Theory of Reproducing Kernel Hilbert Spaces

Rohan Joy

A Thesis Submitted to

Indian Institute of Technology Hyderabad

In Partial Fulfillment of the Requirements for

The Degree of Master of Science

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Mathematics

May 2018

Approval Sheet

This Thesis entitled Theory of Reproducing Kernel Hilbert Spaces by Rohan Joy is approved for the degree of Master of Science from IIT Hyderabad.

(Dr. G Ramesh) Adviser
Dept. of Mathematics
IITH

11/5/18

(Dr. C S Sastry) Faculty Adviser
Dept. of Mathematics
IITH

11/5/18

(Dr. J Balasubramaniam) H.O.D
Dept. of Mathematics
IITH

Declaration

I declare that this written submission represents my ideas in my own words, and where ideas or words of others have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

Rohan Joy

(Signature)

ROHAN JOY

(Rohan Joy)

MA16MSCST11011

(Roll No.)

Acknowledgements

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I am highly indebted to my supervisor Dr.G. Ramesh for his guidance and constant supervision during my project. I would like to express my gratitude towards my parents for their love and encouragement,without whom I would never have enjoyed so many oppurtunities. My thanks and appreciation also goes to my classmates and seniors who have willingly helped me out with their abilities.

Abstract

A Hilbert Space $\mathcal{H}$ is a real or complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product. A Reproducing Kernel Hilbert Space(RKHS) is a Hilbert Space of functions in which point evaluations are continuous.

An RKHS is associated with a kernel that reproduces every function in the space. This means that for any element in the set on which the functions are defined, evaluation at element can be performed by taking an inner product with a function determined by the kernel. Such a reproducing kernel exists if and only if every evaluation functional is continuous.

In his 1907 work concerning boundary value problems for harmonic and biharmonic functions, Stanislaw Zaremba introduced reproducing kernels. James Mercer simultaneously examined functions which satisfy the reproducing property in the theory of integral equations. Later further work on this topic was done by mathematicians like Gabor Szego,Stefan Bergman, and Salomon Bochner. The subject was systemtically developed in the early 1950's by Nachman Aronszan and Stefan Bergman.

These spaces have wide applications, including complex analysis, harmonic analysis and quantum mechanics. Reproducing kernel Hilbert spaces are particularly important in the field of statistical learning theory because of the celebrated representer theorem which states that every function in an RKHS that minimises an empirical risk function can be written as a linear combination of the kernel function evaluated at the training points. This is a practically useful result as it effectively simplifies the empirical risk minimizaton problem from an infinite dimensional to a finite dimensional optimization problem.

Contents

Chapter 1

Notations

1. $C[a, b]$: the class of all real-valued continuous functions on [0,1]
2. $\Re[z]$: the real part of the complex number z
3. $\Im[z]$: the imaginary part of the complex number z
4. ${l_p}^n$: the space of all n-tuples $(x_1, x_2, ..., x_n) \in \mathbb{C}$ with the norm $\|x\| = (\sum_{i=1}^{i=n} |x_i|^p)^{\frac{1}{p}}$
5. l_p: the space of all sequences (x_n) such that $\sum_{n=1}^{\infty} |x_n|^p < \infty$
6. $L^p(E)$: the space of all measurable p-integrable functions on E

Chapter 2

Hilbert Spaces

Definition 2.0.1. *A complex vector space X is called an inner product space if to each pair of elements x, y of X is associated a complex number $\langle x, y\rangle$ called the inner product of x and y that satisfies the following four conditions:*

1. $\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle \forall x, y, z \in X$.
2. $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle \ \forall \alpha \in \mathbb{C}$.
3. $\langle x, y\rangle = \overline{\langle y, x\rangle}$. *(the bar denotes complex conjugation)*
4. $\langle x, x\rangle \geq 0$ *and* $\langle x, x\rangle = 0$ *iff* $x = 0$.

   Note that conditions 1, 2 *and* 3 *imply that*

5. $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$.

   Proof.

$$\begin{aligned} \langle x, y+z\rangle &= \overline{\langle y+z, x\rangle} \qquad \text{from} \quad (3) \\ &= \overline{\langle y+x\rangle + \langle z, x\rangle} \qquad \text{from} \quad (1) \\ &= \overline{\langle y, x\rangle} + \overline{\overline{\langle x, z\rangle}} \\ &= \langle x, y\rangle + \langle x, z\rangle. \end{aligned}$$

□

6. $\langle x, \alpha y\rangle = \bar{\alpha}\langle x, y\rangle$.

   Proof.

$$\begin{aligned} \langle x, \alpha y\rangle &= \overline{\langle \alpha y, x\rangle} \qquad \text{from} \quad (3) \\ &= \overline{\alpha\langle y, x\rangle} \qquad \text{from} \quad (2) \\ &= \overline{\alpha\overline{\langle x, y\rangle}} \\ &= \overline{\alpha}\langle x, y\rangle. \end{aligned}$$

□

Note 2.0.2. *The function* $\langle , \rangle$ *is linear in the first variable and conjugate linear in the second.* $\langle ., . \rangle$ *is linear in the first variable because*

1. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.

2. $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$.

$\langle , \rangle$ *is conjugate linear in the second because*

1. $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.

2. $\langle x, \alpha y\rangle = \overline{\alpha}\langle x, y\rangle$.

Definition 2.0.3. *A real vector space* X *is called an Inner Product Space if there is defined a real function* $\langle , \rangle$ *on* $X \times X$ *satisfying the properties* $1 - 4$*. Then the properties* $3, 6$ *are reduced to*

3. $\langle x, y\rangle = \langle y, x\rangle$.

6. $\langle x, \alpha y\rangle = \langle \alpha y, x\rangle = \alpha\langle x, y\rangle$.

Example 2.0.4. *The space* $\mathbb{C}^n$ *is an inner product space with the usual definition* $\langle x, y\rangle = \sum_{j=1}^{n} x_j\overline{y_j}$.

Example 2.0.5. *The space* $C[a, b]$ *is an inner product space if we define* $\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}dx$.

Definition 2.0.6. *Let* X *be a vector space over the field* $\mathbb{F}$*, where* $\mathbb{F}$ *is either the field* $\mathbb{R}$ *of real numbers or the field* $\mathbb{C}$ *of complex numbers. The norm* $\|.\|$ *on* X *is a function that assigns to each element of* X *a non negative real value and has the following properties :*

1. $\|x\| = 0$ *iff* $x = 0$.

2. $\|\alpha x\| = |\alpha|\|x\| \quad \forall \alpha \in \mathbb{F}, x \in X$.

3. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X$. *(is called the triangle inequality).*

Definition 2.0.7. *A vector space equipped with a norm is called a normed vector space (or a normed linear space)*

The norm induces a metric on X given by $d(x, y) = \|x - y\|$.

Definition 2.0.8. *A Banach Space is a complete normed linear space (complete in the metric induced by the norm).*

Theorem 2.0.9. *(Cauchy-Schwarz inequality)*
For $x, y \in X$*, we have*

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

Proof. If $y = 0$, $\|y\| = 0$ and $\langle 0, x\rangle = \langle 0\theta, x\rangle = 0\langle \theta, x\rangle$ where θ is the zero element of X and $\langle x, 0\rangle = \overline{\langle 0, x\rangle} = \bar{\theta} = \theta$ so both sides of the inequality vanish and the inequality is true.
So let us assume that $y \neq 0$ and take any scalar $\alpha \in \mathbb{C}$. Then we have

$$0 \leq \langle x + \alpha y, x + \alpha y\rangle$$

.

But

$$\begin{aligned}\langle x-\alpha y, x-\alpha y\rangle &= \langle x,x\rangle - \langle x,\alpha y\rangle - \langle \alpha y, x\rangle + \langle \alpha y, \alpha y\rangle \\ &= \langle x,x\rangle - \langle x,\alpha y\rangle - \langle \alpha y, x\rangle + \langle \alpha y, \alpha y\rangle \\ &= \langle x,x\rangle - \overline{\alpha}\langle x,y\rangle - \alpha\langle y,x\rangle + \alpha\overline{\alpha}\langle x,y\rangle \\ &= \|x\|^2 - \alpha\langle y,x\rangle - \overline{\alpha}\langle x,y\rangle + |\alpha|^2\|y\|^2.\end{aligned}$$

Hence we have $\|x\|^2 - \alpha\langle y,x\rangle - \overline{\alpha}\langle x,y\rangle + |\alpha|^2\|y\|^2 \geq 0$.
Since $y \neq 0, \|y\| \neq 0$. Choosing $\alpha = \dfrac{\langle x,y\rangle}{\|y\|^2}$ and using $\langle y,x\rangle = \overline{\langle x,y\rangle}$,
we get,

$$\|x\|^2 - \frac{\langle x,y\rangle\overline{\langle x,y\rangle}}{\|y\|^2} - \frac{\overline{\langle x,y\rangle}\langle x,y\rangle}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^4}\|y\|^2 \geq 0$$

which gives,

$$\|x\|^2 - \frac{|\langle x,y\rangle|^2}{\|y\|^2} - \frac{|\langle x,y\rangle|^2}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^2} \geq 0.$$

Hence we get $\|x\|^2 - \dfrac{|\langle x,y\rangle|^2}{\|y\|^2} \geq 0$ from which we have

$$|\langle x,y\rangle| \leq \|x\|\|y\|.$$

□

Theorem 2.0.10. *Given an inner product on X, put*

$$\|x\| := \sqrt{\langle x,x\rangle}.$$

Then $\|.\|$ defines a norm on X.

Proof. We will show $\|.\|$ satisfies the three properties of norm.

1.

$$\begin{aligned}\|x\| = 0 &\implies \sqrt{\langle x,x\rangle} = 0 \\ &\implies \langle x,x\rangle = 0 \\ &\implies x = 0.\end{aligned}$$

Conversly:

$$\begin{aligned}x = 0 &\implies \langle x,x\rangle = 0 \\ &\implies \sqrt{\langle x,x\rangle} = 0 \\ &\implies \|x\| = 0.\end{aligned}$$

2.

$$
\begin{aligned}
\|\alpha x\| &= \sqrt{\langle \alpha x, \alpha x\rangle} \\
&= \sqrt{\alpha\langle x, \alpha x\rangle} \\
&= \sqrt{\alpha\bar{\alpha}\langle x, x\rangle} \\
&= \sqrt{|\alpha|^2\langle x, x\rangle} \\
&= |\alpha|\|x\|.
\end{aligned}
$$

3. Let $x, y \in X$. Then we have,

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle.$$

Since ,

$$\langle y, x\rangle = \overline{\langle x, y\rangle} \text{ and } \langle x, y\rangle + \overline{\langle x, y\rangle} = 2\Re\langle x, y\rangle$$

We get ,

$$\|x+y\|^2 = \|x\|^2 + 2\Re\langle x, y\rangle + \|y\|^2. \tag{2.1}$$

Since

$$2\Re\langle x, y\rangle \leq 2|\langle x, y\rangle|, \tag{2.2}$$

using (2.1) in (2.2) and applying Schwarz inequality,we get

$$\|x+y\|^2 \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\|.$$

That is $\|x+y\|^2 \leq (\|x\| + \|y\|)^2$ which proves $\|x+y\| \leq \|x\| + \|y\|$.

□

Definition 2.0.11 (Hilbert Space)**.** *If the inner product space X with norm (induced by the inner product) is a complete metric space , we say it is a Hilbert Space. We use the symbol $\mathcal{H}$ for a Hilbert Space.*

An inner product space is sometimes called a pre Hilbert Space. If it is not complete, then its completion is a Hilbert Space.

Example 2.0.12. *The space l_2 is Hilbert Space with the inner product $\langle x, y\rangle = \sum_{j=1}^{\infty} x_j\overline{y_j}$.*

Example 2.0.13. *The space $L_2[a,b]$ is a Hilbert Space with the inner product $\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}dx$.*

Theorem 2.0.14 (Parallelogram Law)**.** *If x and y are two vectors in a Hilbert Space, then*

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

Proof. For any $x, y \in X$, we get

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle \\ &= \langle x, x+y\rangle + \langle y, x+y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\ &= \|x\|^2 + \langle x,y\rangle + \langle y,x\rangle + \|x\|^2.\end{aligned}$$

Also,

$$\begin{aligned}\|x-y\|^2 &= \langle x-y, x-y\rangle \\ &= \langle x, x-y\rangle - \langle y, x-y\rangle \\ &= \langle x,x\rangle - \langle x,y\rangle - \langle y,x\rangle + \langle y,y\rangle \\ &= \|x\|^2 - \langle x,y\rangle - \langle y,x\rangle + \|x\|^2.\end{aligned}$$

Therefore we get

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

. □

Note 2.0.15. *In a Hilbert Space, the norm induced by the inner product satisfies the parallelogram law. This is not true in general in a Banach Space. That is, the norm in a Banach Space need not necessarily satisfy the parallelogram law as illustrated by the following example.*

Example 2.0.16. ***In the Banach Space*** $l_1{}^n$ ***where*** $n \geq 1$ ***the parallelogram law is not true.*** *Consider the set* $e_1, e_2, ...e_n$ *in* $l_1{}^n$ *where*

$$e_1 = (1,0,0,...,0), e_2 = (0,1,0,...,0) \text{ and } e_1 = (0,0,0,...,1).$$

We know that $l_1{}^n$ *in a Banach Space with the norm*

$$\|x\| = \sum_{i=1}^{n} |x_i|, \text{ if } x = (x_1, x_2, ..., x_n) \in l^n{}_1.$$

Let us take $x = e_1$ *and* $y = e_2$. *Now* $x + y = e_1 + e_2 = (1,1,0,0,....0)$ *and* $x - y = e_1 - e_2 = (1,-1,0,0,....0)$.
So $\|x\| = 1, \|y\| = 1, \|x+y\| = 2$ *and* $\|x-y\| = 2$. *Hence,* $\|x+y\|^2 + \|x-y\|^2 = 8$ *and* $2\|x\|^2 + 2\|y\|^2 = 2 + 2 = 4$.
Thus $\|x+y\|^2 + \|x-y\|^2 \neq 2(\|x\|^2 + \|y\|^2)$.

Definition 2.0.17. *Let* $\mathcal{H}$ *be the given Hilbert Space. Then* $x, y \in \mathcal{H}$ *are said to be orthogonal denoted by* $x \perp y$ *if* $\langle x,y\rangle = 0$.

From the definition we have the following consequences.

1. The relation of orthogonality is symmetric. That is $x \perp y$ implies $y \perp x$.
   $x \perp y$ gives $\langle x,y\rangle = 0$ which gives $\overline{\langle x,y\rangle} = 0$. Now $\overline{\langle x,y\rangle} = \langle y,x\rangle = 0$ so that $y \perp x = 0$.

2. If $x \perp y$, then $\alpha x \perp y$ for every scalar $\alpha \in \mathbb{C}$. Thus $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle = 0$ so that $x \perp y$ implies $\alpha x \perp y$ for any scalar.

3. Since $\langle 0, x\rangle = 0$ for any $x \in \mathcal{H}$, $0 \perp x\ \forall x \in \mathcal{H}$.

4. If $x \perp x$, then x must be zero. For $x \perp x$, then $\langle x, x\rangle = 0$ which implies $\langle x, x\rangle = \|x\|^2 = 0$. That is $x = 0$.

