AN INTRODUCTION TO THE THEORY OF REPRODUCING KERNEL HILBERT SPACES

VERN I. PAULSEN

Abstract. These notes give an introduction to the theory of reproducing kernel Hilbert spaces and their multipliers. We begin with the material that is contained in Aronszajn's classic paper on the subject. We take a somewhat algebraic view of some of his results and discuss them in the context of pull-back and push-out constructions. We then prove Schoenberg's results on negative definite functions and his characterization of metric spaces that can be embedded isometrically in Hilbert spaces. Following this we study multipliers of reproducing kernel Hilbert spaces and use them to give classic proofs of Pick's and Nevanlinna's theorems. In later chapters we consider generalizations of this theory to the vector-valued setting.

1. Introduction

These are the lecture notes for a course on reproducing kernel Hilbert spaces first given at the University of Houston in the Spring of 2006.

Reproducing kernel Hilbert spaces arise in a number of areas, including approximation theory, statistics, machine learning theory, group representation theory and various areas of complex analysis.

2. Definitions and Basic Examples

We will consider Hilbert spaces over either the field of real numbers, $\mathbb{R}$, or of 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 result that is true for either the real or complex numbers, we will use $\mathbb{F}$.

Given a set X, if we equip the set of all functions from X to $\mathbb{F}, \mathcal{F}(X, \mathbb{F})$ with the usual operations of addition, $(f+g)(x) = f(x) + g(x)$, and scalar multiplication, $(\lambda \cdot f)(x) = \lambda \cdot (f(x))$, then $\mathcal{F}(X, \mathbb{F})$ is a vector space over $\mathbb{F}$.

Definition 2.1. *Given a set X, we will say that $\mathcal{H}$ is a* **reproducing kernel Hilbert space(RKHS)** *on X over $\mathbb{F}$, provided that:*

(i) *$\mathcal{H}$ is a vector subspace of $\mathcal{F}(X, \mathbb{F})$,*
(ii) *$\mathcal{H}$ is endowed with an inner product, $\langle , \rangle$, making it into a Hilbert space,*

Date: September 30, 2009.
Key words and phrases. positive definite, reproducing kernel.
Research supported in part by a grant from the NSF.

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

If $\mathcal{H}$ is a RKHS on X, then since every bounded linear functional is given by the inner product with a unique vector in $\mathcal{H}$, we have that for every $y \in X$, there exists a unique vector, $k_y \in \mathcal{H}$, such that for every $f \in \mathcal{H}$, $f(y) = \langle f, k_y \rangle$.

Definition 2.2. *The function k_y is called the* **reproducing kernel for the point** *y. The 2-variable function defined by*

$$K(x, y) = k_y(x)$$

is called the **reproducing kernel for** $\mathcal{H}$.

Note that we have,

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

and

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

Problem 2.3. *Show that if $\mathcal{H}$ is a reproducing kernel Hilbert space on X with reproducing kernel $K(x, y)$, then $K(y, x) = \overline{K(x, y)}$.*

Definition 2.4. *Let $\mathcal{H}$ be a reproducing kernel Hilbert space on X. We say that $\mathcal{H}$* **separates points** *provided that for $x \neq y$ there exists $f \in \mathcal{H}$ such that $f(x) \neq f(y)$.*

Problem 2.5. *Let $\mathcal{H}$ be a reproducing kernel Hilbert space on X. Prove that setting $d(x, y) = \sup\{|f(x) - f(y)| : f \in \mathcal{H}, \|f\| \leq 1\}$ defines a metric on X if and only if $\mathcal{H}$ separates points. Give a formula for $d(x, y)$ in terms of the reproducing kernel.*

Problem 2.6. *Show that if $\mathcal{H}$ is a RKHS on X and $\mathcal{H}_0 \subseteq \mathcal{H}$ is a closed, subspace, then $\mathcal{H}_0$ is also a RKHS on X. Prove that the reproducing kernel for $\mathcal{H}_0$ for a point y is the function $P_0(k_y)$ where k_y is the reproducing kernel function for $\mathcal{H}$ and $P_0 : \mathcal{H} \to \mathcal{H}_0$ denotes the orthogonal projection of $\mathcal{H}$ onto $\mathcal{H}_0$.*

Thus, there is a reproducing kernel, $K_0(x, y)$, for the subspace $\mathcal{H}_0$. One of the problems that we shall study later is determining the relationship between $K_0(x, y)$ and the reproducing kernel $K(x, y)$ for the whole space. We will see that although it is fairly easy to write down some general theorems about this relationship, computing specific examples is much trickier.

Sometimes to fix ideas it helps to look at a non-example. Suppose that we take the continuous functions on [0,1], $C([0, 1])$, define the usual 2-norm on this space, i.e., $\|f\|^2 = \int_0^1 |f(t)|^2 dt$, and complete to get the Hilbert space, $L^2[0, 1]$. Given any point $x \in [0, 1]$ it is easy to construct a sequence, $f_n \in C([0, 1])$, such that $\lim_n \|f_n\| = 0$, and $\lim_n f_n(x) = +\infty$. Thus, there is no way to extend the values of functions at points in $C([0, 1])$ to regard functions in $L^2[0,1]$ as having values at points. This is just another way to see that we can't think of elements of $L^2[0,1]$ as functions, in particular, it is not a RKHS on [0,1]. One of the remarkable successes of measure theory is showing that this completion can be regarded as equivalences of functions, modulo sets of measure 0.

Thus, reproducing kernel Hilbert spaces are quite different from L^2-spaces.

We now look at a few key examples.

The Hardy Space of the Unit Disk, $H^2(\mathbb{D})$.

This space plays a key role in function theoretic operator theory.

To construct $H^2(\mathbb{D})$, we first consider formal complex power series, $f \sim \sum_{n=0}^{\infty} a_n z^n, g \sim \sum_{n=0}^{\infty} b_n z^n$ and endow them with the inner product, $\langle f, g\rangle = \sum_{n=0}^{\infty} a_n \bar{b_n}$. Thus, we have that $\|f\|^2 = \sum_{n=0}^{\infty} |a_n|^2$. The map $L : H^2(\mathbb{D}) \to \ell^2(\mathbb{N})$ defined by $L(f) = (a_0, a_1, \ldots)$ is a linear inner product preserving isomorphism and hence we see that $H^2(\mathbb{D})$ can be identified with the Hilbert space, $\ell^2(\mathbb{Z}^+)$, where $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$ the natural numbers, $\mathbb{N}$, together with 0, and hence is itself a Hilbert space. Thus, we see that (ii) of the above definition is met.

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

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

Thus, each power series defines a function on $\mathbb{D}$ and the vector space operations on formal power series, clearly agrees with their vector space operations as functions on $\mathbb{D}$, and so (i) is met.

The above inequality also shows that the map, E_z is bounded with $\|E_z\| \leq \frac{1}{\sqrt{1-|z|^2}}$ and so $H^2(\mathbb{D})$ is a RKHS on $\mathbb{D}$.

To compute the kernel, for a point $w \in \mathbb{D}$, 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.

Finally, we can compute, $\|E_z\| = K(z, z) = \frac{1}{\sqrt{1-|z|^2}}$, so that the above inequality was sharp.

Problem 2.7. *Let $\mathcal{H}_N \subseteq H^2(\mathbb{D})$ denote the subspace consisting of all functions of the form $f(z) = \sum_{n=N}^{\infty} a_n z^n$.Find the reproducing kernel for $\mathcal{H}_N$.*

Sobolev Spaces on [0,1]

These are very simple examples of the types of Hilbert spaces that arise in differential equations.

Let $\mathcal{H} = \{f : [0,1] \to \mathbb{R} : \text{ f is absolutely continuous}, f(0) = f(1) = 0, f' \in L^2[0,1]\}$. Recall that if a function is absolutely continuous, then it is differentiable almost everywhere and is equal to the integral of its derivative. Clearly, $\mathcal{H}$ is a vector space of functions on [0,1].

We endow $\mathcal{H}$ with the non-negative, sesquilinear form, $\langle f, g\rangle = \int_0^1 f'(t)g'(t)dt$.

Since f is absolutely continuous and $f(0) = 0$, for any $0 \le x \le 1$, we have that $f(x) = \int_0^x f'(t)dt = \int_0^1 f'(t)\chi_{[0,x]}(t)dt$. Thus, by the Cauchy-Schwartz inequality,

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

This last inequality shows that $\|f\| = 0$ if and only if $f = 0$. Thus, $\langle , \rangle$ is an inner product on $\mathcal{H}$ and that for every $x \in [0,1], E_x$ is bounded with $\|E_x\| \le \sqrt{x}$.

All that remains to show that $\mathcal{H}$ is a RKHS is to show that it is complete. If $\{f_n\}$ is a Cauchy sequence in this norm, then $\{f_n'\}$ is Cauchy in $L^2[0,1]$ and hence there exists $g \in L^2[0,1]$ that this sequence converges to. By the above inequality, $\{f_n\}$ must be pointwise Cauchy and hence we may define a function by setting $f(x) = \lim_n f_n(x)$. Since, $f(x) = \lim_n f_n(x) = \lim_n \int_0^x f_n'(t)dt = \int_0^x g(t)dt$, it follows that f is absolutely continuous and that $f' = g, a.e.$ and hence, $f' \in L^2[0,1]$. Finally, $f(0) = \lim_n f_n(0) = 0 = \lim_n f_n(1) = f(1)$, and so $f \in \mathcal{H}$.

Thus, $\mathcal{H}$ is a RKHS on [0,1].

It remains to find the kernel function. To do this we first formally solve a differential equation and then show that the function we obtain by this formal solution, belongs to $\mathcal{H}$. To find $k_y(t)$, we apply integration by parts to see that, $f(y) = \langle f, k_y\rangle = \int_0^1 f'(t)k_y'(t)dt = f(t)k_y'(t) - \int_0^1 f(t)k_y''(t)dt = -\int_0^1 f(t)k_y''(t)dt$.

If we 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

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

After formally obtaining this solution, it can now be checked that it indeed satisfies the necessary equations to be the reproducing kernel for $\mathcal{H}$.

Note that, $\|E_y\|^2 = \|k_y\|^2 = K(y,y) = y(1-y)$, which is a better estimate than obtained above.

Problem 2.8. *Let $\mathcal{H}$ be the same set of functions as in the above problem, but define a new inner product by $\langle f, g\rangle = \int_0^1 f(t)g(t) + f'(t)g'(t)dt$. Prove that $\mathcal{H}$ is still a Hilbert space in this new inner product, show that the kernel function is the formal solution to $-k_y'' + k_y = \delta_y, k_y(0) = k_y(1) = 0$ and find $K(x,y)$.*

Problem 2.9. *Let $\mathcal{H}_1 = \{f : [0,1] \to \mathbb{R} : f$ is absolutely continuous,$f' \in L^2[0,1], f(0) = f(1)\}$ and set $\langle f, g\rangle = f(0)g(0) + \int_0^1 f'(t)g'(t)dt$. Prove that $\mathcal{H}_1$ is a RKHS, show that the kernel function is the formal solution to $-k_y'' = \delta_y, k_y'(1) - k_y'(0) + k_y(0) = 0, k_y(0) = k_y(1)$ and that the kernel, $K(x,y) = K_0(x,y) + 1$, where $K_0(x,y)$ denotes the kernel of the last example. Note that $\mathcal{H}_1$ is equal to the span of $\mathcal{H}$ and the constant functions.*

Bergman Spaces on Complex Domains

Let $G \subset \mathbb{C}$ be open and connected. We let

$$B^2(G) = \{f : G \to \mathbb{C} | f \text{is analytic on G and} \int\int_G |f(x+iy)|^2 dxdy < +\infty\},$$

where $dxdy$ denotes area measure. We define a sesquilinear form on $B^2(G)$ by $\langle f, g\rangle = \int\int_G f(x+iy)\overline{g(x+iy)}dxdy$. It is easily seen that this defines an inner product on $B^2(G)$, so that $B^2(G)$ is an inner product space.

