for-function-spaces/>
- [2] Steven Roman. Advanced Linear Algebra. <https://www.springer.com/gp/book/9780387728285>
- [3] S. Roman. Advanced Linear Algebra 3e. Springer Press, 2008.
- [4] H. Brezis. Functional Analysis, Sobolev Spaces, and Partial Differential Equations. Springer Press, 2011.
- [5] A. Berline, C. Thomas-Agnan. Reproducing Kernel Hilbert Spaces in Probability and Statistics. Springer Press, 2004.
- [6] B. Schölkopf, A. Smola. Learning with kernels. MIT Press, 2001.