Theorem 2.0.18. *(Pythagorean Theorem)If x and y are two orthogonal vectors in a Hilbert Space $\mathcal{H}$, then*

$$\|x + y\|^2 = \|x - y\|^2 = \|x\|^2 + \|y\|^2.$$

Proof. Since $x \perp y$, we have $\langle x, y\rangle = 0$. Now,
$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2$ by the hypothesis and
$\|x - y\|^2 = \langle x - y, x - y\rangle = \langle x, x\rangle - \langle x, y\rangle - \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2$.
Hence

$$\|x + y\|^2 = \|x - y\|^2 = \|x\|^2 + \|y\|^2.$$

□

Lemma 2.0.19. *(Polarization Identity): If x,y are any two vectors in a Hilbert Space $\mathcal{H}$, then*

$$4\langle x, y\rangle = \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2.$$

Proof.

$$\|x + y\|^2 = \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2. \tag{2.3}$$

$$\|x - y\|^2 = \|x\|^2 - \langle x, y\rangle - \langle y, x\rangle + \|y\|^2. \tag{2.4}$$

Subtracting (2.4) from (2.3) ,we get

$$\|x + y\|^2 - \|x - y\|^2 = 2\langle x, y\rangle + 2\langle y, x\rangle. \tag{2.5}$$

Replacing y by iy in (2.5), we get

$$\|x + iy\|^2 - \|x - iy\|^2 = 2\langle x, iy\rangle + 2\langle iy, x\rangle = 2\bar{i}\langle x, y\rangle + 2i\langle y, x\rangle$$

$$\implies \|x + iy\|^2 - \|x - iy\|^2 = -2i\langle x, y\rangle + 2i\langle y, x\rangle.$$

Multiplying both sides of by i, we get

$$i\|x + iy\|^2 - i\|x - iy\|^2 = 2\langle x, y\rangle - 2\langle y, x\rangle. \tag{2.6}$$

Adding (2.4) and (2.5), we get

$$\|x + y\|^2 - \|x - y\|^2 - i\|x + iy\|^2 - i\|x - iy\|^2 = 4\langle x, y\rangle.$$

which proves the polarization identity. □

Theorem 2.0.20. *Let $\|.\|$ be norm on a Banach space. Then there exists an inner product $\langle , \rangle$ on B such that $\langle x, x\rangle = \|x\|^2$ for all $x \in B$ if and only if the norm satisfies the parallelogram law $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$. In this case such a inner product is unique and it is given by polarisation identity given in Lemma above.*

Note 2.0.21. *The above theorem asserts that not all Banach Spaces are Hilbert Spaces. If a norm does not satisfy the parallelogram law,it cannot be a Hilbert Space. The following example illustrates this point.*

Example 2.0.22. *The space l_p with $p \neq 2$ is not an inner product space and hence it is not a Hilbert Space.*
Let $x = (1, 1, 0, 0, ...)$ and $y = (1, -1, 0, 0, ...)$. Then we note that $x, y \in l_p$.
Now, $\|x\| = \|y\| = 2^{\frac{1}{p}}, \|x+y\| = (2^p + 0^p + ...)^{\frac{1}{p}}$ and $\|x-y\| = (0^p + 2^p + ...)^{\frac{1}{p}}$ so that $\|x+y\| = \|x-y\| = 2$. Hence the parallelogram law is not satisfied for $p \neq 2$. Hence the complete space l_p for $p \neq 2$ is a Banach Space which is not a Hilbert Space.

In the case of a real inner product space the polarisation identity reduces to

$$4\langle x, y\rangle = \|x+y\|^2 + \|x-y\|^2$$

Theorem 2.0.23. *(Appolonius Theorem): For x, y, z in an inner product space, we have*

$$\|x-y\|^2 + \|x-z\|^2 = 2(\|x - \frac{1}{2}(y+z)\|^2 + \|\frac{1}{2}(y-z)\|^2.$$

It generalises the theorem with this name in plane geometry: if ABC is a triangle, and D is the mid-point of the side BC, then

$$(AB)^2 + (AC)^2 = 2[(AD)^2 + (BD)^2$$

Theorem 2.0.24. *If N is a normed linear space then the norm $\|.\| : N \to \mathbb{R}$ is continuous function on $\mathbb{R}$.*

Proof. We shall show that if $x_n \to x$ implies, then $\|x_n\| \to \|x\|$.
Now,

$$|\|x\| - \|y\|| \leq \|x-y\| \quad \forall x, y \in N. \tag{2.7}$$

Let $\epsilon > 0$, since $x_n \to x$ by definition of convergence of sequences in normed linear space.

$$\text{there exists } n_0 \in N \text{ such that } \|x_n - x\| < \epsilon \quad \forall n \geq n_0. \tag{2.8}$$

Using (2.7) and (2.8), we get $|\|x_n\|| - |\|x\|| < \epsilon$ for all $n \geq n_0$ so that $\|x_n\| \to \|x\|$.
This proves that the norm is a continuous function on N. □

Theorem 2.0.25. *The inner product in a Hilbert Space is jointly continuous,that is if $x_n \to x$ and $y_n \to y \implies \langle x_n, y_n\rangle \to \langle x, y\rangle$ as $n \to \infty$.*

Proof.

$$\begin{aligned}|\langle x_n, y_n\rangle - \langle x, y\rangle| &= |\langle x_n, y_n\rangle - \langle x_n, y\rangle + \langle x_n, y\rangle - \langle x, y\rangle| \\ &= |\langle x_n, y_n - y\rangle + \langle x_n - x, y\rangle| \\ &\leq |\langle x_n, y_n - y\rangle| + |\langle x_n - x, y\rangle|.\end{aligned}$$

We get,

$$|\langle x_n, y_n\rangle - \langle x, y\rangle| \leq |\langle x_n, y_n - y\rangle| + |\langle x_n - x, y\rangle|. \tag{2.9}$$

By the Schwarz inequality,

$$|\langle x_n, y_n - y\rangle| \leq \|x_n\|\|y_n - y\|, \tag{2.10}$$

and

$$|\langle x_n - x, y\rangle| \leq \|x_n - x\|\|y\|. \tag{2.11}$$

Using (2.10) and (2.11) in (2.9), we get

$$|\langle x_n, y_n\rangle - \langle x, y\rangle| \leq \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\|.$$

Since $x_n \to x$ and $y_n \to y$, $\|x_n - x\| \to 0$ and $\|y_n - y\| \to 0$. Further since $\langle x_n\rangle$ is a convergent sequence, it is bounded so that $\|x_n\| \leq M$ for all n. □

Theorem 2.0.26. *Let S be any subset of $\mathcal{H}$. Let*

$$S^{\perp} = \{x \in \mathcal{H} : x \perp y \;\forall\; y \in S\}.$$

Then

1. $S \bigcap S^{\perp} \subset 0$.
2. *$S^{\perp}$ is a closed subspace of $\mathcal{H}$.*
3. $0^{\perp} = \mathcal{H}, \mathcal{H}^{\perp} = \{0\}$.
4. *If $S_1 \subset S_2$, then $S_2^{\perp} \subset S_1^{\perp}$.*
5. $S \subset S^{\perp\perp}$.

Proof. 1. Let $x \in S \bigcap S^{\perp} \implies x \in S$ and $x \in S^{\perp}$.Then in particular, $\|x\|^2 = \langle x, x\rangle = 0$ so $x = 0$.

2. Since $\langle 0, y\rangle = 0 \;\forall\; y \in S$, then $0 \in S^{\perp}$. Hence $S^{\perp} \neq \emptyset$. Let $x_1, x_2 \in S^{\perp}$ and $\alpha, \beta \in \mathbb{C}$. Then $\langle x_1, y\rangle = 0$ and $\langle x_2, y\rangle = 0$ for every $y \in S$. Hence $\forall y \in S$, we get

$$\langle \alpha x_1 + \beta x_2, y\rangle = \alpha\langle x_1, y\rangle + \beta\langle x_2, y\rangle = \alpha.0 + \beta.0$$

This proves that $\alpha x_1 + \beta x_2 \in S^{\perp}$. Hence $S^{\perp}$ is a subspace of $\mathcal{H}$.
Now to show $S^{\perp}$ is closed. Let $(x_n) \in S^{\perp}$ and $(x_n) \to x$ in $\mathcal{H}$. Then we have to show that $x \in S^{\perp}$ i.e we need to show $\langle x, y\rangle = 0 \forall y \in S$. Since $x_n \in S^{\perp}, \langle x_n, y\rangle = 0 \;\forall y \;\in S \;\forall\; n \in \mathbb{N}$.

Since the inner product is a continuous function, we get

$$\langle x_n, y\rangle \to \langle x, y\rangle \text{ as } n \to \infty.$$

Since $\langle x_n, y\rangle = 0 \; \forall n, \langle x, y\rangle = 0$. Thus $x \in S^{\perp}$. Hence $S^{\perp}$ is a closed subset of $\mathcal{H}$. Now $S^{\perp}$ is a closed subspace of complete space implies $S^{\perp}$ is complete.

3. Let $x \in \mathcal{H} \implies \langle x, y\rangle = 0 \; \forall \; y \in \mathcal{H}$.In particular, $\langle x, x\rangle = 0$.That is $x = 0$. Let $x \in \{0\}^{\perp} \implies \langle x, y\rangle = 0 \; \forall y \in \{0\} \implies \langle x, 0\rangle = 0 \implies x \in \mathcal{H}$.

4. Let $x \in S_2^{\perp} \implies \langle x, y\rangle = 0 \; \forall \; y \in S_2$. Since $S_1 \subset S_2$ we get $\langle x, y\rangle = 0 \; \forall \; y \in S_1 \implies x \in S_1^{\perp}$.

5. Let $x \in S \implies \langle x, y\rangle = 0 \; \forall \; y \in S^{\perp}$. If $y \in S^{\perp} \implies x \in S^{\perp\perp}$(by definition). Thus $x \in S \implies x \in S^{\perp\perp} \implies S \subset S^{\perp\perp}$.

□

Theorem 2.0.27. *Let S be any closed convex subset of $\mathcal{H}$. Then for each x in $\mathcal{H}$ there existsa unique point x_0 in S such that*

$$\|x - x_0\| = dist(x, S) := \inf_{y \in S} \|x - y\|$$

.

Proof. Let $d = dist(x, S)$

a. **Existence Claim:** There exists a sequence y_n in S such that $\|x - y_n\| \to d$. Now $d = dist(x, S) \implies d = \inf_{y \in S} \|x - y\|$(by definition of $dist$). Let $\epsilon > 0$ be given . Take $\epsilon = \frac{1}{n}$, now by definition of infimum there exists$y_n \in S$ such that $\|x - y_n\| \le d + \frac{1}{n}$. Already it is given that $d = \inf_{y \in S} \|x - y\|$ and therefore since $y_n \in S$ we have $d \le \|x - y_n\|$.
From the above the two statements we have, $\forall n \in \mathbb{N}$, there exists $y_n \in S$ such that $d \le \|x - y_n\| \le d + \frac{1}{n} \implies \|x - y_n\| \to d$.
Now by using Appolonius theorem we get:

$$\|x - y_n\|^2 + \|x - y_m\|^2 = 2(\|x - \frac{1}{2}(y_n + y_m)\|^2 + \|\frac{1}{2}(y_n - y_m)\|^2).$$

Now S is a convex set and hence $\frac{1}{2}(y_n + y_m) \in S$. This implies that $\|x - \frac{1}{2}(y_n + y_m)\| \ge d^2$ $\implies \|x - y_n\|^2 + \|x - y_m\|^2 \ge 2d^2 + \frac{1}{2}\|y_n - y_m\|^2$.
Now as $n, m \to \infty$,$\|x - y_n\|^2 + \|x - y_m\|^2 \to 2d^2 \implies \frac{1}{2}\|y_n - y_m\|^2 \to 0$. This implies $\langle y_n\rangle$ is a Cauchy sequence and since $\mathcal{H}$ is a Hilbert Space the sequence converges to a limit.
Since S is closed $x_0 := \lim y_n$ is in S and $\|x - x_0\| = \lim \|x - y_n\| = d$.

b. **Uniqueness**
We assume that if there exists another point x_1 in S for which $\|x - x_1\| = d$, then we show $x_1 = x_0$.

By Parallelogram equality,

$$\begin{aligned}\|x_0 - x_1\|^2 &= \|(x_0 - x) - (x_1 - x)\|^2 \\ &= 2\|x_0 - x\|^2 + 2\|x_1 - x\|^2 - \|(x_0 - x) + (x_1 - x)\|^2 \\ &= 2d^2 + 2d^2 - 2^2\|\frac{1}{2}(x_0 + x_1) - x\|^2.\end{aligned}$$

As S is a convex set, $\frac{1}{2}(x_0 + x_1) \in S$, so we have

$$\|\frac{1}{2}(x_0 + x_1) - x\| \geq d.$$

This implies the right hand side is less than or equal to $2d^2 + 2d^2 - 4d^2 = 0$. Hence we have the inequality $\|x_0 - x_1\| \leq 0$. Clearly, $\|x_0 - x_1\| \geq 0$, so we must have the equality, and $x_0 = x_1$.

□

An interesting is the case when S is a closed linear subspace.
For each x in $\mathcal{H}$ let

$$P_S(x) = x_0.$$

where x_0 is the unique point in S closest to x. Then P_S is a well defined map with range S.
Claim: P_S is idempotent i.e $P_S^2 = P_S$. $x \in \mathcal{H} \implies P_S(x) = x_0$, where $x_0 \in S$. Then $P_S^2(x) = P_S(x_0) = x_0 = P_S(x_0)$.

For each y in S and $t \in R$, we have

$$\|x - (x_0 + ty)\|^2 \geq \|x - x_0\|^2 \text{ (by best approximant property).}$$

From this we get,

$$\|x - x_0\|^2 + t^2\|y\|^2 - 2t\Re\langle x - x_0, y\rangle \quad \geq \|x - x_0\|^2.$$

That is

$$t^2\|y\|^2 \geq 2t\Re\langle x - x_0, y\rangle.$$

Since this is true for all real t we must have

$$\Re\langle x - x_0, y\rangle = 0.$$

Replacing y by ιy,we get

$$\Im\langle x - x_0, y\rangle = 0.$$

Hence $\langle x - x_0, y\rangle = 0 \implies x - x_0 \in S^\perp$. Now, since $S \bigcap S^\perp = \{0\}$ and $x = x - x_0 + x_0$ where $x - x_0 \in S^\perp$ and $x_0 \in S$. We have by definition of direct sum decomposition,

$$\mathcal{H} = S \bigoplus S^\perp.$$

Theorem 2.0.28 (Riesz's Theorem)**.** *Every bounded linear functional f on a Hilbert Space $\mathcal{H}$ can* *be represented in terms of the inner product, namely*

$$f(x) = \langle x, y \rangle. \tag{2.12}$$

where y depends on f, is uniquely determined by f and has norm

$$\|y\| = \|f\|. \tag{2.13}$$

Proof. We prove the theorem in the following three steps.

1. **In this step we show that any $f \in \mathcal{H}^*$ has the representation as in** (2.12).
   If $f = 0$, we take $y = 0$ so that (2.12) and (2.13) are true. So let us take $f \neq 0$. Let us consider the null space $N(f)$ of f. Since f is continuous we know that $N(f)$ is a proper closed subspace and since $f \neq 0, N(f) \neq \mathcal{H}$ and so $N(f)^{\perp} \neq 0$. Hence by the orthogonal decomposition theorem proved above, there exists a $y_0 \neq 0$ in $N(f)^{\perp}$. Let us define for any arbitrary $x \in \mathcal{H}$,
   $$z = f(x)y_0 - f(y_0)x.$$
   Now, $f(z) = f(x)f(y_0) - f(y_0)f(x) = 0$. This implies $z \in N(f)$. Since $y_0 \in N(f)^{\perp}$, we get
   $$0 = \langle z, y_0 \rangle = \langle f(x)y_0 - f(y_0)x, y_0 \rangle = f(x)\langle y_0, y_0 \rangle - f(y_0)\langle x, y_0 \rangle$$
   $$\Longrightarrow \qquad f(x)\langle y, y_0 \rangle - f(y_0)\langle x, y_0 \rangle = 0.$$
   Noting that $\langle y_0, y_0 \rangle = \|y_0\|^2 \neq 0$, we get $f(x) = (\frac{f(y_0)}{\|y_0\|^2})\langle x, y_0 \rangle$. We use property 6 of inner product spaces and write $f(x) = \langle x, \overline{\frac{f(y_0)}{\|y_0\|^2}} y_0 \rangle$.
   Now taking $\overline{\frac{f(y_0)}{\|y_0\|^2}} y_0$ as y, we have established that there exists a y such that $f(x) = \langle x, y \rangle$ for any $x \in \mathcal{H}$.

2. **In this step we show that $\|f\| = \|y\|$.**
   If $f = 0$, then $y = 0$ and (2.13) holds good. Hence we let $f \neq 0$. Then $y \neq 0$.
   From (2.12) and the Schwarz inequality we have
   $$|f(x)| = |\langle x, y \rangle| \leq \|x\|\|y\|.$$
   This implies $\sup_{x \neq 0} \frac{|f(x)|}{\|x\|} \leq \|y\|$.
   Using the definition of norm of f, we get from the above,
   $$\|f\| \leq \|y\|. \tag{2.14}$$
   To prove the reverse inequality, let us take $x = y$ in (2.12), then we obtain
   $$\|y\|^2 = \langle y, y \rangle = f(y) \leq \|f\|\|y\|.$$

Since $y \neq 0$, we get

$$\|f\| \geq \|y\|. \tag{2.15}$$

From (2.14) and (2.15) ,we get $\|f\| = \|y\|$ which proves (2.13).

3. **In this step we establish the uniqueness of y in** (2.12).
   Let us suppose that y is not unique in the representation (2.12) Suppose for all $x \in \mathcal{H}$, there exists y_1 and y_2 such that $f(x) = \langle x, y_1 \rangle = \langle x, y_2 \rangle$. Then $\langle x, y_1 \rangle - \langle x, y_2 \rangle = 0$ which implies $\langle x, y_1 - y_2 \rangle = 0$ for all $x \in \mathcal{H}$. Let us choose x to be $(y_1 - y_2)$ so that

$$\langle y_1 - y_2, y_1 - y_2 \rangle = \|y_1 - y_2\|^2 = 0.$$

   Hence $y_1 - y_2 = 0$ so that $y_1 = y_2$ which proves that y is unique in the representation for f in (2.12). This completes the proof of Riesz Representation theorem for continuous linear functionals on $\mathcal{H}$.