Theorem 2.10. *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 $w, B(w;R)^-$, is contained in G, then by Cauchy's integral formula for any $0 \le r \le R$, we have $f(w) = \frac{1}{2\pi}\int_0^{2\pi} f(w + re^{i\theta})d\theta$.

Therefore,

$$f(w) = \frac{1}{\pi R^2}\int_0^R r(2\pi f(w))dr = \\ \frac{1}{\pi R^2}\int_0^R r\int_0^{2\pi} f(w+re^{i\theta})d\theta = \frac{1}{\pi R^2}\int\int_{B(w;R)} f(x+iy)dxdy.$$

Thus, by Cauchy-Schwartz, it follows that $|f(w)| \le \frac{1}{\pi R^2}\|f\|(\int\int_{B(w;R)} dxdy)^{1/2} = \frac{1}{R\sqrt{\pi}}\|f\|$.

This proves that for $w \in G$ the evaluation functional is bounded. So all that remains to prove that $B^2(G)$ is a RKHS is to show that $B^2(G)$ is complete in this norm. To this end let $\{f_n\} \subseteq B^2(G)$ be a Cauchy sequence. For any $w \in G$ pick R as above and pick $0 < \delta < d(B(w;R), G^c)$, where $d(\cdot,\cdot)$ denotes the distance between the two 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 is 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 convergent on every closed ball contained in G. If we let $f(z) = \lim_n f_n(z)$ denote the pointwise limit of this sequence, then we have that $\{f_n\}$ converges uniformly to f on each closed ball contained in G and so by Montel's theorem f is analytic.

Since $B^2(G) \subseteq L^2(G)$ 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 $h(z) = \lim_k f_{n_k}(z)$ a.e., but this implies that $h(z) = f(z)$ a.e. and so $\|f - f_n\|_2 \to 0$. Thus, $f \in B^2(G)$ and so $B^2(G)$ is complete. $\square$

Definition 2.11. *Given any open connected subset $G \subseteq \mathbb{C}$, the reproducing kernel for $B^2(G)$ is called the* **Bergman kernel** *for the G.*

The result that we proved above extends to open connected susbets of $\mathbb{C}^n$ and to many complex manifolds. Knowledge of the Bergman kernel of such domains has many applications in complex analysis and the study of this kernel is still an active area of research.

Note also that the above inequality shows that, $B^2(\mathbb{C}) = (0)$, since in this case R could be taken arbitrarily large, 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.

Problem 2.12. *Let $U = \{x + iy : 0 < x < +\infty, 0 < y < 1\}$. Give an example of a non-zero function in $B^2(U)$.*

When $A = area(G) < +\infty$, then the constant function 1 is in $B^2(G)$ and $\|1\| = \sqrt{A}$. In this case it is natural to re-normalize so that $\|1\| = 1$, to do this we just re-define the inner product to be,

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

Often when books refer to the Bergman space on such a domain they mean this *normalized* Bergman space. We shall adopt this convention too. 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,

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

Problem 2.13. *Show that the Bergman kernel for $B^2(\mathbb{D})$ is given by $K(z, w) = \frac{1}{(1-z\bar{w})^2}$.*

Weighted Hardy Spaces

We now look at another family of reproducing kernel Hilbert spaces. Given a sequence $\beta = \{\beta_n\}$, with $\beta_n > 0$, consider the space of all formal power series, $f(z) = \sum_{n=0}^{\infty} a_n z^n$, such that the norm,

$$\|f\|_\beta^2 = \sum_{n=0}^{\infty} \beta_n^2 |a_n|^2$$

is finite. This is a Hilbert space with inner product, $\langle f, g\rangle = \sum_{n=0}^{\infty} \beta_n^2 a_n \bar{b_n}$, where $f(z)$ is as above and $g(z) = \sum_{n=0}^{\infty} b_n z^n$. This Hilbert space is denoted H^2_β and is called a **weighted Hardy space.**

Thus, the usual Hardy space is the weighted Hardy space corresponding to all weights $\beta_n = 1$.

A power series $f(z) = \sum_{n=0}^{\infty} a_n z^n$ in H^2_β will satisfy $\lim_n \beta_n |a_n| = 0$. Hence, for n sufficiently large, we have that $|a_n| \leq \beta_n^{-1}$. Thus, the radius of convergence R_f of f satisfies, $R_f^{-1} = \limsup_n |a_n|^{1/n} \leq \limsup_n \beta_n^{-1/n} = (\liminf_n \beta_n^{1/n})^{-1}$. Hence, f will have radius of convergence greater than,

$$\liminf_{n\to\infty} (\beta_n)^{-1/n} \equiv R.$$

Thus, provided $R > 0$, every function in H^2_β will converge to define an analytic function on the disk of radius R and H^2_β can be viewed as a space of analytic functions on this disk.

It is easy to see that for any $|w| < R, f(w) = \langle f, k_w\rangle$ where

$$k_w(z) = \sum_{n=0}^{\infty} \frac{\bar{w}^n z^n}{\beta_n^2},$$

is in H^2_β. Thus, given the constraint on the sequence β that $R > 0$, we see that H^2_β is a RKHS on the disk of radius R, with reproducing kernel, $K_\beta(z, w) = k_w(z)$.

Problem 2.14. *Show that $B^2(\mathbb{D})$ is a weighted Hardy space.*

In addition to the Hardy space and Bergman space of the disk, another widely studied weighted Hardy spaces is the *Segal-Bargmann space,* which is the weighted Hardy space that one obtains by taking weights $\beta_n = \sqrt{n!}$. Since $\liminf_{n\to\infty} (n!)^{-1/(2n)} = +\infty$, this is a space of entire functions and the reproducing kernel for this space is easily computed to be $K(z, w) = e^{z\overline{w}}$.

In the definition of weighted Hardy space it is not really necessary to demand that every $\beta_n > 0$. Instead one considers a space of power series such that $\beta_n = 0$ implies that the coefficient of z^n is 0.

The *Dirichlet space* is an example of this type of weighted Hardy space. The Dirichlet space is obtained by taking weights $\beta_n = \sqrt{n}$, so that every function in the space is 0 at 0. The resulting space of power series is seen to define functions that are analytic on the unit disk and has a reproducing kernel given by $K(z, w) = log(1 - z\overline{w})$, where the branch of the logarithm is defined by taking the negative real axis as a cut line.

Multi-Variable Examples

Given a natural number n, by a **multi-index** we mean a point, $I = (i_1, \ldots, i_n) \in (\mathbb{Z}^+)^n$. Given $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$, we set $z^I = z_1^{i_1} \cdots z_n^{i_n}$. By a **power series in n variables**, we mean a formal expression of the form $f(z) = \sum_{I\in(\mathbb{Z}^+)^n} a_I z^I$, where $a_I \in \mathbb{C}$ are called the **coefficients** of f.

We define the **n-variable Hardy space,** $H^2(\mathbb{D}^n)$ to be the set of all power series, $f \sim \sum_{I\in(\mathbb{Z}^+)^n} a_I z^I$, such that $\|f\|^2 = \sum_{I\in(\mathbb{Z}^+)^n} |a_I|^2 < +\infty$.

Reasoning as for the one-variable Hardy space one can see that for each $z \in \mathbb{D}^n$ the power series converges and defines an analytic function, $f(z)$ and that $H^2(\mathbb{D}^n)$ is a RKHS on $\mathbb{D}^n$ with kernel given by

$$K(z,w) = \sum_{I\in(\mathbb{Z}^+)^n} \bar{w}^I z^I = \prod_{i=1}^{n} \frac{1}{1-\bar{w}_i z_i}.$$

Similarly, we can define **multi-variable Bergman spaces,** $B^2(G)$ for $G \subset \mathbb{C}^n$ an open connected subset by using 2n-dimensional Lebesgue measure. As in the one variable case, if the Lebesgue measure of G is finite then one often uses normalized Lebesgue measure to define the norm on $B^2(G)$, so that the constant function 1 has norm 1.

Problem 2.15. *Find the reproducing kernels for the space $B^2(\mathbb{D}^2)$ with respect to the normalized and ordinary Lebesgue measure.*

2.1. **The Complexification of a RKHS of Real-Valued Functions.** Let $\mathcal{H}$ be a RKHS of real-valued functions on the set X with reproducing kernel, $K(x,y)$. Let $\mathcal{W} = \{f_1 + if_2 : f_1, f_2 \in \mathcal{H}\}$, which is a vector space of complex-valued functions on X. If we set,

$$\langle f_1 + if_2, g_1 + ig_2\rangle_{\mathcal{W}} = \langle f_1, g_1\rangle_{\mathcal{H}} + i\langle f_2, g_1\rangle_{\mathcal{H}} - i\langle f_1, g_2\rangle_{\mathcal{H}} + \langle f_2, g_2\rangle_{\mathcal{H}},$$

then it is easily checked that this defines an inner product on $\mathcal{W}$, with corresponding norm,

$$\|f_1 + if_2\|^2_{\mathcal{W}} = \|f_1\|^2_{\mathcal{H}} + \|f_2\|^2_{\mathcal{H}}.$$

Hence, $\mathcal{W}$ is a Hilbert space and since,

$$f_1(y) + if_2(y) = \langle f_1 + if_2, k_y\rangle_{\mathcal{W}},$$

we have that $\mathcal{W}$ equipped with this inner product is a RKHS of complex-valued functions on X with reproducing kernel, $K(x,y)$.

We call $\mathcal{W}$ the **complexification of $\mathcal{H}$.** Since every real-valued RKHS can be complexified in a way that still preserves the reproducing kernel, we shall from this point on, only consider the case of complex-valued reproducing kernel Hilbert spaces.

3. General Theory

Let X be any set and let $\mathcal{H}$ be a RKHS on X with kernel K. In this section we will show that K completely determines the space $\mathcal{H}$ and characterize the functions that are the kernel functions of some RKHS.

Proposition 3.1. *Let $\mathcal{H}$ be a RKHS on the set X with kernel K. Then the linear span of the functions, $k_y(\cdot) = K(\cdot, y)$ 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 $f = 0$. $\square$

Lemma 3.2. *Let $\mathcal{H}$ be a RKHS on X and let $\{f_n\} \subseteq \mathcal{H}$. If $\lim_n \|f_n - f\| = 0$, then $f(x) = \lim_n f_n(x)$ for every $x \in X$.*

Proof. We have that $|f_n(x) - f(x)| = |\langle f_n - f, k_x \rangle| \leq \|f_n - f\| \|k_x\| \to 0$. $\square$

Proposition 3.3. *Let $\mathcal{H}_i, i = 1, 2$ be RKHS's on X with kernels, $K_i(x, y), i = 1, 2$. 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 let $\mathcal{W}_i = span\{k_x \in \mathcal{H}_i : x \in X\}, i = 1, 2$. By the above result, $\mathcal{W}_i$ is dense in $\mathcal{H}_i, i = 1, 2$. Note that for any $f \in \mathcal{W}_i$, we have that $f(x) = \sum_j \alpha_j k_{x_j}(x)$ and so it's values as a function 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 $f \in \mathcal{W}_1 = \mathcal{W}_2$.

Finally, if $f \in \mathcal{H}_1$, then there exists a sequence of functions, $\{f_n\} \subseteq \mathcal{W}_1$ with $\|f - f_n\|_1 \to 0$. Since, $\{f_n\}$ is Cauchy in $\mathcal{W}_1$ it is also Cauchy in $\mathcal{W}_2$, so there exists $g \in \mathcal{H}_2$ with $\|g - f_n\|_2 \to 0$. By the above Lemma, $f(x) = \lim_n f_n(x) = g(x)$. 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, since $\|f\|_1 = \|f\|_2$ for every f in a dense subset, we have that the norms are equal for every f. $\square$

We now look at another consequence of the above Lemma. This result gives another means of calculating the kernel for a RKHS that is very useful.