□

Lemma 2.0.29. *If $\langle v_1, w \rangle = \langle v_2, w \rangle$ for all w in an inner product space X, then $v_1 = v_2$. In particular, $\langle v, w \rangle = 0 \quad \forall w \in X$ implies $v_1 = 0$.*

Proof. By assumption, for all w,

$$\langle v_1 - v_2, w \rangle = \langle v_1, w \rangle - \langle v_2, w \rangle = 0.$$

For $w = v_1 - v_2$ this gives $\|v_1 - v_2\| = 0$. Hence $v_1 - v_2 = 0$, so that $v_1 = v_2$. In particular, $\langle v_1, w \rangle = 0$ with $w = v_1$ gives $\|v_1\| = 0$, so that $v_1 = 0$. □

Definition 2.0.30 (Sesquilinear form)**.** *Let X and Y be vector spaces over the same field $\mathbb{K}(\mathbb{R}$ or $\mathbb{C})$. Then a sesquilinear form (or sesquilinear functional) h on $X \times Y$ is a mapping*

$$h : X \times Y \to \mathbb{K}$$

such that $\forall x, x_1, x_2 \in X$ and $y, y_1, y_2 \in Y$ and all scalars α, β,

(a) $h(x_1 + x_2, y) = h(x_1, y) + h(x_2, y)$.

(b) $h(x, y_1 + y_2) = h(x, y_1) + h(x, y_2)$.

(c) $h(\alpha x, y) = \alpha h(x, y)$.

(d) $h(x, \beta y) = \bar{\beta} h(x, y)$.

Hence h is linear in the first argument and conjugate linear in the second one. If X and Y are real ($\mathbb{K} = \mathbb{R}$), then (d) is simply

$$h(x, \beta y) = \beta h(x, y).$$

and h is bilinear since it is linear in both arguments.
If X and Y are normed spaces and if there is a real number c such that $\forall x, y$

$$|h(x, y)| \leq \|x\| \|y\|.$$

then h is said to be bounded and the number

$$\|h\| = \sup_{x \in X - \{0\}, y \in Y - \{0\}} \frac{|h(x,y)|}{\|x\|\|y\|} = \sup_{\|x\|=1, \|y\|=1} |h(x,y)|.$$

is called the norm of h.

Example 2.0.31. *The inner product is sesquilinear and bounded.*

Theorem 2.0.32. *Riesz Representation Theorem*
Let $\mathcal{H}_1, \mathcal{H}_2$ be Hilbert Spaces and $h : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{K}$ a bounded sesquilinear form. Then h has a representation $h(x,y) = \langle Sx, y\rangle$ where $S : \mathcal{H}_1 \to \mathcal{H}_2$ is a bounded linear operator. S is uniquely determined by h and has norm $\|S\| = \|h\|$.

Proof. We consider $\overline{h(x,y)}$. This is linear in y, because of the bar. To make the previous theorem applicable, we keep x fixed. Then theorem yields a representation in which y is variable, say,

$$\overline{h(x,y)} = \langle y, z\rangle.$$

Hence,

$$h(x,y) = \langle z, y\rangle.$$

Here $z \in H_2$ is unique but, of course, depends on our fixed $x \in H_1$. Consider the operator,

$$S : \mathcal{H}_1 \to \mathcal{H}_2 \text{ given by } z = Sx$$

Claim: S is linear

$$\begin{aligned} \langle S(\alpha x_1 + \beta x_2), y\rangle &= h(\alpha x_1 + \beta x_2, y) \\ &= \alpha h(x_1, y) + \beta h(x_2, y) \\ &= \alpha \langle Sx_1, y\rangle + \beta \langle Sx_2, y\rangle \\ &= \langle \alpha Sx_1 + \beta Sx_2, y\rangle. \end{aligned}$$

for all y in H_2, so that by above lemma $S(\alpha x_1 + \beta x_2) = \alpha Sx_1 + \beta Sx_2$.
Claim: S is bounded. Indeed, leaving aside the trivial case $S = 0$, we have

$$\|h\| = \sup_{x \neq 0, y \neq 0} \frac{|\langle Sx, y\rangle|}{\|x\|\|y\|} \geq \sup_{x \neq 0, Sx \neq 0} \frac{|\langle Sx, Sx\rangle|}{\|x\|\|Sx\|} = \sup_{x \neq 0} \frac{\|Sx\|}{\|x\|} = \|S\|.$$

This proves boundedness. Moreover, $\|h\| \geq \|S\|$. **Claim:** $\|S\| = \|h\|$. Now,

$$\|h\| = \sup_{x \neq 0, y \neq 0} \frac{|\langle Sx, y\rangle|}{\|x\|\|y\|} \leq \sup_{x \neq 0} \frac{\|Sx\|\|y\|}{\|x\|\|y\|} = \|S\| \text{ (using the Schwarz inequality)}$$

We have already proved that $\|h\| \geq \|S\|$ and therefore we get

$$\|S\| = \|h\|.$$

Claim: S is unique.

Assume there exists a linear operator $T : \mathcal{H}_1 \to \mathcal{H}_2$ such that for all $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$ we have

$$h(x, y) = \langle Sx, y \rangle = \langle Tx, y \rangle,$$

we see that $Sx = Tx$ by the above Lemma for all $x \in \mathcal{H}_1$. Hence $S = T$ by definition.

$\square$

Definition 2.0.33. *Let $T : \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator, where $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert Spaces. Then the Hilbert Adjoint operator T^* of T is the operator $T^* : \mathcal{H}_1 \to \mathcal{H}_2$ such that for all $x \in \mathcal{H}_1$,$y \in \mathcal{H}_2$ we have $\langle Tx, y \rangle = \langle x, T^* y \rangle$*

Definition 2.0.34. *An isomorphism T of an inner product space X onto an inner product space $\tilde{X}$ over the same field is a bijective linear operator $T : X \to \tilde{X}$ which preserves the inner product, that is for all $x, y \in X$ $\langle Tx, Ty \rangle = \langle x, y \rangle$,where we denoted inner product on X and $\tilde{X}$ by the same symbol for simplicity. $\tilde{X}$ is then called isomorphic with X. $\tilde{X}$ and X are called isomorphic inner product spaces.*

Note that bijectivity and linearity guarantees that T is a vector space isomorphism of X onto $\tilde{X}$, so that T preserves the whole structure of an inner product space. T is also an isometry of X onto $\tilde{X}$ because distances in X and $\tilde{X}$ are determined by the norms defined by the inner product on X and $\tilde{X}$. Since Hilbert Spaces are also inner product spaces the same definition will always work for Hilbert Spaces.

Theorem 2.0.35. *Every operator A from X to Y gives rise, in a natural way to an operator A^* from the dual space Y^* to X^* and $\|A\| = \|A^*\|$.*

Proof. Let A be an operator from X to Y. For $f \in Y$. For $f \in Y^*$ let

$$(A^* f)(x) = f(Ax)$$

for all $x \in X$. Then $A * f$ is bounded linear functional on X, i.e, $A^* f \in X^*$. The equation is some times written as

$$\langle A^* f, x \rangle = \langle f, Ax \rangle, x \in X, f \in Y^*.$$

A^* is called the adjoint of A.

If $f \in Y^*$, and $\|f\| = 1$, then

$$\|A^* f\| = \sup_{\|x\|=1} |(A^* f)(x)| = \sup_{\|x\|=1} |f(Ax)| \leq \sup_{\|x\|=1} \|Ax\| = \|A\|.$$

Thus $\|A^*\| \leq \|A\|$, and A^* is a bounded linear operator from Y^* to X^*. In fact, $\|A^*\| = \|A\|$ To prove this we need to show $\|A\| \leq \|A^*\|$. Let x be any element of X and by the Hahn-Banach Theorem, there exists a linear functional f on Y such that $\|f\| = 1$ and $f(Ax) = \|Ax\|$. Thus

$$\|Ax\| = f(Ax) = (A^* f)(x) \leq \|A^*\| \|f\| \|x\| = \|A^*\| \|x\|$$

This shows that $\|A\| \leq \|A^*\|$

. $\square$

Some properties:

1. Let $A, B \in \mathbb{B}(X, Y)$. Then
$$(\alpha A + \beta B)^* = \alpha A^* + \beta B^*$$
for all $\alpha, \beta \in \mathbb{C}$. From this we conclude that the map $A \mapsto A^*$ from $\mathcal{B}(X, Y)$ to $\mathcal{B}(Y^*, X^*)$ is linear.

2. Let $A \in \mathbb{B}(X, Y), B \in \mathbb{B}(Y, Z)$. Then
$$(BA)^* = A^* B^*.$$

3. The adjoint of the identity of operator on X is the identity operator on X^*, i.e,
$$I^* = I.$$

4. If A is an invertible operator from X to Y then A^* is an invertible operator from Y^* to X^*, and $(A^*)^{-1} = (A^{-1})^*$.

The map $A \mapsto A^*$ is an isometry. In general, it not surjective.

Note: Let $\mathbb{H}$ be a Hilbert Space. Now $\mathcal{H}^*$ is isomorphic to $\mathbb{H}^*$ via a conjugate map linear map R that associates to $y \in \mathbb{H}$. The linear functional f_y defined as $f_y(x) = \langle x, y \rangle$ for all $x \in \mathbb{H}$ for every $A \in \mathbb{B}(\mathbb{H})$, its adjoint A^* can be identified with an operator on $\mathbb{H}$. Call this operator $A^\dagger$, so we have $A^\dagger = R^{-1} A^* R$.

If $A^* f_y = f_z$, then $A^\dagger y = z$. We have
$$\langle Ax, y \rangle = f_y(Ax) = (A^* f_y)(x) = f_z(x) = \langle x, z \rangle = \langle x, A^\dagger y \rangle$$
for all x, y. Thus
$$\langle Ax, y \rangle = \langle x, A^\dagger y \rangle$$
for all $x, y \in \mathbb{H}$. This equation determines $A^\dagger$ uniquely, i.e if there is another linear operator B on $\mathbb{H}$ such that
$$\langle Ax, y \rangle = \langle x, By \rangle \text{ for all } x, y,$$
then $B = A^\dagger$. It is customary to call this operator $A^\dagger$ the adjoint of A. We will do so too and use the symbol A^* for this operator. Thus A^* is the unique operator associated with A by the condition
$$\langle Ax, y \rangle = \langle x, A^* y \rangle \text{for all } x, y \in \mathbb{H}.$$
This correspondence $A \mapsto A^*$ is conjugate linear.

Definition 2.0.36. *If $\mathbb{H}, \mathbb{K}$ are Hilbert Spaces and A is linear operator from $\mathbb{H}$ to $\mathbb{K}$, then A^* is a linear operator from $\mathbb{K}$ to $\mathbb{H}$ defined by*

$$\langle Ax, y\rangle = \langle x, A^*y\rangle \text{ for all } x, y \in \mathbb{H}$$

with $x \in \mathbb{H}, y \in \mathbb{K}$.

Theorem 2.0.37. *The map $A \mapsto A^*$ on $\mathbb{B}(\mathbb{H})$ has the following properties:*

1. *it is conjugate linear*
2. *it is isometric, $\|A^*\| = \|A\|$ for all A.*
3. *it is surjective.*
4. *$A^{**} = A$ for all A.*
5. *$(AB)^* = B^*A^*$ for all A, B.*
6. *$I^* = I$.*
7. *If A is invertible, then so is A^* and $(A^*)^{-1} = (A^{-1})^*$.*

Self-adjoint operators

An operator A on $\mathbb{H}$ is said to be self-adjoint or Hermitian, if $A = A^*$.
If A is self-adjoint, then for all $x \in \mathcal{H}$

$$\langle Ax, x\rangle = \langle x, Ax\rangle = \overline{\langle Ax, x\rangle}.$$

So, $\langle Ax, x\rangle$ is real. Conversely if $\mathbb{H}$ is a complex Hilbert Space and $\langle Ax, x\rangle$ is real for all x, then A is self-adjoint.
For every operator A on $\mathbb{H}$, we have

$$\sup_{\|y\|=1} |\langle Ax, y\rangle| = \|Ax\|,$$

and hence,

$$\sup_{\|x\|=1, \|y\|=1} |\langle Ax, y\rangle| = \sup_{\|x\|=1} \|Ax\| = \|A\|.$$

If A is self-adjoint, then

$$\|A\| = \sup_{\|x\|=1} |\langle Ax, x\rangle|.$$

Definition 2.0.38. *A linear map P on $\mathbb{H}$ is called a projection if it is idempotent. If $S =$ ran P and $S' =$ ker P, then $\mathbb{H} = S + S'$, and P is the projection on S along S'. The operator $I - P$ is also a projection, its range is S' and kernel S. A special property characterizes orthogonal projections: those for which $S' = S^\perp$.*

Proposition 2.0.39. *An idempotent operator P on $\mathbb{H}$ is an orthogonal projection if and only if it is self-adjoint.*

Proof. Let $x \in S, y \in S'$, then $Px = x, Py = 0$ if $P^* = P$, we have $\langle x, y\rangle = \langle Px, y\rangle = \langle x, Py\rangle = 0$. This shows $S' = S^{\perp}$. Conversely let z be any vector in $\mathbb{H}$, and split as $z = x + y$ with $x \in S, y \in S^{\perp}$. Let $Pz = x$. Then for any two vectors z_1, z_2

$$\begin{aligned}\langle Pz_1, z_2\rangle =&\langle x_1, x_2 + y_2\rangle\\ =&\langle x_1, x_2\rangle\\ =&\langle x_1 + y_1, x_2\rangle = \langle z_1, Pz_2\rangle.\end{aligned}$$

This shows $P^* = P$. □

When we talk about Hilbert Spaces we usually mean an orthogonal projection when we say a projection. To each closed linear subspace S in $\mathbb{H}$ there corresponds a unique(orthogonal) projection P and vice versa.

Theorem 2.0.40. *Prove* $\mathcal{R}(A) = \mathcal{N}(A^*)^{\perp}$.

Proof. A matrix A for which A and A^* commutes is called normal. A basic fact about normal matrices is that for all $v \in \mathbb{C}^n$ we have $\|Av\| = \|A^*v\|$. The reason is that

$$\|Av\|^2 = \langle Av, Av\rangle = \langle v, A^*Av\rangle = \langle A^*v, A^*v\rangle = \|A^*v\|^2.$$

In particular, this implies that $Av = 0$ if and only if $A^*v = 0$ so A and A^* has the same kernel. Then

$$\mathcal{R}(A) = \mathcal{N}(A^*)^{\perp} = \mathcal{N}(A)^{\perp} = \mathcal{R}(A^*)$$

so A, A^* also have the same image. □

Chapter 3

Theory of Reproducing Kernel Hilbert Spaces

3.1 Definition

We will consider Hilbert spaces over the field of either real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$. We will use $\mathbb{F}$ to denote either $\mathbb{R}$ or $\mathbb{C}$, so that when we wish to state a definition or a result that is true for either real or complex numbers, we will use $\mathbb{F}$.
Let X be a set. We denote by $\mathcal{F}(\mathcal{X}, \mathbb{F})$ the set of functions from X to $\mathbb{F}$. The set $\mathcal{F}(\mathcal{X}, \mathbb{F})$ is a vector space over the field $\mathbb{F}$ with operations of addition, $(f+g)(x) = f(x)+g(x)$ and scalar multiplication $(\lambda.f)(x) = \lambda.(f(x))$.

Definition 3.1.1. *Given a set X, we will say $\mathcal{H} \subseteq \mathcal{F}(\mathcal{X}, \mathbb{F})$ is a Reproducing Kernel Hilbert Space (RKHS) on X over F, provided that:*

1. *$\mathcal{H}$ is a vector space of $\mathcal{F}(\mathcal{X}, \mathbb{F})$;*

2. *$\mathcal{H}$ is endowed with an inner product $\langle , \rangle$ making it into a Hilbert Space;*

3. *$\forall y \in X$, the linear evaluation functional, $E_y : \mathcal{H} \to \mathbb{F}$, defined by $E_y(f) = f(y)$, is bounded.*

If $\mathcal{H}$ is an RKHS on X, then an application of the Riesz representation theorem shows that the linear evaluation functional is given by the inner product with a unique vector in $\mathcal{H}$. Therefore $\forall x \in X$, there exists a unique vector $k_x \in \mathcal{H}$ such that

$$f(x) = E_x(f) = \langle f, k_x \rangle, \text{ for all } f \in \mathcal{H}.$$

Definition 3.1.2. *The function k_x is called the Reproducing Kernel for the point x. The 2-variable function $K : X \times X \to \mathbb{F}$ defined by*

$K(x, y) = k_y(x)$ is called the Reproducing Kernel for $\mathcal{H}$.

Note that we have:

$$K(x, y) = k_y(x) = \langle k_y, k_x \rangle.$$

so that

$$K(x,y) = \langle k_y, k_x \rangle = \overline{\langle k_x, k_y \rangle} = \overline{K(y,x)}.$$

in the complex case and $K(x,y) = K(y,x)$ in the real case. Also,

$$\|E_y\|^2 = \|k_y\|^2 = \langle k_y, k_y \rangle = K(y,y).$$

We now look at some examples of Reproducing Kernel Hilbert Spaces. Our examples are drawn from function theory, differential equations and statistics.

3.2 Basic examples

3.2.1 $\mathbb{C}^n$ as an RKHS

We let $\mathbb{C}^n$ denote the vector space of complex n-tuples and for $v = (v_1, ..., v_n), w = (w_1, ..., w_n) \in \mathbb{C}$, we let

$$\langle v, w \rangle = \sum_{i=1}^{n} v_i \overline{w_i},$$

to denote the usual inner product. $\mathbb{C}^n$ with this inner product defined on it is a Hilbert space.
If we let $X = \{1, ..., n\}$, then we could also think of a complex n-tuple as a function $v : X \to \mathbb{C}$, where $v(j) = v_j$. With this identification, $\mathbb{C}^n$ becomes the vector space of all functions on X. If we let $\{e_j\}_{j=1}^n$ denote the "canonical" orthonormal basis for $\mathbb{C}^n$, that is e_j be the function,

$$e_j(i) = \begin{cases} 1, & i = j \\ 0, & i \neq j, \end{cases}$$

then for every $v \in \mathbb{C}^n$ we have

$$v(j) = v_j = \langle v, e_j \rangle.$$

Thus, we see that the "canonical" basis for $\mathbb{C}^n$ is precisely the set of kernel functions for point evaluations when we regard $\mathbb{C}^n$ as a space of functions. This also explains why this basis seems so much more natural than other orthonormal bases for $\mathbb{C}^n$. Note that the reproducing kernel for $\mathbb{C}^n$ is given by

$$K(i,j) = \langle e_j, e_i \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j, \end{cases}$$

which can be thought of as the identity matrix.
More generally, given any (finite or countably infinite) set X, we set

$$l^2(X) = f : X \to \mathbb{C} : \sum_{x \in X} |f(x)|^2 < \infty.$$

Given $f, g \in l^2(X)$, we define $\langle f, g\rangle = \sum_{x\in X} f(x)\overline{g(x)}$. With these definitions $l^2(X)$ becomes a Hilbert space of functions on X. If for a fixed $y \in X$, we let $e_y \in l^2(X)$ denote the function given by

$$e_y(x) = \begin{cases} 1, & i = j \\ 0, & i \neq j, \end{cases}$$

then it is easily seen that $\{e_y\}_{y\in X}$ is an orthonormal basis for $l^2(X)$ and that $\langle f, e_y\rangle = f(y)$, so that these functions are also the reproducing kernels and as before

$$K(x, y) = \langle e_y, e_x\rangle = \begin{cases} 1, & x = y \\ 0, & x \neq y. \end{cases}$$

3.3 Examples from analysis

3.3.1 Sobolev spaces on [0,1]

These are very simple examples of the types of Hilbert Spaces that arise in different equations.

Definition 3.3.2. *A function $f : [0, 1] \to \mathbb{R}$ is absolutely continuous provided that for every $\epsilon > 0$ there exists $\delta > 0$, so that when $(x_1, y_1), ...(x_n, y_n)$ are any non-overlapping intervals contained in [0,1] with $\sum_{j=1}^{n} |y_j - x_j| < \delta$, then $\sum_{j=1}^{n} |f(y_j) - f(x_j)| < \epsilon$.*

It is a well known fact that f absolutely continuous if and only if $f'(x)$ exists for almost all x, the derivative is integrable and up to a constant, f is equal to the integral of tis derivative. Thus, absolutely continuous functions are the functions for which the first fundamental theorem of calculus applies.
Let,

$$\mathcal{H} = \{f | f : [0, 1] \to \mathbb{R}, \text{f is absolutely continuous, } f' \text{ is square integrable and satisfies} f(0) = f(1) = 0\}.$$

The set $\mathcal{H}$ is a vector space of functions on $[0, 1]$. In order to make $\mathcal{H}$ a Hilbert space we endow $\mathcal{H}$ with the nonnegative, sesquilinear form,

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

Let $0 \leqslant x \leqslant 1$ and let $f \in \mathcal{H}$. Since f is absolutely continuous we have

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

Thus by the Cauchy- Schwartz inequality,

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

This last inequality shows that $\langle f, f\rangle = 0$ if and only if $f = 0$.

Thus, $\langle , \rangle$ is an inner product on $\mathcal{H}$. Also, for every $x \in [0,1]$, E_x is bounded with $\|E_x\| \leqslant \sqrt{x}$.
All that remains to show that $\mathcal{H}$ is an RKHS is to show that is it is complete in the norm induced by its inner product. If (f_n) is a Cauchy sequence in this norm, then (f_n') is Cauchy in $L^2[0.1]$ and hence there exists $g \in L^2[0.1]$ to which this sequence converges in the $L^2[0.1]$-sense. By the above inequality. (f_n) must be pointwise Cauchy (and therefore pointwise convergent, since $\mathbb{R}$ is complete) and hence we may define a function by setting $f(x) = \lim_n f_n(x)$. Since

$$f(x) = \lim_n \int_0^x f_n'(t) = \int_0^x g(t)dt.$$

It follows that f is absolutely continuous and that $f' = g$ a.e. Note that even though g was only an equivalence class of functions, $\int_0^x g(t)dt$ was independent of the particular function chosen from the equivalence class. Hence, $f' \in L^2[0,1]$. Finally, $f(0) = \lim_n f_n(0) = 0 = \lim_n f_n(1) = f(1)$. Thus .
We now wish to find the kernel function.
We know that $f(x) = \int_0^1 f'(t)\chi_{[0,x]}(t)dt$. Thus if could solve the boundary-value problem,

$$g'(t) = \chi_{[0,x]}(t), g(0) = g(1) = 0$$

then $g \in \mathcal{H}$ with $f(x) = \langle f, g \rangle$ and so $g = k_x$.
Unfortunately, this boundary-value problem has no solution .Yet we know the function $k_x(t)$ exists and is continuous. Instead, to find the kernel function we formally derive a different boundary-value problem. Then we will show that the function we obtain by this formal solution to $\mathcal{H}$ and we will verify that it is the kernel function. To find $k_x(t)$, we first apply integration by parts. We have

$$\begin{aligned} f(x) = \langle f, k_x \rangle &= \int_0^1 f'(t)k_x'(t)dt \\ &= f(t)k_x'(t)|_0^1 - \int_0^1 f(t)k_x''(t)dt = \int_0^1 f(t)k_x''(t)dt. \end{aligned}$$

Let, δ_y denote the formal Dirac-delta function, then

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

Thus we need to solve the boundary value problem,

$$-k_y''(t) = \delta_y(t), k_y(0) = k_y(1) = 0.$$

The solution to this system of equations is called the Green's function for the differential equation. Solving formally, by integrating twice and checking the boundary conditions, we find

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

3.4 Function theoretic examples

We now consider some examples of reproducing kernel Hilbert Spaces that arise in complex analysis. The Bergman spaces appearing below are named after Stefan Bergman, who originated the theory of reproducing kernel Hilbert spaces and was the earliest researcher to obtain knowledge about spaces from their kernel functions.

3.4.1 The Hardy space of the unit disk $H^2(\mathbb{D})$

This plays a key role in function theory, operator theory and in the theory of stochastic processes.

To construct $H^2(\mathbb{D})$, we first consider formal complex power series,

$$f \sim \sum_{n=0}^{\infty} a_n z^n$$

such that $\sum_{n=0}^{\infty} |a_n|^2 < \infty$. Using the usual definitions for sums and scalar multiples, the set of all such power series clearly forms a vector space. Given another such power series $g \sim \sum_{n=0}^{\infty} b_n z^n$ we define the inner product,

$$\langle f, g \rangle = \sum_{n=0}^{\infty} a_n \overline{b_n}.$$

Thus, we have that $\|f\|^2 = \sum_{n=0}^{\infty} |a_n|^2$.

The map $L : H^2(\mathbb{D}) \to l^2(\mathbb{Z}^+)$ where $\mathbb{Z}^+ = \{0, 1, 2, ...\} = \mathbb{N} \cup \{0\}$, defined by

$$L(f) = (a_0, a_1, a_2, a_3, ...)$$

is a linear inner product preserving isomorphism. Hence we see that $H^2(\mathbb{D})$ can be identified with the Hilbert Space $l^2(\mathbb{Z}^+)$, hence it is itself a Hilbert Space. Thus, we see that the second condition in the definition of an RKHS is met.

Next we show that every power series in $H^2(\mathbb{D})$, converges to define a function on the disk. To see that if $z \in \mathbb{D}$, then

$$|E_z(f)| = |\sum_{n=0}^{\infty} a_n z^n| \leq \sum_{n=0}^{\infty} |a_n||z|^n|$$
$$\leq (\sum_{n=0}^{\infty} |a_n|^2)^{1/2} (\sum_{n=0}^{\infty} |a_n|^2)^{1/2} = \|f\|. \frac{1}{\sqrt{1 - |z|^2}}.$$

Thus,each power series defines a function on $\mathbb{D}$.

Now we want to see that if two power series define the same function on $\mathbb{D}$, then they are the same power series i.e that their coefficients must all be equal. To see this recall that the functions that the power series define are infinitely differentiable on their radius of convergence. By differentiating

the function f, n times and evaluating at 0, we obtain $(n!)a_n$. Thus, if two power series are equal as functions on $\mathbb{D}$, then they are the same power series.
Also the vector space operations on formal power series clearly agree with their vector space operations s functions on $\mathbb{D}$ and so (1) is met.
The above inequality also shows that the map E_z is bounded with $\|E_z\| \leq \dfrac{1}{\sqrt{1-|z|^2}}$ and so $H^2(\mathbb{D})$ is an RKHS on $\mathbb{D}$.

We now compute the kernel function for $H^2(\mathbb{D})$.

Let $w \in \mathbb{D}$, note that $g(z) = \sum_{n=0}^{\infty} \bar{w^n} z^n \in H^2(\mathbb{D})$ and for any $f(z) = \sum_{n=0}^{\infty} a_n z^n \in H^2(\mathbb{D})$, we have that $\langle f, g \rangle = \sum_{n=0}^{\infty} a_n w^n = f(w)$.
Thus g is the reproducing kernel for w and so

$$K(z, w) = k_w(z) = g(z) = \sum_{n=0}^{\infty} \bar{w^n} z^n = \frac{1}{1 - \bar{w} z}.$$

This function is called the **Szego Kernel** on the disk.Note that

$$\|E_z\| = K(z, z) = \frac{1}{\sqrt{1-|z|^2}}.$$

3.4.2 Bergman Spaces on complex Domains

Stefan Bergman introduced the concept of reproducing kernel Hilbert Spaces and used them to study various problems in complex analysis. The spaces that he introduced now bear his name. Let $G \subset \mathbb{C}$ be open and connected. We let

$$B^2(G) = \{f : G \to \mathbb{C} | f \text{is analytic on } G \text{ and } \int \int_G |f(x + \imath y)|^2 dx dy < \infty\}.$$

where $dxdy$ denotes the area measure. We define a sesquilinear form on $B^2(G)$ by

$$\langle f, g \rangle = \int \int_G f(x + iy) \overline{g(x + iy} dx dy,$$

If $f \in B^2(g)$ is nonzero, then since f is analytic, f is continuous and consequently there will be an open set on which $|f|$ is bounded away from 0. Hence,$f \neq 0$ implies that $\langle f, f \rangle \ > 0$, and so $B^2(G)$ is an inner product space.
Also, if $f, g \in B^2(g)$ and $f = g$ a.e. then the continuous function, $|f - g|$ can not be bounded away from 0 on any open set, and so $f(z) = g(z)$ for every $z \in G$. Thus, $B^2(G)$ can be regarded as a vector subspace of $L^2(G)$.

Theorem 3.4.3. *Let $G \subseteq \mathbb{C}$ be open and connected. Then $B^2(G)$ is a RKHS on G.*

Proof. If we fix $w \in G$ and choose $R > 0$ such that the closed ball of radius R centered at w, $\overline{B(w; R)}$ is contained in G (since G is open).

Then by Mean Value property we have

$$f(w) = \frac{1}{2\pi}\int_0^{2\pi} f(w + re^{i\theta})d\theta,$$

for any $0 \le r \le R$.
Multiplying r on both sides and integrating in $r \in [0, R]$, we obtain,

$$\int_0^R rf(w) = \frac{1}{2\pi}\int_0^R \int_0^{2\pi} rf(w + re^{i\theta})d\theta dr$$

$$\implies f(w) = \frac{1}{\pi R^2}\int\int_{B(w;R)} f(x + iy)dxdy.$$

Thus by Cauchy-Schwartz inequality, it follows that

$$\begin{aligned} |f(w)| &\le \frac{1}{\pi R^2}\|f\|(\int\int_{B(w;R)} dA)^{\frac{1}{2}} \\ &= \frac{1}{\pi R^2}\|f\|\sqrt{\pi}R \\ &= \frac{1}{\sqrt{\pi}R}\|f\|. \end{aligned}$$

This proves that for $w \in G$ the evaluation functional is bounded.
So all that remains to prove that $B^2(G)$ is an RKHS is to show that $B^2(G)$ is complete in this norm. Let f_n be a Cauchy sequence in $B^2(G)$. For any $w \in G$. Pick R as above and pick $0 < \delta < d(B(w; R), G^c)$ where $d(.,.)$ denotes the distance between 2 sets.
Then for any z in the closed ball of radius R centered at w we have that the closed ball of radius δ centered at z contained in G. Hence by the above estimate,

$$|f_n(z) - f_m(z)| \le \frac{1}{\delta\sqrt{\pi}}\|f_n - f_m\|.$$

Thus,the sequence of functions is uniformly covergent on every closed ball contained in G. If we let $f(z) = \lim f_n(z)$ denote the pointwise limit of this sequence, then we have that f_n converges uniformly to f on each closed ball contained in G. As a consequence of Cauchy's integral theorem, a sequence of holomorphic functions that converges uniformly on compact sets must converge to a holomorphic function and therefore f is analytic.

Since $B^2(G) \subseteq L^2(G)$ (by definition) and $L^2(G)$ is complete , There exists $h \in L^2(G)$ such that $\|h - f_n\|_2 \to 0$. Moreover, we may choose a subsequence f_{n_k} such that $h(z) = \lim f_{n_k}(z)$ almost everywhere but this implies that $h(z) = f(z)$ almost everywhere and so $\|f - f_n\|_2 \to 0$. Thus $f \in B^2(G)$ and so $B^2(G)$ is complete. □

Definition 3.4.4. *Given any open connected subset $G \subseteq \mathbb{C}$, the reproducing kernel for $B^2(G)$ is called the Berman Kernel for G.*

Note that the above inequality shows $B^2(\mathbb{C}) = (0)$, since in this case R could be taken arbitrarily and so $|f(w)| = 0$ for any $f \in B^2(\mathbb{C})$. Thus, the only analytic function defined on the whole complex plane that is square integrable is the 0 function.

When $A = area(G) < \infty$.
Consider $f(z) = 1 \ \forall \ z \in \mathbb{C}$, then $\int\int_G |f(x+iy)|^2 dxdy = \int\int_G |1|^2 dxdy = area(G)$
$\implies f(z) = 1 \in B^2(G)$ and $\|f(z)\|_{B^2} = \|1\|_{B^2} = A^{\frac{1}{2}}$
In this case, it is natural to re-normalize so that$\|1\| = 1$. To do this we just redefine the inner product to be

$$\langle f, g\rangle = \frac{1}{\pi}\int\int_G f(x+\iota y)\overline{g(x+\iota y)}dxdy.$$

When we will refer to the Bergman Space on such a domain $(G : A = area(A) < \infty)$ we mean the normalized Bergman Space. So, in particular, by the space, $B^2(\mathbb{D})$, we mean the space of square-integrable analytic functions on $\mathbb{D}$, with inner-product

$$\langle f, g\rangle = \frac{1}{\pi}\int\int_{\mathbb{D}} f(x+\iota y)\overline{g(x+\iota y)}dxdy.$$

Chapter 4

Fundamental results

Let X be any set and $\mathcal{H}$ be an RKHS on X with kernel K. In this chapter we begin with a few results that show that K completely determines the space $\mathcal{H}$. We will introduce the concept of a Parseval frame and show that, given any Parseval frame for an RKHS, the kernel can be constructed as a power series. Conversely, any series that yields the kernel in this fashion must be a Parseval frame for the RKHS. Next, we will prove Moore's theorem, which characterizes the functions that are the kernel functions of some RKHS. Such functions are often called either *Positive definite* or *Positive Semidefinite*. Thus, every such function yields an RKHS by Moore's theorem, but it is often quite difficult to obtain a concrete description of the induced RKHS. We call the problem of obtaining the RKHS from the function the *Reconstruction Problem* and we illustrate this process by some important examples.