Recall that given vectors $\{h_s : s \in S\}$ in a normed space $\mathcal{H}$, indexed by an arbitrary set S. We say that $h = \sum_{s \in S} h_s$ provided that for every $\epsilon > 0$, there exists a finite subset $F_0 \subseteq S$, such that for any finite set $F, F_0 \subseteq F \subseteq S$, we have that $\|h - \sum_{s \in F} h_s\| < \epsilon$. Two examples of this type of convergence are given by the two **Parseval identities.** When $\{e_s : s \in S\}$ is an orthonormal basis for a Hilbert space, $\mathcal{H}$, then for any $h \in \mathcal{H}$, we have that

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

and

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

Note that these types of sums do not need S to be an ordered set to be defined. Perhaps, the key example to keep in mind is that if we set $a_n = \frac{(-1)^n}{n}, n \in \mathbb{N}$ then the series, $\sum_{n=1}^{\infty} a_n$ converges, but $\sum_{n \in \mathbb{N}} a_n$ does not converge. In fact, for complex numbers, one can show that $\sum_{n \in \mathbb{N}} z_n$ converges if and only if $\sum_{n=1}^{\infty} |z_n|$ converges. Thus, this convergence is equivalent to absolute convergence in the complex case.

Theorem 3.4. *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 that $\langle k_y, e_s\rangle = \overline{\langle e_s, k_y\rangle} = \overline{e_s(y)}$. Hence, $k_y = \sum_{s\in S} \overline{e_s(y)} e_s$, where these sums converge in the norm on $\mathcal{H}$.

But since they converge in the norm, they converge at every point. Hence, $K(x,y) = k_y(x) = \sum_{s\in S} \overline{e_s(y)} e_s(x)$. □

For a quick example of this theorem, we can easily see that 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

$$\sum_{n\in\mathbb{Z}^+} e_n(z)\overline{e_n(w)} = \sum_{n=0}^{\infty} (z\overline{w})^n = \frac{1}{1 - z\overline{w}} \ .$$

Returning to our earlier example of a Sobolev space on[0,1], $\mathcal{H} = \{f : [0,1] \to \mathbb{R} :$ f is absolutely continuous,$f(0) = f(1) = 0, f' \in L^2[0,1]\}$, it is easily checked that for $n \neq 0$, the functions

$$e_n(t) = \frac{e^{2\pi int} - 1}{2\pi n}$$

belong to $\mathcal{H}$ and are orthonormal. If $f \in \mathcal{H}$ and $0 = \langle f, e_n\rangle = -i\int_0^1 f'(t)e^{2\pi int}dt$ for all $n \in \mathbb{N}$, then since the functions, $\{e^{2\pi int}\}_{n\neq 0}$ together with the constants span $L^2[0,1]$, we see that $f'(t)$ is constant and hence that $f(t)$ is a first degree polynomial. But the boundary conditions, $f(0) = f(1) = 0$ are easily seen to imply that this polynomial is 0.

Hence, we have shown that these functions are an orthonormal basis for $\mathcal{H}$. Applying the above theorem and our earlier calculation of the reproducing kernel, we have that,

$$\sum_{n\neq 0} \tfrac{(e^{2\pi inx}-1)(e^{2\pi iny}-1)}{4\pi^2 n^2}$$
$$= \sum_{n\neq 0} \frac{cos(2\pi n(x-y)) - cos(2\pi nx) - cos(2\pi ny) + 1}{2\pi^2 n^2}$$
$$= \begin{cases} (1-y)x, & x \leq y \\ y(1-x), & x > y \end{cases}.$$

Problem 3.5. *Let $\mathcal{H}$ be a RKHS on a disk of radius $R > 0$. Prove that $\mathcal{H} = H^2_\beta$ for some sequence β if and only if $z^n \in \mathcal{H}$ for all $n \geq 0$, the set $\{z^n : n \geq 0\}$ is total and $z^n \perp z^m, m \neq n$. Give a formula for the weight sequence β in this case.*

If a set of functions sums to the kernel as in the above theorem, it need not be an orthonormal basis for the RKHS, but such sets do have a very nice and useful characterization. The following definition is motivated by the first Parseval identity.

Definition 3.6. *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 that*

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

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 $\{u_s/\sqrt{2} : s \in S\} \cup \{v_t/\sqrt{2} : t \in T\}$ are both Parseval frames for $\mathcal{H}$.

Problem 3.7. *Show that the set of three vectors, $\{(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{2}}), (\frac{1}{\sqrt{3}}, 0), (\frac{1}{\sqrt{3}}, \frac{-1}{\sqrt{2}})\}$ is a Parseval frame for $\mathbb{C}^2$.*

Problem 3.8. *Show that for any $N \geq 3$, the set of vectors, $\{\sqrt{\frac{2}{N}}(cos(\frac{2\pi j}{N}), sin(\frac{2\pi j}{N})) : j = 1, ..., N\}$, is a Parseval frame for $\mathbb{C}^2$.*

Thus, in particular, we see that Parseval frames do not need to be linearly independent sets.

The following result shows one of the most common ways that Parseval frames arise.

Proposition 3.9. *Let $\mathcal{H}$ be a Hilbert space, let $\mathcal{H}_0 \subseteq \mathcal{H}$ be a closed subspace and let P_0 denote the orthogonal projection of $\mathcal{H}$ onto $\mathcal{H}_0$. If $\{e_s : s \in S\}$ is an orthonormal basis for $\mathcal{H}$, then $\{P_0(e_s) : s \in S\}$ is a Parseval frame for $\mathcal{H}_0$.*

Proof. Let $h \in \mathcal{H}_0$, then $h = P_0(h)$ and hence, $\langle h, e_s \rangle = \langle P_0(h), e_s \rangle = \langle h, P_0(e_s) \rangle$. Thus, $\|h\|^2 = \sum_{s \in S} |\langle h, P_0(e_s) \rangle|^2$ and the result follows. □

The following result shows that either of the Parseval identities could have been used to define Parseval frames.

Proposition 3.10. *Let $\mathcal{H}$ be a Hilbert space and let $\{f_s : s \in S\} \subseteq \mathcal{H}$. Then $\{f_s : s \in S\}$ is a Parseval frame if and only if $h = \sum_{s \in S} \langle h, f_s \rangle f_s$ for every $h \in \mathcal{H}$. Moreover, if $\{f_s : s \in S\}$ is a Parseval frame, then for any $h_1, h_2 \in \mathcal{H}$, we have that $\langle h_1, h_2 \rangle = \sum_{s \in S} \langle h_1, f_s \rangle \langle f_s, h_2 \rangle$.*

Proof. Let $\ell^2(S) = \{g : S \to \mathbb{C} : \sum_{s \in S} |g(s)|^2 < \infty\}$ denote the Hilbert space of square-summable functions and let $e_t : S \to \mathbb{C}, e_t(s) = \begin{cases} 1 & t = s \\ 0 & t \neq s \end{cases}$, be the canonical orthonormal basis. Define, $V : \mathcal{H} \to \ell^2(S)$, by $(Vh)(s) = \langle h, f_s \rangle$, so that in terms of the basis, $Vh = \sum_{s \in S} \langle h, f_s \rangle e_s$.

We have that $\{f_s : s \in S\}$ is a Parseval frame if and only if V is an isometry. Note that $\langle h, V^* e_t \rangle = \langle Vh, e_t \rangle = \langle h, f_t \rangle$, and hence, $V^* e_t = f_t$.

Recall from the general theory of Hilbert spaces that V is an isometry if and only if $V^*V = I_{\mathcal{H}}$. But $V^*V = I_{\mathcal{H}}$ if and only if,

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

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

Thus, we have that $\{f_s : s \in S\}$ is a Parseval frame if and only if V is an isometry if and only if $V^*V = I_{\mathcal{H}}$ if and only if

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

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

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

$$\begin{aligned}\langle h_1, h_2\rangle_{\mathcal{H}} = \langle V^*Vh_1, h_2\rangle_{\mathcal{H}} &= \langle Vh_1, Vh_2\rangle_{\ell^2(S)} \\ &= \sum_{s\in S}(Vh_1)(s)\overline{(Vh_2)(s)} = \sum_{s\in S}\langle h_1, f_s\rangle\langle f_s, h_2\rangle\end{aligned}$$

$\square$

The proof of the above result shows that our first proposition about how one could obtain a Parseval frame is really the most general example.

Proposition 3.11 (Larson). *Let $\{f_s : s \in S\}$ be a Parseval frame for a Hilbert space $\mathcal{H}$, then there is a Hilbert space $\mathcal{K}$ containing $\mathcal{H}$ as a subspace and an orthonormal basis $\{e_s : s \in S\}$ for $\mathcal{K}$, such that $f_s = P_{\mathcal{H}}(e_s), s \in S$, where $P_{\mathcal{H}}$ denotes the orthogonal projection of $\mathcal{K}$ onto $\mathcal{H}$.*

Proof. Let $\mathcal{K} = \ell^2(S)$ and let $V : \mathcal{H} \to \ell^2(S)$ be the isometry of the last proposition. Identifying $\mathcal{H}$ with $V(\mathcal{H})$ we may regard $\mathcal{H}$ as a subspace of $\ell^2(S)$. Note that $P = VV^* : \ell^2(S) \to \ell^2(S)$, satisfies $P = P^*$ and $P^2 = (VV^*)(VV^*) = V(V^*V)V^* = VV^* = P$. Thus, P is the orthogonal projection onto some subspace of $\ell^2(S)$. Since $Pe_s = V(V^*e_s) = Vf_s \in V(\mathcal{H})$, we see that P is the projection onto $V(\mathcal{H})$ and that when we identify "$h \equiv Vh$", we have that P is projection onto $\mathcal{H}$ with $Pe_s = Vf_s \equiv f_s$. $\square$

The following result gives us a more general way to compute reproducing kernels than 3.4. It was first pointed out to us by M. Papadakis.

Theorem 3.12 (Papadakis). *Let $\mathcal{H}$ be a RKHS on X with reproducing kernel $K(x, y)$. Then $\{f_s : s \in S\} \subseteq \mathcal{H}$ is a Parseval frame for $\mathcal{H}$ if and only if $K(x, y) = \sum_{s\in S} f_s(x)\overline{f_s(y)}$, where the series converges pointwise.*

Proof. Assuming that the set is a Parseval frame we have that, $K(x, y) = \langle k_y, k_x\rangle = \sum_{s\in S}\langle k_y, f_s\rangle\langle f_s, k_x\rangle = \sum_{s\in S} = \sum_{s\in S} \overline{f_s(y)}f_s(x)$.

Conversely, assume that the functions sum to 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

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

By Proposition 3.1, the set of such vectors, h, is dense in $\mathcal{H}$. Now it is easily seen that if we take a limit of a norm convergent sequence of vectors on both sides of this identity, then we obtain the identity for the limit vector, too. Thus, the condition to be a Parseval frame is met by the set $\{f_s : s \in S\}$. □

3.1. **Characterization of Reproducing Kernels.**

We now turn our attention to obtaining necessary and sufficient conditions for a function $K(x, y)$ to be the reproducing kernel for some RKHS. We first recall some facts about matrices.

Let $A = (a_{i,j})$ be a $n \times n$ complex matrix. Then A is **positive**(written: $A \geq 0$) if and only if for every $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ we have that $\sum_{i,j=1}^n \overline{\alpha_i} \alpha_j a_{i,j} \geq 0$.

Some remarks are in order. If we let $\langle , \rangle$ 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 the vector x whose i-th component is the number α_i.