4.1 Hilbert Space Structure

Proposition 4.1.1. *Let $\mathcal{H}$ be an RKHS on the set X with kernel K. Then the linear span of the functions, $k_y(.) = K(., y)$ is dense in $\mathcal{H}$.*

Proof. A function $f \in \mathcal{H}$ is orthogonal to the span of the functions $k_y : y \in X$ if and only if $\langle f, k_y \rangle = f(y) = 0$ for every $y \in X$, which is if and only if $f = 0$. □

Lemma 4.1.2. *Let H be a RKHS on X and $\{f_n\} \subseteq \mathcal{H}$. If $\lim_n \|f_n - f\| = 0$, then $f(x) = \lim_n f_n(x)$ for every $x \in X$.*

Proof. As we know $\lim_n \|f_n - f\| = 0$ implies $\|f_n - f\|\|k_n\| \to 0$ as $n \to \infty$. Thus our result follows from the following inequality,

$$|f_n(x) - f(x)| = |\langle f_n - f, k_x \rangle| \leq \|f_n - f\|\|k_x\| \text{ (Cauchy Schwarz inequality).}$$

□

Proposition 4.1.3. *Let $\mathcal{H}_i$ be RKHS's on X with kernels, $K_i(x, y) fori = 1, 2$. If $K_1(x, y) = K_2(x, y)$ for all $x, y \in X$, then $\mathcal{H}_1 = \mathcal{H}_2$ and $\|f\|_1 = \|f\|_2$ for every f.*

Proof. Let $K(x, y) = K_1(x, y) = K_2(x, y)$ and $\mathcal{W}_i = span\{k_x\} \in \mathcal{H}_i fori = 1, 2$. By the Proposition 4.1.1, $\mathcal{W}_i$ is dense in $\mathcal{H}_i, i = 1, 2$. Note that for any $f \in \mathcal{W}_i$, we have $f(x) = \sum_j \alpha_j k_{x_j}(x) = \sum_j \alpha_j K(x, x_j)$,which means that values of f are independent of whether we regard it as in $\mathcal{W}_1$ or $\mathcal{W}_2$.

Also, for such an f,

$$\|f\|_1^2 = \sum_{i,j} \alpha_i \overline{\alpha}_j \langle k_{x_i} k_{x_j} \rangle = \sum_{i,j} \alpha_i \overline{\alpha}_j K(x_j, x_i) = \|f\|_2^2.$$

Thus,$\|f\|_1 = \|f\|_2$, for all $f \in W_1 = W_2$.

Finally, if $f \in \mathcal{H}_\infty$, then there exists a sequence of functions, $\{f_n\} \subseteq \mathcal{W}_\infty$ with $\|f - f_n\|_1 \to 0$. Since, f_n is Cauchy in $\mathcal{W}_\infty$ it is also Cauchy in $\mathcal{W}_\in$, so there exists $g \in \mathcal{H}_2$ such that $\|g - f_n\|_2 \to 0$. By the Lemma 4.1.2, $f(x) = \lim_n f_n(x) = g(x)$. Thus, every $f \in \mathcal{H}_1$ is also in $\mathcal{H}_2$ and by an analogous argument, every $g \in \mathcal{H}_2$ is in $\mathcal{H}_1$. Hence $\mathcal{H}_1 = \mathcal{H}_2$.

Finally, $\|f\|_1 = \|f\|_2$ for every f in a dense subset. Thus we get that the norms are equal for every f. □

Definition 4.1.4. *Given vectors $\{h_s : s \in S\}$ in a normal space $\mathcal{H}$, indexed by an arbitrary set S. We say that $h = \sum_{s \in S} h_s$ if for every $\epsilon > 0$, there exists a finite subset $F_0 \subseteq S$ such that for any finite set F with $F_0 \subseteq F \subseteq S$, we have $\|h - \sum_{s \in F} h_s\| < \epsilon$.*

Two examples of this type of convergence are given by the two Parseval identities.
If $\{e_s : s \in S\}$ is an orthonormal basis for a Hilbert Space $\mathcal{H}$, then for any $h \in \mathcal{H}$, we have

$$\|h\|^2 = \sum_{s \in S} |\langle h, e_s \rangle|^2$$

and

$$h = \sum_{s \in S} \langle h, e_s \rangle e_s.$$

Note that here we do not need S to be an ordered set. Perhaps,the key example to keep in mind is the following.
Take $a_n = \dfrac{(-1)^n}{n}, n \in N$, then the series, $\sum_{n=1}^{\infty} a_n$ converges, but $\sum_{n \in N} a_n$ does not converge.

For complex numbers, one can show that $\sum_{n \in N} z_n$ converges if and only if $\sum_{n=1}^{\infty} |z_n|$ converges. Thus convergence is equivalent to absolute convergence in the complex case.

Theorem 4.1.5. *Let $\mathcal{H}$ be a RKHS on X with reproducing kernel, $K(x, y)$. If $\{e_s : s \in S\}$ is an orthonormal basis for $\mathcal{H}$, then $K(x, y) = \sum_{s \in S} \overline{e_s(y)} e_s(x)$ where this series converges pointwise.*

Proof. For any $y \in X$, we have $\langle k_y, e_s \rangle = \overline{\langle e_s, k_y \rangle} = \overline{e_s(y)}$. Hence $k_y = \sum_{s \in S} \langle k_y, e_s \rangle e_s = \sum_{s \in S} \overline{e_s(y)} e_s$, where these sums converge in the norm on $\mathcal{H}$.

But norm convergence implies point wise convergence. Hence, $K(x, y) = k_y(x) = \sum_{s \in S} \overline{e_s(y)} e_s(x)$. □

Example 4.1.6. *In the Hardy Space, the functions $e_n(z) = z^n, n \in \mathbb{Z}^+$ form an orthonormal basis and hence, the reproducing kernel for the Hardy Space is given by*

$$\sum_{s \in S} \overline{e_s(y)} e_s(x) = \sum_{n=0}^{\infty} (z\bar{w})^n = \frac{1}{1 - z\bar{w}}$$

Theorem 4.1.7. *Let $\mathcal{H}$ be an RKHS on X with reproducing kernel K, $\mathcal{H}_0 \subseteq \mathcal{H}$ be a a closed subspace and define(linear map) $P_0 : \mathcal{H} \to \mathcal{H}_0$. Then $\mathcal{H}_0$ is an RKHS on X with reproducing kernel $K_0(x, y) = \langle P_0(k_y), k_x \rangle$*

Proof. Since evaluation of a point in X defines a bounded linear functional on $\mathcal{H}$, thus it remains bounded when restricted to the subspace $\mathcal{H}_0$. Hence, $\mathcal{H}_0$ is an RKHS on X.

For $f \in \mathcal{H}_0$,

$$f(x) = \langle f, k_x \rangle = \langle P_0(f), k_x \rangle = \langle f, P_0^*(k_x) \rangle = \langle f, P_0(k_x) \rangle.$$

Hence, $P_0(k_x)$ is the kernel function for $\mathcal{H}_0$ and we have

$$K_0(x, y) = \langle P_0(k_y), P_0(k_x) \rangle = \langle P_0(k_y), k_x \rangle.$$

□

Definition 4.1.8. *Let $\mathcal{H}$ be a Hilbert Space with inner product $\langle \cdot, \cdot \rangle$. A set of vectors $\{f_s : s \in S\} \subseteq \mathcal{H}$ is called a Parseval Frame for $\mathcal{H}$ provided*

$$\|h\|^2 = \sum_{s \in S} |\langle h, f_s \rangle|^2$$

for every $h \in \mathcal{H}$.

For example if $\{u_s : s \in S\}$ and $\{v_t : t \in T\}$ are two orthonormal bases for $\mathcal{H}$, then the sets $\{u_s : s \in S\} \cup 0$ and $\{(u_s)/\sqrt{2} : s \in S\} \cup \{\frac{v_t}{\sqrt{2}} v : t \in T\}$ are both Parseval frames for $\mathcal{H}$.

In particular, we see that Parseval frames need not be linearly independent sets. The following result shows one of the most common ways that Parseval frames arise.

Proposition 4.1.9. *Let $\mathcal{H}$ be a Hilbert Space and $\mathcal{M} \subseteq \mathcal{H}$ be a closed subspace. Suppose P denote the orthogonal projection of $\mathcal{H}$ onto $\mathcal{M}$. If $\{e_s : s \in S\}$ is an orthonormal basis for $\mathcal{H}$, then $\{P(e_s) : s \in S\}$ is a Parseval Frame for $\mathcal{M}$.*

Proof. For any $h \in \mathcal{M}$, $h = P(h)$ and $\langle h, e_s \rangle = \langle P(h), e_s \rangle = \langle h, P^*(e_s) = \langle h, P(e_s) \rangle$ as $P^* = P$. Thus, $\|h\|^2 = \sum_{s \in S} \|\langle h, P(e_s) \rangle\|^2$ and the result follows. □

The following result shows that any of the Parseval identities can be used to define Parseval frames. Let $l^2(S) := \{g : S \to \mathbb{C} : \sum_{s \in S} \|g(s)\|^2 < \infty\}$ be the Hilbert Space of square- summable functions on S and $e_t : S \to \mathbb{C}$ defined by

$$e_t(s) = \begin{cases} 1 & t = s \\ 0 & t \neq s \end{cases}$$

be the canonical orthonormal basis.

Proposition 4.1.10. *Let $\mathcal{H}$ be a Hilbert Space and $\{f_s : s \in S\} \subseteq \mathcal{H}$. Then the following are equivalent*

1. *The set $\{f_s : s \in S\}$ is a Parseval Frame*
2. *The function $V : \mathcal{H} \to l^2(S)$ given by $(Vh)(s) = \langle h, f_s \rangle$ is a well-defined isometry.*
3. *For all $h \in H$, we have $h = \sum_{s \in S} \langle h, f_s \rangle f_s$.*

Moreover, if $\{f_s : s \in S\}$ is a Parseval Frame, then for any $h_1, h_2 \in \mathcal{H}$ we have $\langle h_1, h_2 \rangle = \sum_{s \in S} \langle h_1, f_s \rangle \langle f_s, h_2 \rangle$.

Proof. (1) $\implies$ (2). First assume that $\{f_s : s \in S\}$ is a Parseval frame and define $V : \mathcal{H} \to l^2(S)$, by $(Vh)(s) = \langle h, f_s \rangle$, , so in terms of the basis, $Vh = \sum_{s \in S} \langle h, f_s \rangle e_s$. Since f_s is a Parseval frame, we see that

$$||Vh||^2 = \langle \sum_{s \in S} \langle h, f_s \rangle e_s, \sum_{s \in S} \langle h, f_s \rangle e_s \rangle = \sum_{s \in S} \langle h, f_s \rangle \overline{\langle h, f_s \rangle} = \sum_{s \in S} |\langle h, f_s \rangle|^2 = ||h||^2$$

(2) $\implies$ (3). Now suppose V is an isometry. Note that $\langle h, V^* e_t \rangle = \langle Vh, e_t \rangle$ and $Vh = \sum_{s \in S} \langle h, f_s \rangle e_s \implies \langle Vh, e_t \rangle = \langle \sum_{s \in S} \langle h, f_s \rangle e_s, e_t \rangle = \langle h, f_t \rangle$ and hence $V^* e_t = f_t$. Since V is an isometry, we have $\langle Vh, Vh \rangle = \langle h, h \rangle$ which is equivalent to saying $\langle Vh, Vh \rangle = \langle h, V^* V h \rangle = \langle h, h \rangle$. Thus $\implies$ $\langle h, (V^* V - I) h \rangle = 0$ from this we get $V^* V = I_{\mathcal{H}}$ Now it follows that

$$h = V^* V h = V^* \left(\sum_{s \in S} \langle h, f_s \rangle e_s \right) = \sum_{s \in S} \langle h, f_s \rangle V^*(e_s) = \sum_{s \in S} \langle h, f_s \rangle f_s$$

for every $h \in \mathcal{H}$.
(3) $\implies$ (1). Finally assume that $\sum_{s \in S} \langle h, f_s \rangle f_s = h$ for all $h \in \mathcal{H}$. We have $\langle h, h \rangle = \sum_{s \in S} \langle h, f_s \rangle \langle f_s, h \rangle = \sum_{s \in S} |\langle h, f_s \rangle|^2$.

Since V is an isometry, for any $h_1, h_2 \in \mathcal{H}$, we have

$$\begin{aligned} \langle h_1, h_2 \rangle_{\mathcal{H}} =& \langle V^* V h_1, h_2 \rangle_{\mathcal{H}} \\ =& \langle V h_1, V h_2 \rangle_{l^2(S)} \\ =& \sum_{s \in S} (V h_1)(s) \overline{(V h_2)(s)} \\ =& \sum_{s \in S} \langle h_1, f_s \rangle \langle f_s, h_2 \rangle \end{aligned}$$

□

Proposition 4.1.11 (Larson). *Let $\{f_s : s \in S\}$ be a Parseval frame for a Hilbert Space $\mathcal{H}$, then there is a Hilbert Space K containing $\mathcal{H}$ as a subspace and an orthonormal basis $\{e_s : s \in S\}$ for K, such that $f_s = P_{\mathcal{H}}(e_s)$ for all $s \in S$, where $P_{\mathcal{H}}$ is the orthogonal projection of K on $\mathcal{H}$*

Proof. Let $K = l^2(S)$ and $V : \mathcal{H} \to l^2(S)$ be the isometry defined in Proposition 4.1.10. By identifying $\mathcal{H}$ with $V(\mathcal{H})$ we may regard $\mathcal{H}$ as a subspace of $l^2(S)$. Note that $P = VV^* : l^2(S) \to l^2(S)$ satisfies $P = P^*$ and $P^2 = (VV^*)(VV^*) = V(V^*V)V^* = VV^* = P$. Thus, P is the orthogonal projection onto some subspace of $l^2(S)$. As $Pe_s = V(V^*e_s) = Vf_s \in V(\mathcal{H})$, we get that P is the projection onto $V(\mathcal{H})$ and when we identify h with Vh, we have P is the projection onto $\mathcal{H}$ with $Pe_s = Vf_s = f_s$. □

The following result was pointed out by M. Papadakis.

Theorem 4.1.12 (Papadakis). *Let $\mathcal{H}$ be an RKHS on X with reproducing kernel K and the $\{f_s : s \in S\}$ is a Parseval Frame for $\mathcal{H}$ iff $K(x,y) = \sum_{s\in S} f_s(x)\overline{f_s(y)}$, where the series converges pointwise.*

Proof. Assuming that the set is a Parseval frame we have that

$$K(x,y) = \langle k_y, k_x\rangle = \sum_{s\in S}\langle k_y, f_s\rangle\langle\overline{f_s(y)}f_s(x)\rangle$$

.

Conversely, assume that the sum of functions give K as above. If α_j are scalars and $h = \sum_j \alpha_j k_{y_j}$ is any finite linear combination of kernel functions, then

$$\begin{aligned}\|h\|^2 &= \sum_{i,j}\alpha_j\overline{\alpha_i}\langle k_{y_j}, k_{y_i}\rangle = \sum_{i,j}\alpha_j\overline{\alpha_i}\langle K(y_i,y_j)\rangle \\ &= \sum_{i,j}\alpha_j\overline{\alpha_i}\sum_{s\in S}\overline{f_s(y_j)}f_s(y_i) = \sum_{i,j}\alpha_j\overline{\alpha_i}\sum_{s\in S}\langle k_y, f_s\rangle\langle f_s, k_y\rangle \\ &= \sum_{s\in S}\langle\sum_j \alpha_j k_{y_j}, f_s\rangle\langle f_s, \sum_i \alpha_i k_{y_i}\rangle = \sum_{s\in S}|\langle h, f_s\rangle|^2\end{aligned}$$

By the Proposition 4.1.1, if we assume that $\mathcal{L}$ denote the linear span of the kernel functions, then $\mathcal{L}$ is dense in $\mathcal{H}$. Thus, $\tilde{V} : \mathcal{L} \to l^2(S)$ defined by $(\tilde{V})h(s) = \langle h, f_s\rangle$ is an isometry on $\mathcal{L}$. Hence, $\tilde{V}$ extends to be an isometry V on $\mathcal{H}$, which is given by the same formula. Thus, the condition to be a Parseval Frame is met by the set $\{f_s : s \in S\}$. □

4.2 Characterization of reproducing kernels

We now try to find the necessary and sufficient conditions for a function $K : X \times X \to \mathbb{C}$ to be the reproducing kernel for some RKHS. We first note some facts about matrices.