Also, $A \geq 0$ if and only if $A = A^*$ and every eigenvalue, λ, of A, satisfies $\lambda \geq 0$. For this reason, some authors might prefer to call such matrices *positive semidefinite* or *non-negative*, but we stick to the notation most common in operator theory and C*-algebras. In the case that $A = A^*$ and every eigenvalue, λ, of A, satisfies $\lambda > 0$ then we will call A **strictly positive**(written: $A > 0$). Since A is a matrix, we see that $A > 0$ is equivalent to $A \geq 0$ and A invertible.

Definition 3.13. *Let X be a set and let $K : X \times X \to \mathbb{C}$ be a function of two variables. Then K is called a* **kernel function***(written: $K \geq 0$) provided that for every n and for every choice of n distinct points, $\{x_1, \ldots, x_n\} \subseteq X$, the matrix, $(K(x_i, x_j)) \geq 0$.*

This terminology is by no means standard. Some authors prefer to call kernel functions, *positive definite functions,* while other authors call them *positive semidefinite functions.* To further confuse matters, authors who call kernel functions, positive semidefinite functions often reserve the term positive definite function for functions such that $(K(x_i, x_j))$ is strictly positive. Thus, some care needs to be taken when using results from different places in the literature. We have adopted our terminology to try avoid these problems. Also because, as we shall shortly prove, a function is a kernel function if and only if there is a reproducing kernel Hilbert space for which it is the reporducing kernel.

Problem 3.14. *Prove that sums of kernel functions are kernel functions and that, if K is a kernel function and $f : X \to \mathbb{C}$ is any function, then $K_2(x, y) = f(x)K(x, y)\overline{f(y)}$ is a kernel function.*

Proposition 3.15. *Let X be a set and let $\mathcal{H}$ be a RKHS on X with reproducing kernel K. Then K is a kernel function.*

Proof. Fix $\{x_1, \ldots, x_n\} \subseteq X$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$. Then we have that $\sum_{i,j} \overline{\alpha_i}\alpha_j K(x_i, x_j) = \langle \sum_j \alpha_j k_{x_j}, \sum_i \alpha_i k_{x_i}\rangle = \| \sum_j \alpha_j k_{x_j}\|^2 \geq 0$, and the result follows. □

We remark that, generally, for a reproducing kernel, $(K(x_i, x_j)) > 0$. For if not, then the above calculation shows that there must exist some non-zero vector such that $\| \sum_j \alpha_j k_{x_j}\| = 0$. Hence, for every $f \in \mathcal{H}$ we have that $\sum_j \overline{\alpha_j} f(x_j) = \langle f, \sum_j \alpha_j k_{x_j}\rangle = 0$. Thus, in this case there is an equation of linear dependence between the values of every function in $\mathcal{H}$ at some 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 case $k_1(t) = k_0(t)$.

Alternatively, many spaces of analytic functions, such as the Hardy or Bergman spaces, contain all polynomials. Note that there is no equation of the form, $\sum_j \beta_j p(x_j) = 0$ that is satisfied by all polynomials. Consequently, the reproducing kernels for these spaces always define matrices that are strictly positive and invertible!

Thus, for example, recalling the Szego kernel for the Hardy space, we see that for any choice of points in the disk, the matrix, $(\frac{1}{1-\overline{\lambda_i}\lambda_j})$ is invertible. For one glimpse into how powerful the theory of RKHS can be, try to show this matrix is invertible by standard linear algebraic methods.

Although the above proposition is quite elementary, it has a converse that is quite deep and this gives us characterization of reproducing kernel functions.

Theorem 3.16 (Moore)**.** *Let X be a set and let $K : X \times X \to \mathbb{C}$ be a function. If K is a kernel function, then there exists a reproducing kernel Hilbert space of functions on X such that K is the reproducing kernel of $\mathcal{H}$.*

Proof. For each $y \in X$, set $k_y(x) = K(x, y)$ and let $W \subseteq \mathcal{F}(X)$ be the space spanned by the set, $\{k_y : y \in X\}$, of these functions.

We claim that there is a well-defined sesquilinear map, $B : W \times W \to \mathbb{C}$ 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 are scalars.

To see that B is well-defined on W, we must show that if $f(x) = \sum_j \alpha_j k_{y_j}(x)$ is identically zero as a function on X, then $B(f,w) = B(w,f) = 0$ for any $w \in W$. Since W is spanned by the functions, k_y it is enough to show that $B(f,k_y) = B(k_y,f) = 0$. But, by the definition, $B(f,k_y) = \sum_j \alpha_j K(y,y_j) = f(y) = 0$. Similarly, $B(k_y,f) = \sum_j \overline{\alpha_j} K(y_j,y) = \sum_j \overline{\alpha_j K(y,y_j)} = \overline{f(y)} = 0$.

Conversely, if $B(f,w) = 0$ for every $w \in W$, then 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 easily checked that it is sesquilinear. Moreover,for any $f \in W$ we have that $f(x) = B(f,k_x)$.

Next since K is positive, for any $f = \sum_j \alpha_j k_{y_j}$, we have that $B(f,f) = \sum_{i,j} \alpha_j \overline{\alpha_i} K(y_i,y_j) \geq 0$.

Thus, we have that B defines a semi-definite inner product on W. Hence, by the same proof as for the Cauchy-Schwarz inequality, one sees that $B(f,f) = 0$ if and only if $B(w,f) = B(f,w) = 0$ for all $w \in W$. Hence we see that $B(f,f) = 0$ if and only if f is the function that is identically 0.

Therefore, B is an inner product on W.

Now given any inner product on a vector space, we may complete the space, by taking equivalence classes of Cauchy sequences from W to obtain a Hilbert space, $\mathcal{H}$.

We must show that every element of $\mathcal{H}$ is actually a function on X(unlike, say, the case of completing the continuous functions on [0,1] to get $L^2[0,1]$). To this end, let $h \in \mathcal{H}$ and let $\{f_n\} \subseteq W$ be a Cauchy sequence that converges to h. By the Cauchy-Schwartz inequality, $|f_n(x) - f_m(x)| = |B(f_n - f_m, k_x)| \leq \|f_n - f_m\| \sqrt{K(x,x)}$. Hence, the sequence is pointwise Cauchy and we may define, $h(x) = \lim_n f_n(x)$. The usual argument shows that this value is independent of the particular Cauchy sequence chosen.

Finally, if we let $\langle \cdot, \cdot \rangle$ denote the inner product on $\mathcal{H}$, then for h as above, we have, $\langle h, k_y \rangle = \lim_n \langle f_n, k_y \rangle = \lim_n B(f_n, k_y) = \lim_n f_n(y) = h(y)$.

Thus, $\mathcal{H}$ is a RKHS on X and since k_y is the reproducing kernel for the point y, we have that $K(x,y) = k_y(x)$ is the reproducing kernel for $\mathcal{H}$. $\square$

Moore's theorem, together with Proposition 3.3 shows that there is a one-to-one correspondence between RKHS's on a set and positive functions on the set.

Definition 3.17. *Given a kernel function $K : X \times X \to \mathbb{C}$, we let $\mathcal{H}(K)$ denote the unique RKHS with reproducing kernel K.*

A reproducing kernel Hilbert space, $\mathcal{H}$, on X is said to **separate points,** provided that for every $x_1 \neq x_2$ there exists $f \in \mathcal{H}$ with $f(x_1) \neq f(x_2)$.

Problem 3.18. *Let $K : X \times X \to \mathbb{C}$ be a kernel function. Prove that for $x_1 \neq x_2$ the 2×2 matrix, $(K(x_j,x_i))$ is strictly positive if and only if k_{x_1} and k_{x_2} are linearly independent. Deduce that if for every $x_1 \neq x_2$, this 2×2 matrix is strictly positive, then $\mathcal{H}(K)$ separates points on X.*

One of the more difficult problems in the theory of RKHS's is starting with a positive definite function, K, to give a concrete description of the space, $\mathcal{H}(K)$. We shall refer to this as the **reconstruction problem**. For example, if we started with the Szego kernel on the disk, $K(z,w) = 1/(1-\overline{w}z)$, then the space W that we obtain in the proof of Moore's theorem consists of linear combinations of the functions, $k_w(z)$, which are rational functions with a single pole of order one outside the disk. Thus, the space W will contain no polynomials. Yet the space $\mathcal{H}(K) = H^2(\mathbb{D})$, which contains the polynomials as a dense subset. In later chapters we will prove theorems that, at least in the analytic case, will allow us to determine when $\mathcal{H}(K)$ contains polynomials.

We close this chapter with a simple application of Moore's Theorem.

Proposition 3.19. *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 positive, $\mathcal{H}(K)$ is the span of f and $\|f\| = 1$.*

Proof. To see that K is positive, we compute,
$\sum_{i,j} \alpha_i\overline{\alpha_j}K(x_i,x_j) = |\sum_i \alpha_i f(x_i)|^2 \geq 0.$

To find, $\mathcal{H}(K)$, note that every function $k_y = \overline{f(y)}f$ and hence, W is just the one-dimensional space spanned by f. Since finite dimensional spaces are automatically complete, $\mathcal{H}$ is just the span of f.

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

4. Interpolation and Approximation

One of the primary applications of the theory of reproducing kernel Hilbert spaces is to problems of interpolation and approximation. It turns out that it is quite easy to give concrete formulas for interpolation and approximation in these spaces.

Definition 4.1. *Let X and Y be sets and let $\{x_1,\ldots,x_n\} \subseteq X$ and $\{y_1,\ldots,y_n\} \subseteq Y$ be subsets, with the x_i's distinct points. We say that a function $g : X \to Y$* **interpolates** *these points, provided that $g(x_i) = y_i, i = 1,\ldots,n$.*

Let $\mathcal{H}$ be a RKHS on X with reproducing kernel, K. Assume that $\{x_1,\ldots,x_n\} \subseteq X$ is a set of distinct points and that $\{\lambda_1,\ldots,\lambda_n\} \subseteq \mathbb{C}$ is a collection of possibly non-distinct numbers. We will give necessary and sufficient conditions for there to exist a function $g \in \mathcal{H}$ that interpolates these values and then we will give a concrete formula for the unique such function of minimum norm. We will then use this same technique to give a solution to the reconstruction problem.

Before proceeding, we adopt the following notation. Given a finite set $F = \{x_1,\ldots,x_n\} \subseteq X$ of distinct points, we will let $\mathcal{H}_F \subseteq \mathcal{H}$ denote the subspace spanned by the functions, $\{k_{x_1},\ldots,k_{x_n}\}$.

Note that $dim(\mathcal{H}_F) \leq n$ and its dimension would be strictly less if and only if there is some non-zero equation of linear dependence among these functions. To understand what this means, suppose that $\sum_{j=1}^{n} \alpha_j k_{x_j} = 0$, then for every $f \in \mathcal{H}$,

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

Thus, 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 can also be seen to be equivalent to the fact that the linear map $T_F : \mathcal{H} \to \mathbb{C}^n$ defined by $T_F(f) = (f(x_1), \ldots, f(x_n))$ is not onto.

Thus, when $dim(\mathcal{H}_F) < n$, there will exist $(y_1, \ldots, y_n) \in \mathbb{C}^n$, which can not be interpolated by any $f \in \mathcal{H}$.

We've seen that it is possible for there to be equations of linear dependence between the kernel functions and sometimes even desirable, such as in the case of the Sobolev space, where the boundary conditions implied that $k_1 = k_0 = 0$.

We shall let P_F denote the orthogonal projection of $\mathcal{H}$ onto $\mathcal{H}_F$.

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

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

Proposition 4.2. *Let $\{x_1, \ldots, x_n\}$ be a set of distinct points in X and let $\{\lambda_1, \ldots, \lambda_n\} \subseteq \mathbb{C}$. If there exists $g \in \mathcal{H}$ that interpolates these values, then $P_F(g)$ is the unique function of minimum norm that interpolates these values.*

Proof. By the remarks, if g_1 and g_2 are any two functions that interpolate these points, then $(g_1 - g_2) \in \mathcal{H}_F^{\perp}$. Thus, all possible solutions of the interpolation problem are of the form $g + h, h \in \mathcal{H}_F^{\perp}$ and the unique vector of minimum norm from this set is $P_F(g)$. ☐

We now give necessary and sufficient conditions for the existence of any solution to this interpolation problem.