1. Let $A = (a_{i,j})$ be a $n \times n$ complex matrix. Then A is *Positive* if and only if for every $\{\alpha_1, \alpha_2, ..., \alpha_n\} \in \mathbb{C}$ we have $\sum_{i,j=1}^n \bar{\alpha}_i\alpha_j a_{i,j} \geq 0$. We denote this by $A \geq 0$.

2. If we assume $\langle\cdot,\cdot\rangle$ denote the usual inner product on $\mathbb{C}^n$, then in terms of the inner product, $A \geq 0$ if and only if $\langle Ax, x\rangle \geq 0$ for every $x \in \mathbb{C}^n$. In fact the sum in the definition is $\langle Ax, x\rangle$ for a vector x whose i-th component is the number α_i.

3. If $A \geq 0$ and $B \geq 0$ both are $n \times n$ matrices, then $A + B \geq 0$ and $rA \geq 0$ for any $r \in \mathbb{R}^+$.

4. A matrix $A \geq 0$ if and only if $A = A^*$ and every eigenvalue of A is nonnegative.

5. A matrix $A > 0$ if and only if $A = A^*$ and every eigenvalue of A is strictly positive.

6. Since A is a $n \times n$ matrix, we see that $A > 0$ is equivalent to $A \geq 0$ and A is invertible.

Definition 4.2.1. *Let X be a set and $K : X \times X \to \mathbb{C}$ be a function of two variables. Then K is called a Kernel function provided that for every n and the choice of n distinct points, $\{x_1, x_2, ..., x_n\} \subseteq X$, the matrix $(K(x_i, x_j)) \geq 0$. The kernel function is denoted as $K \geq 0$.*

Proposition 4.2.2. *Let X be a set and $\mathcal{H}$ be an RKHS on X with reproducing kernel K. Then K is a kernel function.*

Proof. Fix $\{x_1, x_2, ..., x_n\} \subseteq X$ and $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{C}$. Then we have

$$\sum_{i,j} \bar{\alpha}_i \alpha_j K(x_i, x_j) = \langle \sum_j \alpha_j k_{x_j}, \sum_i \alpha_i k_{x_i} \rangle = \| \sum_j \alpha_j k_{x_j} \|^2 \geq 0,$$

and the result follows. □

In general, for a reproducing kernel Hilbert space the matrix $P = (K(x_i, x_j))$ is not strictly positive,but the above calculation shows that $\langle P\alpha, \alpha \rangle = 0$ if and only if $\| \sum_j \alpha_j k_{x_j} \| = 0$. Hence, for every $f \in \mathcal{H}$ we have $\| \sum_j \alpha_j k_{x_j} \| = 0$. As $\sum_j \bar{\alpha}_j f(x_j) = \langle f, \sum_j \alpha_j k_{x_j} \rangle = 0$, in this case there is an equation of linear dependence between the values of every function in $\mathcal{H}$ at the finite set of points.

Such examples do naturally exist. Recall that in the Sobolev spaces on [0,1], we were interested in spaces with boundary conditions, like $f(0) = f(1)$, in which $k_1(t) = k_0(t)$.

On the other hand, many spaces of analytic functions, such as the Hardy or Bergmann spaces, contain all polynomials. Note that there is no equation of the form $\sum_j \beta_j p(x_j) = 0$, with β_j not all zero, which is satisfied by all polynomials. Consequently, the reproducing kernel that is satisfied by all polynomials. Hence, the reproducing kernels for these spaces always define matrices that are strictly positive and invertible.

Thus for example, the Szego kernel for the Hardy space, we see that for any choice of points $\lambda_1, ..., \lambda_n$ in the disk, the matrix $\dfrac{1}{1 - \overline{\lambda_i} \lambda_j}$ is invertible, by standard linear algebraic methods.

Theorem 4.2.3 (Moore)**.** *Let X be a set and $K : X \times X \to \mathbb{C}$ be a function. If K is a kernel function, then there exists a reproducing kernel Hilbert space $\mathcal{H}$ of functions on X such that K is the reproducing kernel of $\mathcal{H}$.*

Proof. Let $k_y : X \to \mathbb{C}$ be a function defined by $k_y(x) = K(x, y)$. From Theorem 4.1.1 we know that if K is the kernel function of an RKHS then the span of these functions is dense in it. So it is a natural attempt to define a sesquilinear form on the vector space that is the span of the functions k_y, with $y \in X$.

Let $W \subseteq \mathcal{F}(X)$ be the vector space of functions spanned by the set $\{k_y : y \in X\}$ and $B : W \times W \to \mathbb{C}$ given by $B(\sum_j \alpha_j k_{y_j}, \sum_i \beta_i k_{y_i}) = \sum_{i,j} \alpha_j \overline{\beta_i} K(y_i, y_j)$, where α_j and β_i both are

scalars.
Since a given function could be expressed in many different ways as sum of the functions k_y, our first task to show that B is well defined. For showing that B is well-defined on W, it is enough to show that if $f = \sum_j \alpha_j k_{y_j}$ is identically zero as a function on X, then $B(f,g) = B(g,f) = 0$ for every $g \in W$. Since W is spanned by the functions k_y, to prove the last equation it will be enough to show that $B(f,k_y) = B(k_y,f) = 0$. By definition $B(f,k_x) = \sum_j K(x,y_j) = f(x) = 0$. Similarly,

$$B(k_x, f) = \sum_j \overline{\alpha_j} K(y_j, x) = \sum_j \overline{\alpha_j K(x, y_j)} = \overline{f(x)} = 0.$$

Conversely, if $B(f,w) = 0$ for every $w \in W$, then by taking $w = k_y$, we see that $f(y) = 0$. Thus, $B(f,w) = 0$ for all $w \in W$ if and only if f is identically zero as a function on X.

Thus, B is well-defined and it is easy to check that it is sesquilinear. Moreover, for any $f \in W$ we have $f(x) = B(f,k_x)$.

As $(K(y_i,y_j))$ is positive(or semi-definite), for any $f = \sum_j \alpha_j k_{y_j}$, we have $B(f,f) = \sum_{i,j} \alpha_j \overline{\alpha_i} K(y_i,y_j) \geq 0$. Thus B defines a semidefinite inner product on W. Hence, by the same proof as for the Cauchy-Schwarz inequality, one can observe that $B(f,f) = 0$ if and only if $f(y) = B(f,k_y) = 0$ for every $y \in X$, i.e f is identically 0. Therefore, B is an inner product on W.
Now given any inner product on a vector space, we can complete the space by taking equivalence classes of Cauchy sequences from W to obtain a Hilbert space. We have to show that every element of $\mathcal{H}$ can be identified uniquely with a function on X.

For, given $h \in \mathcal{H}$ define

$$\hat{h}(x) = \langle h, k_x \rangle$$

and take

$$\hat{\mathcal{H}} = \{\hat{h} : h \in \mathcal{H}\}$$

so that $\hat{H}$ is a set of functions on X. If $L : \mathcal{H} \to \mathcal{F}(X;\mathbb{C})$ is defined by $L(h) = \hat{h}$, then L is clearly linear and so $\hat{H}$ is a vector space of functions on X. Moreover for any function $f \in W$, we have $\hat{f}(x) = f(x)$.

We have to show that the map that sends $h \to \hat{h}$ is one-to-one. That is, $\hat{h}(x) = 0$ for all $x \in X$ if and only if $h = 0$.

Suppose that $\hat{h}(x) = 0$ for every $x \in X$. Then $h \perp k_x$ for every $x \in X$, so $h \perp W$. Since W is dense in $\mathcal{H}$, we have $h = 0$ and so the map $L : \mathbb{H} \to \hat{\mathcal{H}}$ is one-to-one and onto. Thus, if we define an inner product on $\hat{\mathbb{H}}$ by $\langle \hat{h_1}, \hat{h_2} \rangle$, then $\hat{\mathcal{H}}$ will be a Hilbert space of functions on X. Since

$$E_x(\hat{h}) = \hat{h}(x) = \langle h, k_x \rangle = \langle \hat{h}, \hat{k}_x \rangle$$

we see that every point evaluation is bounded and $\hat{k}_x = k_x$ is the reproducing kernel for the point x. Thus, $\hat{H}$ is an RKHS on X. Since $\hat{k}_y$ is the reproducig kernel for the point y, we get $\hat{k}_y(x) = \langle k_y, k_x \rangle = K(x,y)$ is the reproducing kernel for $\hat{H}$.

□

4.2.4 The Reconstruction Problem

If we start with the Szejo Kernel on the disk, $K(z,w) = \dfrac{1}{1-\overline{w}z}$, then the space W that we have obtain in the proof of Moore's Theorem consists of linear combination of the functions $k_w(z)$, which are rational functions with a single pole of order one outside the disk.

Thus the space W doesn't contain polynomials. Yet the space $\mathcal{H}(K) = H^2(\mathbb{D})$ contains the polynomials as a dense subset and has the set $\{z^n : n \geq 1\}$ as an orthonormal basis.

Theorem 4.2.5. *Let X be a topological space, with product topology on $X \times X$ and $K : X \times X \to \mathbb{C}$ be a kernel function. If K is continuous then every function in $\mathcal{H}(K)$ is continuous.*

Proof. Let $f \in \mathcal{H}(K)$ and fix $y_0 \in X$. Given $\epsilon > 0$, we have to prove that there is a neighborhood U of y_0 such that for every $y \in U$, $|f(y) - f(y_0)| \leq \epsilon$. By the continuity of K, we can choose a neighborhood $V \subseteq X \times X$ of (y_0, y_0) such that $(x,y) \in V$ implies that

$$|K(x,y) - K(y_0,y_0)| < \frac{\epsilon^2}{3(\|f\|^2+1)}$$

Since $X \times X$ is endowed with the product topology, we can pick a neighborhood $U \subseteq X$ of y_0 such that $U \times U \subseteq V$. For $y \in U$, we have

$$\begin{aligned}\|k_y - k_{y_0}\|^2 &= K(y,y) - K(y,y_0) - K(y_0,y) + K(y_0,y_0)\\ &= [K(y,y) - K(y_0,y_0)] - [K(y,y_0) - K(y_0,y_0)] - [K(y_0,y) - K(y_0,y_0)]\\ &< \frac{\epsilon^2}{\|f\|^2+1}.\end{aligned}$$

Hence,

$$|f(y) - f(y_0)| = |\langle f, k_y - k_{y_0}\rangle| \leq \|f\|\|k_y - k_{y_0}\| < \epsilon,$$

which completes the proof of the theorem. □

Proposition 4.2.6. *Let $K : X \times X \to \mathbb{C}$ be a kernel function and $\mathcal{H}(K)$ be the corresponding RKHS. Then the function $\overline{K}$ is also a kernel function and we have $\mathcal{H}(\overline{K}) = \{\bar{f} : f \in \mathcal{H}(K)\}$. Moreover, the map $C : \mathcal{H}(K) \to \mathcal{H}(\bar{K})$ defined by $C(f) = \bar{f}$ is a surjective conjugate-linear isometry.*

Proof. Given $x_1, x_2, x_3, ...x_n$ and $\alpha_1, ..., \alpha_n \in \mathbb{C}$, we have

$$\sum_{i,j=1}^{n} \bar{\alpha}_i \alpha_j \bar{K}(x_i,x_j) = \overline{\sum_{i,j=1}^{n} \alpha_i \bar{\alpha}_j K(x_i,x_j)} = \overline{\sum_{i,j=1}^{n} \bar{\alpha}_i \alpha_j K(x_i,x_j)} \geq 0$$

as K is a kernel function. Thus, $\bar{K}$ is a kernel function.

From the proof of Moore's Theorem, we see that the linear span of the functions $\bar{K}(\cdot,y) = \overline{k_y}(\cdot)$ is dense in $\mathcal{H}(\overline{K})$. First, we want to show that for any of points $y_1, y_2, ..., y_n \in X$ and scalars $\alpha_1, ..., \alpha_n \in \mathbb{C}$, setting

$$C\left(\sum_{j=1}^{n} \alpha_j k_{y_j}\right) = \sum_{j=1}^{n} \overline{\alpha}_j \bar{k}_{y_j}$$

yields a well-defined isometric conjugate linear map on these dense linear spans.

To show that C is well-defined it is enough to show that if we had two different ways to express the same function in $\mathcal{H}(K)$ as a linear combination of the function k_y, then the corresponding linear combination of the functions $\bar{k}_y$ with the conjugate linear coefficients would give the same function in $\mathcal{H}(\bar{K})$. To show this, it is enough to show that if a linear combination of functions k_y adds up to the 0 function in $\mathcal{H}$, then the corresponding conjugate linear combination of the functions $\bar{k}_y$ adds up to the 0 function in $\mathcal{H}(\bar{K})$.

$$\begin{aligned}
\|\sum_{j=1}^{n}\alpha_j\bar{k}_{y_j}\|^2_{\mathcal{H}(\bar{K})} &= \sum_{i,j=1}^{n}\alpha_j\bar{\alpha}_i\langle\bar{k}_{y_j}\bar{k}_{y_i}\rangle_{\mathcal{H}(\bar{K})}\\
&= \sum_{i,j=1}^{n}\alpha_j\bar{\alpha}_i\bar{K}(y_i,y_j) = \overline{\sum_{i,j=1}^{n}\bar{\alpha}_j\alpha_i K(y_i,y_j)}\\
&= \overline{\sum_{i,j=1}^{n}\bar{\alpha}_j\alpha_i\langle k_{y_j},k_{y_i}\rangle_{\mathcal{H}(k)}}\\
&= \|\sum_{j=1}^{n}\bar{\alpha}_j k_{y_j}\|^2_{\mathcal{H}(K)}
\end{aligned}$$

This calculation shows that the conjugate linear map C is isometry, so that if the first function adds up to 0, so does the second function. Thus, C is well-defined, isometry and conjugate linear on these dense subspaces.

Now it is a standard argument in functional analysis that, by taking limits in the domain, we can extend any bounded conjugate linear map from a dense subspace to the whole space. Moreover, since C is isometry, this extension to the whole space will be isometry. Since C is isometry and its range contains a dense subspace, C will map $\mathcal{H}(K)$ onto $\mathcal{H}(\bar{K})$. Note that on the original linear span, C takes a function to its complex conjugate. Hence, $\mathcal{H}(\bar{K}) = C(\mathcal{H}(K)) = \{\bar{f} : f \in \mathcal{H}(K)\}$. □

4.2.7 The RKHS induced by a function

We start with an example of a kernel function that yields a one-dimensional RKHS.

Proposition 4.2.8. *Let X be a set, let f be a non zero function on X and set $K(x,y) = f(x)\overline{f(y)}$. Then K is a kernel function, $\mathcal{H}(K)$ is one dimensional space spanned by f, $\|f\| = 1$.*

Proof. To see that K is positive, we compute

$$\sum_{i,j}\alpha_j\overline{\alpha_i}K(x_i,x_j).$$

Now,

$$\sum_{i,j}\alpha_j\overline{\alpha_i}K(x_i,x_j) = \sum_{i,j}\alpha_j\overline{\alpha_i}f(x_i)\overline{f(x_j)} = \left|\sum_{i}\overline{\alpha_i}f(x_i)\right|^2 \geq 0$$

To find $\mathcal{H}(K)$, note that every function $k_y = \overline{f(y)}f$. Hence the subspace W, used in the proof of Moore's theorem, is just the one dimensional space spanned by f. Since finite dimensional spaces are automatically complete, $\mathcal{H}(K) = \overline{span} f$.

Finally, we compute the norm of f. Fix any point y such that $f(y) \neq 0$. Then $|f(y)|^2.\|f\|^2 = \|\overline{f(y)}f\|^2 = \|k_y\|^2 = \langle k_y, k_y\rangle = K(y,y) = |f(y)|^2 \implies |f(y)|^2.\|f\|^2 = |f(y)|^2$ and it follows that $\|f\| = 1$.

□

4.2.9 The RKHS of the Min function

We prove that the function

$$K : [0, +\infty) \times [0, +\infty) \to \mathbb{R}$$

defined by $K(x, y) = min(x, y)$, is a kernel function and try to obtain some information about $\mathcal{H}_{\mathbb{R}}(K)$. This kernel function and the corresponding RKHS play an important role in the study of Brownian motion.