Some comments on matrices and vectors are in order. If $A = (a_{i,j})$ is an $n \times n$ matrix and we wish to write a matrix vector equation, $v = Aw$, then we need to regard v and w as column vectors. For typographical reasons it is easier to consider row vectors, so we will write a typical column vector as $v = (v_1, \ldots, v_n)^t$, where, "t", denotes the transpose.

We begin by recalling a calculation from the last section.

Proposition 4.3. *Let X be a set, let $\mathcal{H}$ be a RKHS on X with kernel K and let $\{x_1, ..., x_n\} \subseteq X$ be a finite set of distinct points. If $w = (\alpha_1, ..., \alpha_n)^t$ is a vector in the kernel of $(K(x_i, x_j))$, then the function, $f = \sum_j \alpha_j k_{x_j}$ is identically 0.*

Proof. We have that $f = 0$ if and only if $\|f\| = 0$. Now we compute, $\|f\|^2 = \sum_{i,j} \overline{\alpha_i}\alpha_j \langle k_{x_j}, k_{x_i}\rangle = \sum_{i,j} \overline{\alpha_i}\alpha_j K(x_i, x_j) = \langle (K(x_i, x_j))w, w\rangle_{\mathbb{C}^n} = 0$, and the result follows. □

Note that the above result shows that if $w_1 = (\alpha_1, \dots, \alpha_n)^t$ and $w_2 = (\beta_1, \dots, \beta_n)^t$ 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 $y \in X$.

Theorem 4.4. *(Interpolation in RKHS) Let $\mathcal{H}$ be a RKHS on X with reproducing kernel, K, let $F = \{x_1, \dots, x_n\} \subseteq X$ be distinct and let $\{\lambda_1, \dots, \lambda_n\} \subseteq \mathbb{C}$. Then there exists $g \in \mathcal{H}$ that interpolates these values if and only if $v = (\lambda_1, \dots, \lambda_n)^t$ is in the range of the matrix $(K(x_i, x_j))$. Moreover, in this case if we choose $w = (\alpha_1, ..., \alpha_n)^t$ to be any vector whose image is v, then $h = \sum_i \alpha_i k_{x_i}$ is the unique function of minimal norm in $\mathcal{H}$ that interpolates these points and $\|h\|^2 = \langle v, w\rangle$.*

Proof. First assume that there exists, $g \in \mathcal{H}$ such that $g(x_i) = \lambda_i, i = 1, ..., n$. 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 that $w_1 = (\beta_1, ..., \beta_n)^t$ is a solution of $v = (K(x_i, x_j))w$.

Conversely, if $w = (\alpha_1, ..., \alpha_n)^t$ is any solution of the matrix vector 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 the kernel of the matrix $(K(x_i, x_j))$ and hence by the above proposition, $P_F(g) = h$. Hence, h is the function of minimal norm that interpolates these points. Finally, $\|h\|^2 = \sum_{i,j} \overline{\alpha_i}\alpha_j K(x_i, x_j) = \langle (K(x_i, x_j))w, w\rangle = \langle v, w\rangle$. □

Corollary 4.5. *Let $\mathcal{H}$ be a RKHS on X with reproducing kernel, K, and let $F = \{x_1, \dots, x_n\} \subseteq X$ be distinct. If the matrix $(K(x_i, x_j))$ is invertible, then for any $\{\lambda_1, \dots, \lambda_n\} \subseteq \mathbb{C}$ there exists 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 $v = (\lambda_1, ..., \lambda_n)^t$.*

Assume that $F = \{x_1, ..., x_n\}$, that $P = (K(x_i, x_j))$ is invertible as in the above corollary and write $P^{-1} = (b_{i,j}) = B$. Let $e_j, j = 1, ..., n$ denote the canonical basis vectors for $\mathbb{C}^n$. The columns of B are the unique vectors $w_j, j = 1, ..., n$ that are solutions to $e_j = Pw_j$. Thus, if we set

$$g_F^j = \sum_i b_{i,j} k_{x_i},$$

then $g_F^j(x_i) = \delta_{i,j}$, where $\delta_{i,j}$ denotes the Dirac delta function. Hence,

$$g = \sum_j \lambda_j g_F^j$$

is the unique function in $\mathcal{H}_F$ satisfying $g(x_i) = \lambda_i, i = 1, ...n$.

Definition 4.6. *Let X be a set, $K : X \times X \to \mathbb{C}$ a positive function, and $F = \{x_1, ..., x_n\}$ a finite set of distinct points and assume that $(K(x_i, x_j))$ is invertible. We call the collection of functions, $g_F^j, j = 1, ...n$, a* **partition of unity** *for $\mathcal{H}_F$.*

Problem 4.7. *Let X be a set, $K : X \times X \to \mathbb{C}$ a positive function, and $F = \{x_1, ..., x_n\}$ a finite set of distinct points. Prove that $(K(x_i, x_j))$ is invertible if and only if $k_{x_1}, ..., k_{x_n}$ are linearly independent.*

Problem 4.8. *Let $K : X \times X \to \mathbb{C}$ be a kernel function and let $x_1, ..., x_n$ be a set of n distinct points in X. Prove that the dimension of the span of $\{k_{x_1}, ..., k_{x_n}\}$ is equal to the rank of the matrix $n \times n$ matrix, $(K(x_i, x_j))$.*

Problem 4.9. *Let X be a set, $K : X \times X \to \mathbb{C}$ a positive function, and $F = \{x_1, ..., x_n\}$ a finite set of distinct points and assume that $(K(x_i, x_j))$ is invertible. Assume that the constant function 1, belongs to span$\{k_{x_1}, \dots, k_{x_n}\}$. Prove that $\sum_j g_F^j(x) = 1$ for all $x \in X$.*

Problem 4.10. *Let K denote the Szego kernel and let $z_1 \neq z_2$ be points in the unit disk. Compute, explicitly the functions g_F^1, g_F^2 for $F = \{z_1, z_2\}$. What happens to these functions as $z_1 \to z_2$?*

Problem 4.11. *Repeat the above problem with the Szego kernel on the disk replaced by the Bergman kernel on the disk.*

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 inclusion. That is, setting $F_1 \leq F_2$ if and only if $F_1 \subseteq F_2$ defines a partial order on $\mathcal{F}_X$ and 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, $G = F_1 \cup F_2$. Also, recall that a *net* is a generalization of the concept of a sequence, but it is indexed by an arbitrary directed set. We will also use the concept of *convergence of nets,* which is defined by analogy with convergence of sequences. These concepts are used in a fairly self-explanatory manner in the following results. If the reader is still uncomfortable with these notions after reading the proofs of the following results, a good reference for further reading on nets is [3].

Proposition 4.12. *Let $\mathcal{H}$ be a 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 $k_y(\cdot) = K(\cdot, y)$. Given $\epsilon > 0$, by Proposition 3.1, there exists a finite collection of points, $F_0 = \{x_1, ..., x_n\}$ 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 $F_0 \subseteq F$. Since, g_F is the closest point in $\mathcal{H}_F$ to g and $g_{F_0} \in \mathcal{H}_F$, we have that $\|g - g_F\| < \epsilon$, for every $F_0 \subseteq F$, and the result follows. ☐

Before proving the next result we will need a result about finite matrices. Recall that if A and B are self-adjoint matrices, then we write, $A \leq B$ or $B \geq A$ to mean that $B - A \geq 0$.

Proposition 4.13. *Let $P \geq 0$ be an $n \times n$ matrix, and let $x = (x_1, ..., x_n)^t$ be a vector in $\mathbb{C}^n$. If $xx^* = (x_i\overline{x_j}) \leq cP$, for some scalar, $c > 0$, then x is in the range of P. Moreover, if y is any vector such that $x = Py$, then $0 \leq \langle x, y \rangle \leq c$.*

Proof. Let $\mathcal{R}(P)$ denote the range of P and $\mathcal{N}(P)$ denote the kernel of P. We have that $\mathcal{N}(P) = \mathcal{R}(P)^{\perp}$. 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, $\|w\|^4 = \langle w, x \rangle \langle x, w \rangle = \sum_{i,j} \overline{x_j} w_j x_i \overline{w_i} = \langle (x_i\overline{x_j})w, w \rangle \leq \langle cPw, w \rangle = 0$, 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 $\langle x, y \rangle^2 = \langle y, x \rangle \langle x, y \rangle = \langle (x_i\overline{x_j})y, y \rangle \leq \langle cPy, y \rangle = c\langle x, y \rangle$. Cancelling one factor of $\langle x, y \rangle$ from this inequality, yields the result. $\square$

Problem 4.14. *Let $P \geq 0$ be an $n \times n$ matrix. Prove that $x \in \mathcal{R}(P)$ if and only if there exists a constant $c > 0$ such that $(x_i\overline{x_j}) \leq cP$.*

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

Theorem 4.15. *Let $\mathcal{H}$ be a RKHS on X with reproducing kernel K and let $f : X \to \mathbb{C}$ be a function. Then the following are equivalent:*

(1) *$f \in \mathcal{H}$,*
(2) *there exists a constant, $c > 0$, such that for every finite subset, $F = \{x_1, ...x_n\} \subseteq X$, 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, $c > 0$, such that the function, $c^2K(x, y) - f(x)\overline{f(y)}$ 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, $\sum_{i,j} \overline{\alpha_i}\alpha_j f(x_i)\overline{f(x_j)} = |\sum_i \overline{\alpha_i} f(x_i)|^2 = |\langle f, g \rangle|^2 \leq \|f\|^2\|g\|^2 = \|f\|^2 \sum_{i,j} \overline{\alpha_i}\alpha_j K(x_i, x_j)$. 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 4.13 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 4.13 again.

(2) implies (1) By assumption, for every finite set F there exsits $h_F \in \mathcal{H}$ 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 $M - \frac{\epsilon^2}{8M} < \|g_{F_0}\|$. For any $F \in \mathcal{F}_X$ with $F_0 \subseteq F$ we have that $P_{F_0}(g_F) = g_{F_0}$ 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 $M - \frac{\epsilon^2}{8M} \leq \|g_{F_0}\| \leq \|g_F\| \leq M$.

Therefore, $0 \leq \|g_F\| - \|g_{F_0}\| \leq \frac{\epsilon^2}{8M}$, and we have that $\|g_F - g_{F_0}\|^2 = \|g_F\|^2 - \|g_{F_0}\|^2 = (\|g_F\| + \|g_{F_0}\|)(\|g_F\| - \|g_{F_0}\|) \leq 2M\frac{\epsilon^2}{8M}$. Thus, $\|g_F - g_{F_0}\| < \frac{\epsilon}{2}$ 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 $\|g_{F_1} - g_{F_2}\| < \epsilon$ 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 pointwise, 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 that $\|f\| \leq c$. Hence any c that meets the inequalities in (2) or (3) must be greater than $\|f\|$. □

The following result illustrates some of the surprising conequences of the above theorem.

Corollary 4.16. *Let $f : \mathbb{D} \to \mathbb{C}$ 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 $K(z, w) = \frac{c^2}{1 - z\overline{w}} - f(z)\overline{f(w)}$ is a kernel function on $\mathbb{D}$.*

What is a bit surprising in this last result is that the analyticity of f follows from the kernel function condition, which is just the requirement that certain matrices be positive semidefinite.

Problem 4.17. *Give the reproducing kernel condition for a function $f : [0, 1] \to \mathbb{R}$ to be absolutely continuous, with $f(0) = f(1)$ and f' square-integrable. Can you give a direct proof, without using the above theorem, from this condition that f is absolutely continuous?*

Problem 4.18. *Let $x \neq y$ be points in $\mathbb{D}$. Prove that* $\sup\{|f(y)| : f \in H^2(\mathbb{D}), f(x) = 0, \|f\| \leq 1\} \leq \frac{|x-y|}{|1-y\overline{x}|\sqrt{1-|y|^2}}$. *Is this inequality sharp?*

5. Operations on Kernels

In this section we consider how various algebraic operations on kernels effect the corresponding Hilbert spaces. The idea of examining and exploiting such relations, along with many of the results of this section can be traced back to the seminal work of Aronszajn. The first result characterizes when differences of reproducing kernels is positive.

Theorem 5.1 (Aronszajn)**.** *Let X be a set and let $K_i : X \times X \to \mathbb{C}, i = 1, 2$ be positive with corresponding reproducing kernel Hilbert spaces, $\mathcal{H}(K_i)$ and norms, $\|\cdot\|_i, i = 1, 2$. Then $\mathcal{H}(K_1) \subseteq \mathcal{H}(K_2)$ if and only if there exists a constant, $c > 0$ such that, $K_1(x, y) \le c^2 K_2(x, y)$. Moreover, in this case, $\|f\|_2 \le c\|f\|_1$ for all $f \in \mathcal{H}(K_1)$.*

Proof. First, assume that such a constant $c > 0$ exists. We have that if $f \in \mathcal{H}(K_1)$ with $\|f\|_1 = 1$, then $f(x)\overline{f(y)} \le K_1(x, y) \le c^2K_2(x, y)$, which implies that $f \in \mathcal{H}(K_2)$ and $\|f\|_2 \le c$. Hence, $\mathcal{H}(K_1) \subseteq \mathcal{H}(K_2)$ and $\|f\|_2 \le c\|f\|_1$.

We now prove the converse. Assume that $\mathcal{H}(K_1) \subseteq \mathcal{H}(K_2)$ and let $T : \mathcal{H}(K_1) \to \mathcal{H}(K_2)$ be the inclusion map, $T(f) = f$. If $\{f_n\}$ is a sequence in $\mathcal{H}(K_1)$ and $f \in \mathcal{H}(K_1), g \in \mathcal{H}(K_2)$ with $\|f_n - f\|_1 \to 0$ and $\|T(f_n) - g\|_2 \to 0$, then $f(x) = \lim_n f_n(x) = \lim_n T(f_n)(x) = g(x)$. Thus, $g = T(f)$ and by the closed graph theorem, T is closed and hence bounded. Let $\|T\| = c$ so that, $\|f\|_2 \le c\|f\|_1$ for all $f \in \mathcal{H}(K_1)$.

We claim that, $K_1 \le c^2K_2$. To this end fix, $\{x_1, ..., x_n\} \subseteq X$, and scalars, $\alpha_1, ..., \alpha_n \in \mathbb{C}$. We set $k^1_y(x) = K_1(x, y), k^2_y(x) = K_2(x, y)$.

We now calculate, $0 \le B = \sum_{i,j} \overline{\alpha_i}\alpha_j K_1(x_i, x_j) = \sum_{i,j} \overline{\alpha_i}\alpha_j k^1_{x_j}(x_i) = \sum_{i,j} \overline{\alpha_i}\alpha_j \langle k^1_{x_j}, k^2_{x_i}\rangle_2 = \langle \sum_j \alpha_j k^1_{x_j}, \sum_i \alpha_i k^2_{x_i}\rangle_2 \le \|\sum_j \alpha_j k^1_{x_j}\|_2 \cdot \|\sum_i \alpha_i k^2_{x_i}\|_2 \le c\|\sum_j \alpha_j k^1_{x_j}\|_1 \cdot \|\sum_i \alpha_i k^2_{x_i}\|_2$.

Squaring the first and last terms of this inequality, we have that, $B^2 \le c^2B(\sum_{i,j} \overline{\alpha_i}\alpha_j K_2(x_i, x_j))$. Upon cancelling a factor of B from each side, the result follows. ☐

Definition 5.2. *Given two Hilbert spaces, $\mathcal{H}_i$ with norms, $\|\cdot\|_i, i = 1, 2$, we say that $\mathcal{H}_1$ is* **contractively contained** *in $\mathcal{H}_2$ provided that $\mathcal{H}_1$ is a subspace of $\mathcal{H}_2$ and for every $h \in \mathcal{H}_1, \|h\|_2 \le \|h\|_1$.*

Corollary 5.3. *Let $\mathcal{H}_i, i = 1, 2$ be RKHS's on the set X with reproducing kernels, $K_i, i = 1, 2$, respectively. Then $\mathcal{H}_1$ is contractively contained in $\mathcal{H}_2$ if and only if $K_2 - K_1$ is a kernel function.*

Problem 5.4. *Use the above Corollary to show that $H^2(\mathbb{D})$ is contractively contained in $B^2(\mathbb{D})$.*

Definition 5.5. *A reproducing kernel Hilbert space on the unit disk thqt is contractively contained in the Hardy space $H^2(\mathbb{D})$ is called a* **de Branges space.**

Problem 5.6. *Show that a weighted Hardy space H^2_β is a de Branges space if and only if $\beta_n \le 1$, for all n.*

Properties of de Branges spaces played a key role in L. de Branges solution of the Bieberbach conjecture [4]. These spaces were studied extensively in the book [5].

If A and B are positive matrices, then so is $A+B$. Thus, if $K_i, i = 1, 2$ are kernel functions on a set X, then so is the function $K = K_1 + K_2$. The next result examines the relationship between the three corresponding RKHS's.

Theorem 5.7 (Aronszajn). *Let $\mathcal{H}_i, i = 1, 2$ be RKHS's on X with reproducing kernels, $K_i, i = 1, 2$, and norms, $\|\cdot\|_i, i = 1, 2$. If $K = K_1 + K_2$ and $\mathcal{H}(K)$ denotes the corresponding RKHS with norm, $\|\cdot\|$, then*

$$\mathcal{H}(K) = \{f_1 + f_2 : f_i \in \mathcal{H}_i, i = 1, 2\}$$

and for $f \in \mathcal{H}(K)$, we have

$$\|f\|^2 = min\{\|f_1\|_1^2 + \|f_2\|_2^2 : f = f_1 + f_2, f_i \in \mathcal{H}_i, i = 1, 2\}.$$

Proof. Consider the orthogonal direct sum of the two Hilbert spaces, $\mathcal{H}_1 \oplus \mathcal{H}_2 = \{(f_1, f_2) : f_i \in \mathcal{H}_i\}$ with inner product $\langle (f_1, f_2), (g_1, g_2)\rangle = \langle f_1, g_1\rangle_1 + \langle f_2, g_2\rangle_2$, where $\langle \cdot, \cdot\rangle_i$ denotes the inner product in the Hilbert space, $\mathcal{H}_i, i = 1, 2$. Note that with this inner product $\|(f_1, f_2)\|^2_{\mathcal{H}_1 \oplus \mathcal{H}_2} = \|f_1\|_1^2 + \|f_2\|_2^2$. Since $\mathcal{H}_i, i = 1, 2$ are both subspaces of the vector space of all functions on X, the intersection, $F_0 = \mathcal{H}_1 \cap \mathcal{H}_2$, is a well-defined vector space of functions on X. Let $\mathcal{N} = \{(f, -f) : f \in F_0\} \subseteq \mathcal{H}_1 \oplus \mathcal{H}_2$.

Note that $\mathcal{N}$ is a closed subspace, since if $\|(f_n, -f_n) - (f, g)\|_{\mathcal{H}_1 \oplus \mathcal{H}_2} \to 0$, then $\|f_n - f\|_1 \to 0$ and $\| - f_n - g\|_2 \to 0$, and hence, at each point, $f(x) = -g(x)$.

Therefore, decomposing $\mathcal{H}_1 \oplus \mathcal{H}_2 = \mathcal{N} + \mathcal{N}^\perp$, we see that every pair, $(f_1, f_2) = (f, -f) + (h_1, h_2)$ with $f \in F_0$ and $(h_1, h_2) \perp \mathcal{N}$.

Let $\mathcal{H}$ denote the vector space of functions of the form $\{f_1 + f_2 : f_i \in \mathcal{H}_i, i = 1, 2\}$ and define $\Gamma : \mathcal{H}_1 \oplus \mathcal{H}_2 \to \mathcal{H}$ by $\Gamma((f_1, f_2)) = f_1 + f_2$.

The map Γ is a linear surjection with kernel, $\mathcal{N}$ and hence, $\Gamma : \mathcal{N}^\perp \to \mathcal{H}$ is a vector space isomorphism. If we endow $\mathcal{H}$ with the norm that comes from this identification, then $\mathcal{H}$ will be a Hilbert space. If we let $P : \mathcal{H}_1 \oplus \mathcal{H}_2 \to \mathcal{N}^\perp$ denote the orthogonal projection, then for every $f = g_1 + g_2 \in \mathcal{H}$, we will have that

$$\begin{aligned}\|f\|^2 = \|P((g_1, g_2))\|^2_{\mathcal{H}_1 \oplus \mathcal{H}_2} &= min\{\|(g_1 + g, g_2 - g)\|^2_{\mathcal{H}_1 \oplus \mathcal{H}_2} : g \in F_0\}\\ &= min\{\|(f_1, f_2)\|^2_{\mathcal{H}_1 \oplus \mathcal{H}_2} : f = f_1 + f_2, f_i \in \mathcal{H}_i, i = 1, 2\}\\ &= min\{\|f_1\|_1^2 + \|f_2\|_2^2 : f = f_1 + f_2, f_i \in \mathcal{H}_i, i = 1, 2\}.\end{aligned}$$

For any two functions, $f = f_1 + f_2, g = g_1 + g_2$ in $\mathcal{H}$ we will have that $\langle f, g\rangle_{\mathcal{H}} = \langle P((f_1, f_2)), P((g_1, g_2))\rangle$.

It remains to see that $\mathcal{H}$ is a RKHS of functions on X with reproducing kernel K. Let $k_y^i(x) = K_i(x, y)$, so that $k_y^i \in \mathcal{H}_i$ is the kernel function. Note that if $(f, -f) \in \mathcal{N}$, then $\langle (f, -f), (k_y^1, k_y^2)\rangle = \langle f, k_y^1\rangle_1 + \langle -f, k_y^2\rangle_2 = f(y) - f(y) = 0$, so that $(k_y^1, k_y^2) \in \mathcal{N}^\perp$, for every $y \in X$. Thus, for any $f = f_1 + f_2 \in \mathcal{H}$, we have that $\langle f, k_y^1 + k_y^2\rangle_{\mathcal{H}} = \langle P((f_1, f_2)), P((k_y^1, k_y^2))\rangle$ $= \langle P((f_1, f_2)), (k_y^1, k_y^2)\rangle = \langle f_1, k_y^1\rangle_1 + \langle f_2, k_y^2\rangle_2 = f_1(y) + f_2(y) = f(y)$.