Lemma 4.2.10. *Let J_n denote the $n \times n$ matrix with every entry equal to 1. Then $J_n \geq 0$, J_n has n eigenvalue of multiplicity one and every other eigenvalue of J_n is equal to 0.*

Proof. Given $v = (\alpha_1, \alpha_2, \alpha_3, ..., \alpha_n) \in \mathbb{C}^n$. we have

$$\langle J_n v, v\rangle = \sum_{i,j=1}^{n} \alpha_i \bar{\alpha}_i = \left|\sum_{j=1}^{n} \alpha_j\right|^2 \geq 0.$$

Hence, $J_n \geq 0$. If v_1 is a vector with each entry equal to 1, then $J_n v_1 = n v_1$ so one eigenvalue is n.

Note that $J_n^2 = nJ_n$. If w is any non-zero eigenvector with eigenvalue λ then $\lambda^2 w = J_n^2 w = nJ_n w = n\lambda w$. Hence, $\lambda^2 = n\lambda$ so that $\lambda \in \{0, n\}$. However, since the trace of J_n is the sum of all eigenvalues which is equal to n, which implies all other eigenvalues of J_n as equal to 0. □

Proposition 4.2.11. *Let $K : [0, \infty) \times [0, \infty) \to \mathbb{R}$ be defined by the $K(x, y) = min\{x, y\}$. Then K is a kernel function.*

Proof. For $x_1, x_2, x_3, ..., x_n \in [0, \infty)$ we have to show that the corresponding matrix $(K(x_i, x_j))$ is positive.The proof is by induction on the number of points n. Clearly, $min\{x, x\} = x \geq 0$, so the case of $n = 1$ is done. If we permute the points such that $0 \leq x_1 \leq x_2 \leq .... \leq x_n$, this corresponds to conjugating the original matrix by a permutation unitary, which does not affect whether or not the matrix is positive. So we assume that the points are given in this ordering.

The matrix $(K(x_i, y_j))$ has the form

$$\begin{bmatrix} x_1 & x_1 & \cdots & x_1 \\ x_1 & x_2 & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} = x_1 J_n + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_2 - x_1 & \cdots & x_n - x_1 \end{bmatrix}$$

where J_n is the matrix of all ones. Since $x_1 \geq 0$ and $J_n \geq 0$ the first matrix in this sum is positive. The lower $(n-1) \times (n-1)$ block m the second matrix is of the form $K(y_i, y_j)$ where $y_i = x_i - x_1 \geq 0$, for $i \geq 2$. By the induction hypothesis, this $(n-1) \times (n-1)$ positive matrix. Since the sum of two positive matrices is positive the matrix $(K(x_i, x_j)) \geq 0$.

Thus, K induces an RKHS of continuous real-valued functions $\mathcal{H}_{\mathbb{R}}(K)$ on $[0, +\infty)$ and we would like to try to get some information about this space. First, it easy to show that K is continuous and hence, by our earlier result, every function in this space is continuous on $[0, +\infty)$.

We look at a typical function in $W_{\mathbb{R}}$. Choose $y_1 < .... < y_n$ in $[0, +\infty)$ and scalars $a_1, a_2, ..., a_n \in \mathbb{R}$, we see that a typical function is given by

$$\sum_{i=1}^{n} a_i k_{y_i}(x) = \begin{cases} (\sum_{i=1}^{n} a_i)x, & 0 \leq x < y_1 \\ a_1 y_1 + (\sum_{i=2}^{n} a_i y_i), & y_1 \leq x < y_2 \\ \vdots & \vdots \\ (\sum_{i=1}^{n} a_i y_i) + a_n x, & y_{n-1} \leq x < y_n \\ \sum_{i=1}^{n} a_i y_i & y_n \leq x \end{cases}$$

Thus we see that every function in the span of the kernel function is continuous, piecewise linear , 0 at 0 and eventually constant. Conversely, it can be shown that every function belongs to the span of the kernel functions. So $\mathcal{H}_{\mathbb{R}}(K)$ will be a space of continuous functions that is the completion of this space of "sawtooth " functions. □

4.2.12 The RKHS induced by the inner product

Definition 4.2.13. *Let $\mathcal{H}$ be a Hilbert Space and $h_1, h_2, ..., h_n \in \mathcal{H}$. Then the $n \times n$ matrix*

$$(\langle h_i, h_j \rangle)$$

is called the GRAMMIAN of these vectors.

Proposition 4.2.14. *Let $\mathcal{H}$ be a Hilbert Space and $h_1, h_2, ..., h_n \in \mathcal{H}$. Then their Grammian, $G = (\langle h_i, h_j \rangle)$ is a positive semidefinite matrix. Moreover, G is a positive definite matrix if and only if $h_1, ..., h_n$ are linearly independent.*

Proof. Let $y = (y_1, ..., y_n)^t \in \mathbb{C}^n$ and then

$$\langle Gy, y \rangle = \sum_{i,j=1}^{n} \langle h_i, h_j \rangle y_j \bar{y}_i = \| \sum_{i=1}^{n} \bar{y}_i h_i \|^2 \geq 0.$$

Also,$\langle Gy, y \rangle = 0$ if and only if the corresponding linear combination is the 0 vector. Hence, G is positive definite if and only if no nontrivial linear combination of $h_1, ..., h_n$ is 0. □

Proposition 4.2.15. *Let $\mathcal{L}$ be a Hilbert Space with inner product $\langle \cdot, \cdot \rangle$ and $K : \mathcal{L} \times \mathcal{L} \to \mathbb{C}$ be defined by $K(x, y) = \langle x, y \rangle$. Then K is a kernel function on $\mathcal{L}$, $\mathcal{H}(K)$ is the vector space of bounded linear functionals on $\mathcal{L}$ and the norm of a functional in $\mathcal{H}(K)$ is the same as its norm as a bounded linear functional.*

Proof. The Proposition 4.2.14 shows that K is a kernel function. Note that for each $y \in \mathcal{L}, k_y : \mathcal{L} \to \mathbb{C}$ is the bounded linear functional defined as $k_y(\cdot) = \langle \cdot, y \rangle$. Thus, linear combinations of kernel functions are again bounded linear functionals on $\mathcal{L}$. We need to see that every function in $\mathcal{H}(K)$ is of this form.

By the Riesz Representation theorem, each bounded linear functional $f : \mathcal{L} \to \mathbb{C}$ is uniquely determined by a vector $w \in \mathcal{L}$, so that $f = f_w$ where

$$f_w(v) = \langle v, w \rangle.$$

Note that, given a scalar $\lambda \in \mathbb{C}$, the linear functional $\lambda f_w = f_{\bar{\lambda} w}$, i.e. the space of bounded linear functionals is conjugate linearly isomorphic to $\mathcal{L}$ and is itself a Hilbert Space in the inner product

$$\langle f_{w_1}, f_{w_2} \rangle = \langle w_2, w_1 \rangle.$$

Let $\mathcal{H} = \{f_w : w \in \mathcal{L}\}$ be the Hilbert Space of bounded linear functionals on $\mathcal{L}$. For each $x \in \mathcal{L}$, the evaluation map $E_x : \mathcal{H} \to \mathbb{C}$, given by $E_x(f_w) = f_w(x)$ satisfies $|E_x(f_w)| = |\langle x, f_w \rangle| \leq \|x\| \|f_w\|,$. So every evaluation map is bounded on $\mathcal{H}$ and so $\mathcal{H}$ is an RKHS. For each $x \in X$ and each $f_w \in \mathcal{H}$, we have

$$f_w(x) = \langle x, w \rangle = \langle f_w, f_x \rangle.$$

So the kernel function for evaluation at x is $k_x = f_x$. Hence, for $x, y \in X$, the kernel function for $\mathcal{H}$ is

$$K_{\mathcal{H}}(x, y) = k_y(x) = f_y(x) = K(x, y).$$

Thus, $K_{\mathcal{H}} = K$ and the result followed by the uniqueness of the RKHS determined by a kernel function.

Note 4.2.16. *If we define $C : \mathcal{L} \to \mathcal{H}(K)$ by $C(y) = k_y$, then $C(\lambda y) = k_{\lambda y} = \bar{\lambda} k_y$. As the inner product is conjugate linear in the second variable, C is the usual conjugate linear identification between a Hilbert space to its dual.*

□

Chapter 5

Interpolation and approximation

5.1 Interpolation in an RKHS

One of the primary applications of the theory of reproducing kernel Hilbert spaces is the problems of interpolation and approximation.

Definition 5.1.1. *Let X and Y be two sets and $\{x_1, x_2, ..., x_n\} \subseteq X$ be a collection of distinct points, and $\{\lambda_1, \lambda_2, ..., \lambda_n\} \subseteq Y$. We say that a function $g : X \to Y$ Interpolates these points provided that $g(x_i) = \lambda_i$, for all $i = 1, 2, ..., n$.*

Note 5.1.2. *Given a finite set $F = \{x_1, ..., x_n\} \subseteq X$ of distinct points, we will denote the subspace spanned by the kernel functions $\{k_{x_1}, ..., k_{x_n}\}$ as $\mathcal{H}_F \subseteq \mathcal{H}$.*

Theorem 5.1.3. *The $dim(\mathcal{H}(F)) \leq n$ and this is strictly less if and only if there is some nonzero equation of linear dependence among these functions.*

Proof. Suppose that $\sum_{j=1}^{n} \alpha_j k_{x_j} = 0$, then for every $f \in \mathcal{H}$,

$$0 = \langle f, \sum_{j=1}^{n} \alpha_j k_{x_j} \rangle = \sum_{j=1}^{n} \bar{\alpha}_j f(x_j).$$

We see that $dim(\mathcal{H}_F) < n$ if and only if the values of every $f \in \mathcal{H}$ at the points in F satisfy some linear relation. This is equivalent to the fact that the linear map $T_F : \mathcal{H} \to \mathbb{C}^n$ defined by $T_F(f) = (f(x_1), f(x_2), ..., f(x_n))$ is not onto. Since any vector $(\alpha_1, ..., \alpha_n)$ expressing an equation of linear dependence would be orthogonal to the range of T_F.

Thus, if $dim(\mathcal{H}(F)) < n$ then there exist $\lambda_1, ..., \lambda_n \in \mathbb{C}$ which cannot be interpolated by any $f \in \mathcal{H}$. □

Note 5.1.4. *We've seen that it is possible for there to be such equation of linear dependence between the kernel functions, and sometimes when we construct a RKHS this can be desirable property. This was the case for the Sobolev space, where the boundary condition of the differential equation is $f(0) = f(1) = 0$ requires $k_1 = k_0 = 0$. Changing the boundary conditions to $f(0) = f(1)$ would require $k_1 = k_0$.*

Let P_F be the orthogonal projection of $\mathcal{H}$ onto $\mathcal{H}_F$.

Note that $g \in \mathcal{H}_F^{\perp}$ if and only if $g(x_i) = \langle g, k_{x_i} \rangle = 0$, for all $i = 1, 2, ..., n$. Hence, for any $h \in \mathcal{H}$, we have

$$P_F(h)(x_i) = h(x_i), i = 1, ..., n$$

Proposition 5.1.5. *Let $\{x_1, x_2, ..., x_n\}$ be a set of distinct points in X and $\{\lambda_1, \lambda_2, ..., \lambda_n\} \subseteq \mathbb{C}$. If there exists $g \in \mathcal{H}$ that interpolates these values, then $P_F(g)$ is the unique function of minimum norm that interpolates $\{\lambda_1, \lambda_2, ..., \lambda_n\}$.*

Proof. Let $S = \{f : F \to Y | f \text{ interpolates these values}\}$ where $F = \{x_1, x_2, ..., x_n\}$ and $Y = \{\lambda_1, \lambda_2, ..., \lambda_n\}$. If $g_1, g_2 \in S$ then $g_1(x_i) - g_2(x_i) = 0$ for all $i \in 1, 2, ..., n$. This implies that $g_1 - g_2 \in \mathcal{H}^{\perp}$ by the previous statement.

Thus, all possible solutions of the interpolation problem are of the form $g + h \in \mathcal{H}^{\perp}$ and $P_F(g)$ belongs to this set.

Note that for any $h \in \mathcal{H}^{\perp}$ we have $P_F(g) = P_F(g + h)$. Hence $\|P_F(g)\| = \|P_F(g + h)\| \leq \|g + h\|$ and so $P_F(g)$ is the unique vector of minimum norm that interpolates these values. □

Definition 5.1.6.
Null space of A: $\mathcal{N}(A) := \{w \in \mathbb{C}^n : Aw = 0\}$
Range space of A: $\mathcal{R}(A) := \{Aw : w \in \mathbb{C}^n\}$.

Proposition 5.1.7. *Let X be a set, $\mathcal{H}$ be an RKHS on X with kernel K and $\{x_1, x_2, ..., x_n\} \subseteq X$ be a finite set of distinct point. If $w = (\alpha_1, \alpha_2, ..., \alpha_n)^t$ is a vector in the nullspace of $(K(x_i, x_j))$, then the function, $f = \sum_j \alpha_j k_{x_j}$ is identically 0. Consequently, if $w_1 = (\alpha_1, ..., \alpha_n)^t$ and $w_2 = (\beta_1, ..., \beta_n)^t$ are two vectors satisfying $(K(x_i, x_j))w_1 = (K(x_i, x_j))w_2$, then*

$$\sum_{j=1}^{n} \alpha_j k_{x_j}(y) = \sum_{j=1}^{n} \beta_j k_{x_j}(y)$$

for every $y \in X$.

Proof. We have that $f = 0$ if and only if $\|f\| = 0$. Now

$$\|f\|^2 = \sum_{i,j} \bar{\alpha}_i \alpha_j \langle k_{x_j}, k_{x_i} \rangle = \sum_{i,j} \bar{\alpha}_i \alpha_j K(x_i, x_j) = \langle (K(x_i, x_j))w, w \rangle_{\mathbb{C}^n} = 0$$

, and the result follows.

To see this last remark, note that $w_1 - w_2$ is in the nullspace of the matrix and so the function $\sum_{j=1}^{n} (\alpha_j - \beta_j) k_{x_j}$ is identically 0. □

Theorem 5.1.8 (Interpolation in an RKHS)**.** *Let $\mathcal{H}$ be an RKHS on X with reproducing kernel K, $F = \{x_1, x_2, ..., x_n\} \subset X$ be distinct points, and $\{\lambda_1, ..., \lambda_n\} \subset \mathbb{C}$. Then there exists $g \in \mathcal{H}$ that interpolates these values if and only if $v = (\lambda_1, ..., \lambda_n)^t$ is in the range of the matrix $(K(x_i, x_j))$. Moreover, in this case if we choose a vector $w = (\alpha_1, ..., \alpha_n)^t$ whose image is v, then $h = \sum_i \alpha_i k_{x_i}$ is the unique function of minimum norm in $\mathcal{H}$ that interpolates these points. Moreover, $\|h\|^2 = \langle v, w \rangle$.*

Proof. First, assume that there exists $g \in \mathcal{H}$ such that $g(x_i) = \lambda_i$, for all $i = 1, ..., n$. Then the solution of minimal norm is $P_F(g) = \sum_j \beta_j k_{x_j}$ for some scalars, $\beta_1, ..., \beta_n$. Since $\lambda_i = g(x_i) =$

$P_F(g)(x_i) = \sum_j \beta_j k_{x_j}(x_i)$, we have $w_1 = (\beta_1, ..., \beta_n)^t$ is a solution of $v = (K(x_i, x_j))w$.

Conversely, if $w = (\alpha_1, ..., \alpha_n)^t$ is any solution of the equation $v = (K(x_i, x_j))w$ and we set $h = \sum_j \alpha_j k_{x_j}$, then h will be an interpolating function.

Note that $w - w_1$ is in kernel of the matrix $(K(x_i, x_j))$ and by the Proposition 5.1.7, $P_F(g)$ and h are the same functions. Hence, h is the function of minimal norm that interpolates these points. Finally,

$$\|h\|^2 = \sum_{i,j} \bar{\alpha}_i \alpha_j K(x_i, x_j) = \langle (K(x_i, x_j))w, w \rangle = \langle v, w \rangle.$$

□

Corollary 5.1.9. *Let $\mathcal{H}$ be a RKHS on X with reproducing kernel K and $F = \{x_1, ..., x_n\} \subset X$ be distinct. If the matrix $(K(x_i, x_j))$ is invertible, then for any $\{\lambda_1, ...\lambda_n\} \subseteq \mathbb{C}$ there exist a function interpolating these values and the unique interpolating function of minimum norm is given by the formula, $g = \sum_j \alpha_j k_{x_j}$ where $w = (\alpha_1, ..., \alpha_n)^t$ is given by $w = (K(x_i, x_j))^{-1} v$, with $v = (\lambda_1, ..., \lambda_n)^t$.*

5.1.10 Strictly positive kernels

Given a set X, a kernel function $K : X \times X \to \mathbb{C}$ is called *Strictly Positive* if and only if, for every n and every set of distinct points $\{x_1, ..., x_n\} \subseteq X$, the matrix

$$(K(x_i, x_j))$$

is strictly positive definite, i.e.

$$\sum_{i,j=1}^{n} \bar{\alpha}_i \alpha_j K(x_i, x_j) > 0,$$

whenever $\{\alpha_1, ..., \alpha_n\} \subseteq \mathbb{C}$ are not all 0.