Thus, $\mathcal{H}$ is a RKHS with reproducing kernel, $K(x, y) = k_y^1(x) + k_y^2(x) = K_1(x, y) + K_2(x, y)$, and the proof is complete. $\square$

Corollary 5.8. *Let $\mathcal{H}_i, i = 1, 2$ be RKHS's on X with reproducing kernels, $K_i, i = 1, 2$, respectively. If $\mathcal{H}_1 \cap \mathcal{H}_2 = (0)$, then $\mathcal{H}(K_1 + K_2) = \{f_1 + f_2 : f_i \in \mathcal{H}_i\}$ with $\|f_1 + f_2\|^2 = \|f_1\|_1^2 + \|f_2\|_2^2$ is a reproducing kernel Hilbert space with kernel, $K(x, y) = K_1(x, y) + K_2(x, y)$ and $\mathcal{H}_i, i = 1, 2$ are orthogonal subspaces of $\mathcal{H}$.*

Problem 5.9. *Let $H_0^2(\mathbb{D}) = \{f \in H^2(\mathbb{D}) : f(0) = 0\}$. Show that $H_0^2(\mathbb{D})^\perp$ is the set of constant functions. Use this fact to compute the reproducing kernel for $H_0^2(\mathbb{D})$.*

Problem 5.10. *Let $\mathcal{H}$ be a RKHS on X with reproducing kernel, K, fix $x_0 \in X$ and let $\mathcal{H}_0 = \{f \in \mathcal{H} : f(x_0) = 0\}$. Compute the kernel function for $\mathcal{H}_0$.*

Problem 5.11. *Let $\alpha \in \mathbb{D}$ and let $\varphi_\alpha(z) = \frac{z-\alpha}{1-\bar{\alpha}z}$ denote the elementary Mobius transform. Prove that the reproducing kernel for $\{f \in H^2(\mathbb{D}) : f(\alpha) = 0\}$ is $K(z, w) = \frac{\varphi_\alpha(z)\overline{\varphi_\alpha(w)}}{1-\overline{w}z}$.*

Finite Dimensional RKHS's

We illustrate some applications of Aronzajn's theorem by examining finite dimensional RKHS's.

Let $\mathcal{H}$ be a finite dimensional RKHS on X with reproducing kernel, K. If we choose an orthonormal basis for $\mathcal{H}, f_1, ..., f_n$, then by Theorem **??**, $K(x, y) = \sum_{i=1}^n f_i(x)\overline{f_i(y)}$ and necessarily these functions will be linearly independent.

Conversely, let $f_i : X \to \mathbb{C}, i = 1, ..., n$ be linearly independent functions and set $K(x, y) = \sum_i f_i(x)\overline{f_i(y)}$. We shall use the above theorem to describe the space, $\mathcal{H}(K)$. If we let $K_i(x, y) = f_i(x)\overline{f_i(y)}$, and set $L_i(x, y) = \sum_{j\neq i} f_j(x)\overline{f_j(y)}$, then by Proposition 3.19, $\mathcal{H}(K_i) = span\{f_i\}$ and $\|f_i\|_i = 1$, where the norm is taken in $\mathcal{H}(K_i)$. Now, since $K(x, y) = K_i(x, y) + L_i(x, Y)$ and these functions are linearly independent, $\mathcal{H}(K_i) \cap \mathcal{H}(L_i) = (0)$, and by Corollary 5.8, these will be orthogonal subspaces of $\mathcal{H}(K)$. Thus, these functions are an orthonormal basis for $\mathcal{H}(K)$.

By contrast, consider the kernel, $K(x, y) = f_1(x)\overline{f_1(y)} + f_2(x)\overline{f_2(y)} + (f_1(x) + f_2(x))\overline{(f_1(y) + f_2(y))}$, where f_1 and f_2 are linearly independent. By Papadakis' theorem, these functions will be a Parseval frame for $\mathcal{H}(K)$ and, hence, $\mathcal{H}(K)$ will be the 2-dimensional space spanned by, f_1 and f_2. But since the three functions are not linearly independent, they can not be an orthonormal basis. We now use Aronszajn's theorem to figure out the precise relationship between these functions.

Set $L_1(x, y) = f_1(x)\overline{f_1(y)} + f_2(x)\overline{f_2(y)}$ and let $L_2(x, y) = (f_1(x) + f_2(x))\overline{(f_1(y) + f_2(y))}$. By the above reasoning, f_1, f_2 will be an orthonormal basis for $\mathcal{H}(L_1)$ and $\mathcal{H}(L_2)$ will be the span of the unit vector $f_1 + f_2$. Thus, by Aronszajn's theorem, $\|f_1\|_{\mathcal{H}(K)}^2 =$
$min\{\|rf_1 + sf_2\|_{\mathcal{H}(L_1)}^2 + \|t(f_1 + f_2)\|_{\mathcal{H}(L_2)}^2 : f_1 = rf_1 + sf_2 + t(f_1 + f_2)\} =$

$min\{|r|^2+|s|^2+|t|^2 : 1 = r+t, 0 = s+t\} = 2/3$. Similarly, $\|f_2\|^2_{\mathcal{H}(K)} = 2/3$ and from this it follows that $\|f_1+f_2\|^2_{\mathcal{H}(K)} = 2/3$.

Problem 5.12. *Determine, $\langle f_1, f_2\rangle_{\mathcal{H}(K)}$ for the last example. Find an orthonormal basis for this space.*

Pull-Backs, Restrictions and Composition Operators

Let X be a set, let $S \subseteq X$ be a subset and let $K : X \times X \to \mathbb{C}$ be positive definite. Then the restriction of K to $S \times S$ is also positive definite. Thus, we can use K to form a RKHS of functions on X or on the subset, S and it is natural to ask about the relationship between these two RKHS's.

More generally, if S is any set and $\varphi : S \to X$ is a function, then we let, $K\circ\varphi : S\times S \to \mathbb{C}$, denote the function given by, $K\circ\varphi(s,t) = K(\varphi(s),\varphi(t))$. When φ is one-to-one it is easily seen that $K\circ\varphi$ is positive definite on S. We will show below that this is true for a genreal function φ. Thus, it is natural to ask about the relationship between the RKHS $\mathcal{H}(K)$ of functions on X and the RKHS $\mathcal{H}(K\circ\varphi)$ of functions on S. When S is a subset of X, then the case discussed in the first paragraph is the special case of this latter construction where we take φ to be the inclusion of S into X.

Proposition 5.13. *Let $\varphi : S \to X$ and let K be a kernel function on X. Then $K\circ\varphi$ is a kernel function on S.*

Proof. Fix $s_1, ..., s_n \in S$, scalars, $\alpha_1, ..., \alpha_n$ and let $\{x_1, ..., x_p\} = \{\varphi(s_1), ..., \varphi(s_n)\}$, so that $p \le n$. Set $A_k = \{i : \varphi(s_i) = x_k\}$ and let $\beta_k = \sum_{i\in A_k}\alpha_i$. Then

$$\sum_{i,j}\bar{\alpha}_i\alpha_j K(\varphi(s_i),\varphi(s_j)) = \sum_{k,l}\sum_{i\in A_k}\sum_{j\in A_l}\bar{\alpha}_i\alpha_j K(x_k,x_l) = \sum_{k,l}\bar{\beta}_k\beta_l K(x_k,x_l) \ge 0.$$

Hence, $K\circ\varphi$ is a kernel function on S. □

Theorem 5.14. *Let X and S be sets, let $K : X\times X \to \mathbb{C}$ be positive definite and let $\varphi : S \to X$ be a function. Then $\mathcal{H}(K\circ\varphi) = \{f\circ\varphi : f \in \mathcal{H}(K)\}$, and for $u \in \mathcal{H}(K\circ\varphi)$ we have that $\|u\|_{\mathcal{H}(K\circ\varphi)} = inf\{\|f\|_{\mathcal{H}(K)} : u = f\circ\varphi\}$.*

Proof. Let $f \in \mathcal{H}(K)$, with $\|f\|_{\mathcal{H}(K)} = c$, then $f(x)\overline{f(y)} \le c^2K(x,y)$ in the positive definite order. Since this is an inequality of matrices over all finite sets, we see that $f\circ\varphi(s)\overline{f\circ\varphi(t)} \le c^2K(\varphi(s),\varphi(t))$. Hence, by **??**, $f\circ\varphi \in \mathcal{H}(K\circ\varphi)$ with $\|f\circ\varphi\|_{\mathcal{H}(K\circ\varphi)} \le c$.

This calculation shows that there exists a contractive, linear map, $C_\varphi : \mathcal{H}(K) \to \mathcal{H}(K\circ\varphi)$ given by $C_\varphi(f) = f\circ\varphi$.

Set $h_t(\cdot) = K(\varphi(\cdot),\varphi(t))$, so that these are the kernel functions for $\mathcal{H}(K\circ\varphi)$. Note that for any finite set of points and scalars, if $u = \sum_i\alpha_ih_{t_i}$, then $\|u\|_{\mathcal{H}(K\circ\varphi)} = \|\sum_i\alpha_ik_{\varphi(t_i)}\|_{\mathcal{H}(K)}$. It follows that there is a well-defined isometry, $\Gamma : \mathcal{H}(K\circ\varphi) \to \mathcal{H}(K)$, satisfying, $\Gamma(h_t) = k_{\varphi(t)}$.

We have that $C_\varphi\circ\Gamma$ is the identity on $\mathcal{H}(K\circ\varphi)$ and the result follows. □

Definition 5.15. *Given sets, X and S, a function $\varphi : S \to X$, and a positive definite function, $K : X \times X \to \mathbb{C}$, we call $\mathcal{H}(K \circ \varphi)$ the* **pull-back of $\mathcal{H}(K)$ along** *φ and we call the linear map, $C_\varphi : \mathcal{H}(K) \to \mathcal{H}(K \circ \varphi)$ the* **pull-back map**.

Note that $\langle \Gamma(h_t), k_y \rangle_{\mathcal{H}(K)} = k_{\varphi(t)}(y) = K(y, \varphi(t)) = \overline{K(\varphi(t), y)} = \overline{k_y(\varphi(t))} = \overline{C_\varphi(k_y)(t)} = \langle h_t, C_\varphi(k_y) \rangle_{\mathcal{H}(K\circ\varphi)}$. Since the linear spans of such functions are dense in both Hilbert spaces, this calculation shows that $C_\varphi = \Gamma^*$. Since Γ is an isometry its range is a closed subspace and it follows that $\Gamma^* = C_\varphi$ is an isometry on the range of Γ and is 0 on the orthocomplement of the range. Such a map is called a **coisometry.**

Thus, in the case that $S \subseteq X$ and φ is just the inclusion map, so that $K \circ \varphi = K\big|_Y$, we see that C_φ is the coisometry that identifies the closure of the subspace of $\mathcal{H}(K)$ spanned by $\{k_y : y \in Y\}$ which are functions on X, with the same set of functions regarded as functions on Y.

The theory of *composition operators* is a special case of the pull-back construction.

Given sets $X_i, i = 1, 2$ and kernel functions $K_i : X_i \times X_i \to \mathbb{C}, i = 1, 2$, we wish to identify those functions, $\varphi : X_1 \to X_2$ such that there is a well-defined, bounded map, $C_\varphi : \mathcal{H}(K_2) \to \mathcal{H}(K_1)$ given by $C_\varphi(f) = f \circ \varphi$.

Theorem 5.16. *Let $X_i, i = 1, 2$ be sets, $\varphi : X_1 \to X_2$ a function and $K_i : X_i \times X_i \to \mathbb{C}, i = 1, 2$ kernel functions. Then the following are equivalent:*

(1) $\{f \circ \varphi : f \in \mathcal{H}(K_2)\} \subseteq \mathcal{H}(K_1)$,
(2) $C_\varphi : \mathcal{H}(K_2) \to \mathcal{H}(K_1)$ *is a bounded, linear operator,*
(3) *there exists a constant, $c > 0$, such that $K_2 \circ \varphi \leq c^2 K_1$.*

Moreover, in this case, $\|C_\varphi\|$ is the least such constant c.

Proof. Clearly, (2) implies (1). To see that (3) implies (2), let $f \in \mathcal{H}(K_2)$, with $\|f\| = M$. Then $f(x)\overline{f(y)} \leq M^2 K_2(x, y)$, which implies that,

$$f(\varphi(x))\overline{f(\varphi(y))} \leq M^2 K_2(\varphi(x), \varphi(y)) \leq M^2 c^2 K_1(x, y).$$

Thus, it follows that $C_{\varphi(f)} = f \circ \varphi \in \mathcal{H}(K_1)$ with $\|C_\varphi(f)\|_1 \leq c\|f\|_2$. Hence, C_φ is bounded and $\|C_\varphi\| \leq c$.