Theorem 5.1.11. *Let X be a set and $K : X \times X \to \mathbb{C}$ be a kernel. Then the following are equivalent:*

1. *K is strictly positive;*
2. *For any n and set of distinct points $\{x_1, x_2, ..., x_n\}$, the kernel functions $k_{x_1}, ..., k_{x_n}$ are linearly independent;*
3. *For any n, the set of distinct points $\{x_1, ..., x_n\}$, and the set $\{\alpha_1, ..., \alpha_n\} \subseteq \mathbb{C}$ that are not all 0, there exists $f \in \mathcal{H}(K)$ with*
$$\alpha_1 f(x_1) + ... + \alpha_n f(x_n) \neq 0;$$
4. *For any n and the set of distinct points $\{x_1, x_2, ..., x_n\}$. there exists functions, $g_1, ..., g_n \in \mathcal{H}(K)$, satisfying*
$$g_i(x_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Proof. Since $\sum_{i,j=1}^{n} \bar{\alpha}_i \alpha_j K(x_i, x_j) = \| \sum_{j=1}^{n} \alpha_j k_{x_j} \|^2$, the equivalence of (1) and (2) follow.

Note that $\bar{\alpha}_1 k_{x_1} + ... + \bar{\alpha}_n k_{x_n} = 0$ if and only if $\langle f, \sum_{i=1}^{n} \alpha_i k_{x_i} \rangle = 0$ for all $f \in \mathcal{H}$ if and only if $\bar{\alpha}_1 f(x_1) + ... + \bar{\alpha}_n f(x_n) = 0$ for all $f \in \mathcal{H}$. Hence, (2) and (3) are equivalent. Thus, (1),(2) and (3) are equivalent.

To see that statement (4) implies (3), assume that $\alpha_i \neq 0$ and note that we can choose $f = g_i$.

Finally, if (1) holds, then for each i applying Corollary 3.7.9 to the set of numbers $\alpha_j = 0, j \neq i$ and $\alpha_i = 1$, yields the function g_i. □

A set of functions satisfying statement (4) is often called a *Partition of Unity* for the $x_1, x_2, ..., x_n$. Note that if one has a partition of unity for $x_1, x_2, ..., x_n$ then one gets $f \in \mathcal{H}(K)$ satisfying $f(x_i) = \lambda_i$ simply by setting, $f = \lambda_1 g_1 + ... + \lambda_n g_n$.

Definition 5.1.12. *An RKHS $\mathcal{H}$ on X that satisfies any of the equivalent conditions of the above theorem is called* ***Fully Interpolating****.*

We will now show that there is a way to compute a partition of unity it exists. Assume that $F = \{x_1, x_2, ..., x_n\}$ and $P = (K(x_i, x_j))$ is invertible as in the above corollary and write $P^{-1} = (b_{i,j}) = B$. Let $e_j, j = 1, ..., n$ be solutions to $e_j = P w_j$. Thus, if we set

$$g_j = \sum_i b_{i,j} k_{x_i}.$$

then $g_j(x_i) = \delta_{i,j}$,where $\delta_{i,j}$ denotes the Dirac Delta function. Thus, these functions are a partition of unity for F. Moreover, we consider

$$g = \sum_j \lambda_j g_j$$

then g is the unique function in $\mathcal{H}_F$ satisfying $g(x_i) = \lambda_i, i = 1, 2, ..., n$. Hence, g is also the function of minimum norm interpolating these values. Thus, this particular partition of unity gives an easy means for producing the minimum norm interpolant for the given set.

For this reason, this particular partition of unity is well worth computing and is called the *Canonical Partition of Unity* for F.

5.1.13 Best least squares approximants

If $\mathcal{H}$ is an RKHS on X, $F = \{x_1, x_2, ..., x_n\}$ is a finite set of distinct points and $\{\lambda_1, ..., \lambda_n\} \subseteq \mathbb{C}$, and the matrix $(K(x_i, x_j))$ is not invertible, then there might not exist any function $f \in \mathcal{H}$ with $f(x_i) = \lambda_i$, for all i. In these cases one is often interested in finding a function such that the *Least Square Error*,

$$J(f) = \sum_{i=1}^{n} |f(x_i) - \lambda_i|^2$$

is minimized and then, among all such solutions, find the function of minimum norm. As we shall see there is a unique such function and it is called the *Best Least Squares Approximant.*

The following theorem proves the existence of the best least squares approximant and gives a formula for obtaining this function.

Theorem 5.1.14. *Let $\mathcal{H}$ be an RKHS on X with kernel K, $\{x_1, ..., x_n\} \subseteq X$ be a finite set of distinct points, $v = (\lambda_1, ..., \lambda_n)^t \in \mathbb{C}^n$ and $Q = (K(x_i, x_j))$. Then there exists a vector $w = (\alpha_1, ..., \alpha_n)^t$ such that $(v - Qw) \in \mathcal{N}(Q)$. If we assume*

$$g = \alpha_1 k_{x_1} + ... + \alpha_n k_{x_n}.$$

then g minimizes the least square error and among all functions in $\mathcal{H}$ that minimize the least square error, g is the unique function of minimum norm.

Proof. Note that for any function $f \in \mathcal{H}$ there exists a vector $w \in \mathbb{C}^n$ such that $Qw = (f(x_1), ..., f(x_n))^t$. Hence we have

$$J(f) = \sum_{i=1}^{n} |f(x_i) - \lambda_i|^2 = \|Qw - v\|^2.$$

This is minimized for any vector $w = (\alpha_1, ..., \alpha_n)^t$ such that $Qw = P_{\mathcal{R}(Q)}(v) = v_1$ where $\mathcal{R}(Q)$ denotes the range of the matrix Q and $P_{\mathcal{R}(Q)}$ denotes the orthogonal projection onto $\mathbb{R}(Q)$.

If we choose a different vector $w' = (\alpha'_1, ..., \alpha'_n)$ that solves $v_1 = Qw'$ then we will have $\sum_{i=1}^{n} \alpha_i k_{x_i} = \sum_{i=1}^{n} \alpha'_i k_{x_i}$, since $w - w' \in \mathcal{N}(Q)$. Since projecting a function f onto the span of the kernel functions $k_{x_1}, ..., k_{x_n}$ decreases the norm and does not change the value of f at points $x_1, ..., x_n$ we see that g is the unique minimizer of J of smallest norm. □

5.1.15 The elements of $\mathcal{H}(K)$

We will now use the interpolation theory to give a general solution to the reconstruction problem for kernels. We will characterize the functions $f : X \to \mathbb{C}$ that belong to the RKHS $\mathcal{H}(K)$ determined by a kernel K.

Definition 5.1.16. *A **Directed Set** is any set with a partial order that has the property that, given any two elements of the set, there is always at least one element of the set that is greater than or equal to both.*

Given a set X, we let $\mathcal{F}_X$ denote the collection of all finite subsets of X. The set $\mathcal{F}_X$ is a directed set with respect to the partial order given by inclusion. Setting $F_1 \leq F_2$ if and only if $F_1 \subseteq F_2$ defines a partial order on $\mathcal{F}_X$. Given any two finite sets, F_1, F_2, there is always a third finite set G that is larger than both. In particular, we could take $G = F_1 \cup F_2$. Thus, $\mathcal{F}_X$ is a directed set.

Definition 5.1.17. *A **net** is a generalization of the concept of a sequence, but it is indexed by an arbitratry directed set. So, if $(\mathcal{F}, \leq)$ is a directed set,then a net in a Hilbert Space $\mathcal{H}$ is just a collection of vectors $g_{F\,F\in\mathcal{F}} \subseteq \mathcal{H}$. Convergence of nets is defined by analogy with convergence of sequences. The net $g_{F\,F\in\mathcal{F}}$ is said to converge to $g \in \mathcal{H}$, provided that for every $\epsilon > 0$ there is $F_0 \in \mathcal{F}$ such that whenever $F_0 \leq F$, then $\|g - g_F\| < \epsilon$.*

Proposition 5.1.18. *Let $\mathcal{H}$ be an RKHS on the set X, let $g \in \mathcal{H}$ and for each finite set $F \subseteq X$, let $g_F = P_F(g)$, where P_F denotes the orthogonal projection of $\mathcal{H}$ onto $\mathcal{H}_F$. Then the net $g_{F\,F\in\mathcal{F}_X}$ converges in norm to g.*

Proof. Let $K(x, y)$ denote the reproducing kernel for $\mathcal{H}$ and let $k_y(.) = K(., y)$. Given $\epsilon > 0$, by Proposition 4.1.1, there exists a finite collection of points, $F_0 = \{x_1, x_2, ..., x_n\}$ and scalars $\{\alpha_1, ..., \alpha_n\}$, such that $\|g - \sum_i \alpha_i k_{x_i}\| < \epsilon$.

Since g_{F_0} is the closest point in $\mathcal{H}_{F_0}$ to g, we have that $\|g - g_{F_0}\| < \epsilon$. Now let F be any finite set, with $F_0 \subseteq F$. Then $\mathcal{H}_{F_0} \subseteq \mathcal{H}_F$ and since g_F is the closest point in $\mathcal{H}_F$ to g and $g_{F_0} \in \mathcal{H}_F$, we have that $\|g - g_F\| \leq \|g - g_{F_0}\| < \epsilon$, for every $F_0 \subseteq F$, and the result follows. ☐

Proposition 5.1.19. *Let $P \geq 0$ be an $n \times n$ matrix, and let $x = (x_1, x_2, ..., x_n)^t$ be a vector in $\mathbb{C}^n$. If $xx^* = (x_i \bar{x}_j) \leq cP$, for some scalar, $c > 0$, then $x \in \mathcal{R}(P)$. Moreover, if y is any vector such that $x = Py$, then $0 \leq \langle x, y \rangle \leq c$.*

Proof. For any matrix A we have $\mathcal{N}(A^*) = \mathcal{R}(A)^\perp$. Since $P = P^*$ we have $\mathcal{N}(P) = \mathcal{R}(P)^\perp$. Thus, we may write $x = v + w$ with $v \in \mathcal{R}(P)$ and $w \in \mathcal{N}(P)$.

Now, $\langle x, w \rangle = \langle w, w \rangle$, and hence $\|w\|^4 = \langle w, x \rangle \langle x, w \rangle = \sum_{i,j} \bar{x}_j w_j x_i \bar{w}_i = \langle (x_i, \bar{x}_j) w, w \rangle \leq \langle cPw, w \rangle = 0$, since $Pw = 0$.

This inequality shows that $w = 0$ and hence, $x = v \in \mathcal{R}(P)$. Now if we write $x = Py$, then $\langle x, y \rangle = \langle Py, y \rangle \geq 0$. As above, we have that $\langle x, y \rangle^2 = \langle y, x \rangle \langle x, y \rangle = \langle (x_i, \bar{x}_j) y, y \rangle = c\langle x, y \rangle$. Canceling one factor of $\langle x, y \rangle$ from this inequality yields the result.

☐

We are now able to prove a theorem that characterizes the functions that belong to an RKHS in terms of the reproducing kernel.

Theorem 5.1.20. *Let $\mathcal{H}$ be an RKHS on X with reproducing kernel K and let $f : X \to \mathbb{C}$ be a function. Then tbe following are equivalent:*

1. *$f \in \mathcal{H}$*
2. *There exists a constant, $c \geq 0$, such that for every finite subset, $F = \{x_1, x_2, ..., x_n\} \subseteq X$, there exists a function $h \in \mathcal{H}$ with $\|h\| \leq c$ and $f(x_i) = h(x_i), i = 1, ..., n$;*
3. *there exists a constant, $c \geq 0$, such that the function, $c^2 K(x,y) - f(x)\overline{f(y)}$ is a kernel function.*

Moreover, if $f \in \mathcal{H}$ then $\|f\|$ is the least c that satisfies the inequalities in (2) and (3).

Proof. (1) implies (3): Let $F = \{x_1, ..., x_n\} \subseteq X$, let $\alpha_1, ..., \alpha_n$ be scalars and set $g = \sum_j \alpha_j k_{x_j}$. Then

$$\sum_j \bar{\alpha}_i \alpha_j f(x_j)\overline{f(x_j)} = |\sum_i \overline{\alpha}_i f(x_i)|^2 = |\langle f, g\rangle| \leq \|f\|^2 \|g\|^2 = \|f\|^2 \sum_{i,j} \overline{\alpha}_i \alpha_j K(x_i, x_j).$$

Since the choice of the scalars was arbitrary, we have that $(f(x_i)\overline{f(x_j)}) \leq \|f\|^2 (K(x_i, x_j))$ and so (3) follows with $c = \|f\|$.

(3) implies (2): Let $F = \{x_1, ..., x_n\} \subseteq X$ be a finite set. Apply Proposition 5.1.19 to deduce that the vector v whose entries are $\lambda_i = f(x_i)$ is in the range of $(K(x_i, x_j))$. Then use the Interpolation Theorem to deduce that there exists $h = \sum_i \alpha_i k_{x_i}$ in $\mathcal{H}_F$ with $h(x_i) = f(x_i)$. Let w denote the vector whose components are the α_i's and it follows that $\|h\|^2 = \langle v, w\rangle \leq c^2$ by applying Proposition 5.1.19 again.

(2) implies (1): By assumption, for every finite set F there exists $h_F \in \mathcal{H}$ such that $\|h_F\| \leq c$ and $h_F(x) = f(x)$ for every $x \in F$. Set $g_F = P_F(h_F)$, then $g_F(x) = h_F(x) = f(x)$ for every $x \in F$ and $\|g_F\| \leq \|h_F\| \leq c$.

We claim that the net $\{g_F\} F \in \mathcal{F}_X$ is Cauchy and converges to f.

To see that the net is Cauchy, let $M = \sup \|g_F\| \leq c$ and fix $\epsilon > 0$. Choose a set F_0 such that $M - \epsilon^2 < \|g_{F_0}\|$ and hence, $\langle (g_F - g_{F_0}), g_{F_0}\rangle = 0$. Hence, $\|g_F\|^2 = \|g_{F_0}\|^2 + \|g_F - g_{F_0}\|^2$, and so $M - \epsilon^2 \leq \|g_{F_0}\| \leq \|g_F\| \leq M$.

Therefore, $0 \leq \|g_F\| - \|g_{F_0}\| \leq \epsilon^2$, and we have that $\|g_F - g_{F_0}\|^2 = \|g_F\|^2 - \|g_{F_0}\|^2 = (\|g_F\| + \|g_{F_0}\|)(\|g_F\| - \|g_{F_0}\|) \leq 2M\epsilon^2$. Thus, $\|g_F - g_{F_0}\| < \sqrt{2M}\epsilon$ and so for any $F_1, F_2 \in \mathcal{F}_X$ with $F_0 \subseteq F_1, F_0 \subseteq F_2$, it follows that $\|g_{F_1} - g_{F_2}\| < 2\sqrt{2M}\epsilon$ and we have proven that the net is Cauchy.

Thus, there is a function $g \in \mathcal{H}$ that is the limit of this net and hence, $\|g\| \leq M \leq c$. But since any norm convergent net also converges point wise,we have that $g(x) = f(x)$ for any x. Thus, the proof that (2) implies (1) is complete.

Finally, given that $f \in \mathcal{H}$ we have that the conditions of (2) and (3) are met for $c = \|f\|$. So the least c that meets these conditions is less than $\|f\|$. Conversely, in the proof that (3) implies (2), we saw that any c that satisfies (3) satisfies (2). But in the proof that (2) implies (1) we saw $\|f\| \leq c$. Hence many c that meets the inequalities in (2) or (3) must be greater than $\|f\|$. □

Corollary 5.1.21. *Let $f : D \to \mathbb{C}$ be a function. Then f is analytic on $\mathbb{D}$ and has a square summable power series if and only if there exists $c > 0$ such that $K(z, w) = \dfrac{c^2}{1 - z\bar{w}} - f(z)\overline{w}$ is kernel function on $\mathbb{D}$.*

Bibliography

[1] R.Bhatia: Notes on Functional Analysis.

[2] Vern Paulsen: Notes on Reproducing Kernel Hilbert Spaces.