Finally, by the previous theorem, (1), is equivalent to the statement that $\mathcal{H}(K_2 \circ \varphi) \subseteq \mathcal{H}(K_1)$, which is equivalent to the kernel inequality, (3), by Aronszajn's Theorem 5.1.

The statement in Aronszajn's Theorem regarding the norms of inclusion maps proves the last statement. $\square$

Problem 5.17. *Let $X = \mathbb{D}$ and let K be the Szego kernel. Describe the spaces, $\mathcal{H}(K \circ \varphi)$ for $\varphi(z) = z^2$ and for $\varphi(z) = \frac{z-\alpha}{1-\bar{\alpha}z}, \alpha \in \mathbb{D}$ a simple Mobius map.*

Problem 5.18. *Let $R = \{x + iy : x > 0\}$ denote the right-half plane, and let $\varphi : R \to \mathbb{D}$ be defined by $\varphi(z) = \frac{z-1}{z+1}$. Compute the pull-back of the Szego* *kernel. Show that* $h \in \varphi^*(H^2(\mathbb{D}))$ *if and only if the function* $f(z) = h(\frac{1+z}{1-z})$ *is in* $H^2(\mathbb{D})$.

An Application to the Theory of Group Representations

The theory of composition operators is an important tool in the study of unitary representations of groups and gives a very quick proof of a result known as Naimark's dilation theorem. Given a group G with identity e, a Hilbert space $\mathcal{H}$. We call a homomorphism, $\pi : G \to B(\mathcal{H})$, such that $\pi(e) = I_{\mathcal{H}}$ and $\pi(g^{-1}) = \pi(g)^*$, i.e., such that $\pi(g)$ is unitary, for all $g \in G$, a **unitary representation of** G **on** $\mathcal{H}$. A unitary representation is said to be **cyclic**, if there exists, $v_0 \in \mathcal{H}$, such that the linear span of $\pi(G)v_0 = \{\pi(g)v_0 : g \in G\}$ is dense in $\mathcal{H}$.

Definition 5.19. *Let* G *be a group and let* $p : G \to \mathbb{C}$ *be a function. Then* p *is called a* **positive definite function on G** *provided that for every* n *and every* $g_1, ..., g_n \in G$, *the matrix* $(p(g_i^{-1}g_j))$ *is positive semidefinite.*

Note that saying that p is positive definite is the same as requiring that $K_p : G \times G \to \mathbb{C}$ defined by $K_p(g, h) = p(g^{-1}h)$ is a kernel function. Thus, to every positive definite function on G there is associated a RKHS, $\mathcal{H}(K_p)$. Note that in this space, the kernel function for evaluation at the identity element is $k_e(g) = p(g^{-1})$.

Now fix, $g \in G$ and consider the function, $\varphi : G \to G$ defined by $\varphi(h) = g^{-1}h$. We have that $K \circ \varphi(g_1, g_2) = K(g^{-1}g_1, g^{-1}g_2) = p((g^{-1}g_1)^{-1}(g^{-1}g_2)) = K(g_1, g_2)$. Thus, by the above Theorem there is a well-defined contractive linear map, $U_g : \mathcal{H}(K_p) \to \mathcal{H}(K_p)$ with $(U_g f)(h) = f(g^{-1}h)$ for any $f \in \mathcal{H}(K_p)$. Now $(U_{g_1} \circ U_{g_2})(f)(h) = (U_{g_2} f)(g_1^{-1}h) = f(g_2^{-1}g_1^{-1}h) = (U_{g_1 g_2} f)(h)$, and so the map $\pi : G \to B(\mathcal{H}(K_p))$, is a homomorphism. Since $U_{g^{-1}} \circ U_g = I_{\mathcal{H}(K_p)}$, and both of these maps are contractions, it follows that they must both be invertible isometries and hence, unitaries.

Thus, to every positive definite function on p on G, we have associated a unitary representation, $\pi : G \to B(\mathcal{H}(K_p))$, by setting $\pi(g) = U_g$. This construction gives an immediate proof of a theorem of Naimark.

Theorem 5.20 (Naimark's Dilation theorem)**.** *Let* G *be a group and let* $p : G \to \mathbb{C}$ *be a positive definite function. Then there exists a Hilbert space* $\mathcal{H}$, *a unitary representation* $\pi : G \to B(\mathcal{H})$, *and a vector* $v \in \mathcal{H}$, *such that* $p(g) = \langle \pi(g)v, v\rangle$. *Moreover, any function of this form is positive definite.*

Proof. Let $\mathcal{H} = \mathcal{H}(K_p)$, let $\pi(g) = U_g$ and let $v = k_e$. We have that $\langle \pi(g)k_e, k_e\rangle = (\pi(g)k_e)(e) = k_e(g^{-1}) = p(g)$.

Finally, if $f(g) = \langle \pi(g)v, v\rangle$, and we pick $\{g_1, ..., g_n\} \subseteq G$ and scalars $\alpha_1, ..., \alpha_n \in \mathbb{C}$, then we have that

$$\sum_{i,j=1}^{n} \overline{\alpha_i}\alpha_j f(g_i^{-1}g_j) = \langle w, w\rangle$$

where $w = \sum_{j=1}^{n} \alpha_j \pi(g_j)v$ and so p is a positive definite function. $\square$

Given a positive definite function p on a group, the representation that we get by considering composition operators on $\mathcal{H}(K_p)$ is also cyclic, with cycic vector k_e. To see this note that $(U_g k_e)(h) = k_e(g^{-1}h) = K(g^{-1}h, e) = p(h^{-1}g) = K(h, g) = k_g(h)$. Thus, $U_g k_e = k_g$, and hence, the span of $\pi(G)k_e$ is equal to the span of $\{k_g : g \in G\}$ which is always dense in the RKHS.

Conversely, assume that we have a unitary representation γ of G on some Hilbert space $\mathcal{H}$, which has a cyclic vector v_0 and we define $p(g) = \langle \gamma(g)v_0, v_0\rangle$. By Naimark's theorem, we know that p is positive definite.

Now consider the Hilbert space, $\mathcal{H}(K_p)$ and unitary representation π of G. We claim that there is a Hilbert space isomorphism, $W : \mathcal{H}(K_p) \to \mathcal{H}$, such that $Wk_e = v_0$ and $W\pi(g) = \gamma(g)W$, for all $g \in G$.

To define W we set, $Wk_g = \gamma(g)v_0$, and extend linearly. Note that

$$\|\sum_i \alpha_i k_{g_i}\|^2 = \sum_{i,j} \alpha_i \overline{\alpha_j} k_{g_i}(g_j) = \sum_{i,j} \alpha_i \overline{\alpha_j} p(g_j^{-1} g_i)$$
$$= \sum_{i,j} \alpha_i \overline{\alpha_j} \langle \gamma(g_j^{-1} g_i) v_0, v_0\rangle = \|\sum_i \alpha_i \gamma(g_i) v_0\|^2.$$

This equality shows that W is well-defined and an isometry. Thus, W can be extended by continuity to an isometry from all of $\mathcal{H}(K_p)$ onto $\mathcal{H}$. Finally,

$$W\pi(g)k_{g_1} = W\pi(g)\pi(g_1)k_e = W\pi(gg_1)k_e =$$
$$Wk_{gg_1} = \gamma(gg_1)v_0 = \gamma(g)\gamma(g_1)v_0 = \gamma(g)Wk_{g_1}$$

and since these vectors span the space, $W\pi(g) = \gamma(g)W$.

Thus, the representation γ is unitarily equivalent to the representation π, via a map that carries the cyclic vector v_0 to the vector k_e.

These calculations show that if one requires the vector v appearing in the dilation of a positive definite function in Naimark's dilation theorem, then up to a unitary equivalence, we are in the situation where $\mathcal{H} = \mathcal{H}(K_p)$, $\pi(g) = U_g$ and $v = k_e$.

Problem 5.21. *Prove that the function $p(x) = cos(x)$ is a positive definite function on the group $(\mathbb{R}, +)$. Show that $\mathcal{H}(K_p)$, is two-dimensional and explicitly describe the unitary representation of $\mathbb{R}$ on this space.*

Problem 5.22. *Let $p(z) = \sum_{k=0}^{n} a_k z^k$, with $a_k \geq 0, k = 0, ..., n$. Show that p is a positive definite function on the circle group, $(\mathbb{T}, \cdot)$ and that $\mathcal{H}(K_p)$ is $n+1$-dimensional. Explicitly describe the unitary representation.*

Problem 5.23. *Let $p(n) = \begin{cases} 1 & n \ even \\ 0 & n \ odd \end{cases}$. Prove that p is a positive definite function on $(\mathbb{Z}, +)$, find $\mathcal{H}(K_p)$ and the corresponding unitary representation.*

Products of Kernels and Tensor Products of Spaces

Recall that if $\mathcal{H}_i, i = 1, 2$ are Hilbert spaces, then we can form their tensor product, $\mathcal{H}_1 \otimes \mathcal{H}_2$, which is a new Hilbert space. If $\langle \cdot, \cdot \rangle_i, i = 1, 2$, denotes the respective inner products on the spaces, then to form this new space, we first endow the algebraic tensor product with the inner product obtained by setting, $\langle f \otimes g, h \otimes k \rangle = \langle f, h \rangle_1 \langle g, k \rangle_2$ and extending linearly and then completing the algebraic tensor product in the induced norm. One of the key facts about this completed tensor product is that it contains the algebraic tensor product faithfully, that is, the inner product satisfies, $\langle u, u \rangle > 0$, for any $u \neq 0$ in the algebraic tensor product.

Now if $\mathcal{H}$ and $\mathcal{F}$ are RKHS's on sets X and S, respectively, then it is natural to want to identify an element of the algebraic tensor product, $u = \sum_{i=0}^n h_i \otimes f_i$ with the function, $\hat{u}(x, s) = \sum_{i=0}^n h_i(x) f_i(s)$. The following theorem shows that not only is this identification well-defined, but that it also extends to the completed tensor product.

Theorem 5.24. *Let $\mathcal{H}$ and $\mathcal{F}$ be RKHS's on sets X and S, with reproducing kernels, $K_1(x, y)$ and $K_2(s, t)$. Then $K((x, s), (y, t)) = K_1(x, y) K_2(s, t)$ is a kernel function on $X \times S$ and the map $u \to \hat{u}$ extends to a well-defined, linear isometry from $\mathcal{H} \otimes \mathcal{F}$ onto the reproducing kernel Hilbert space $\mathcal{H}(K)$.*

Proof. Set $k_y^1(x) = K_1(x, y)$ and $k_t^2(s) = K_2(s, t)$. Note that if $u = \sum_{i=1}^n h_i \otimes f_i$, then $\langle u, k_y^1 \otimes k_t^2 \rangle_{\mathcal{H} \otimes \mathcal{F}} = \sum_{i=1}^n \langle h_i, k_y^1 \rangle_{\mathcal{H}} \langle f_i, k_t^2 \rangle_{\mathcal{F}} = \hat{u}(y, t)$.

Thus, we may extend the mapping $u \to \hat{u}$ from the algebraic tensor product to the completed tensor product as follows. Given $u \in \mathcal{H} \otimes \mathcal{F}$, define a function on $X \times S$ by setting $\hat{u}(y, t) = \langle u, k_y^1 \otimes k_t^2 \rangle_{\mathcal{H} \otimes \mathcal{F}}$.

It is readily seen that the set $\mathcal{L} = \{\hat{u} : u \in \mathcal{H} \otimes \mathcal{F}\}$ is a vector space of functions on $X \times S$. Moreover, the map $u \to \hat{u}$ will be one-to-one unless there exists a non-zero $u \in \mathcal{H} \otimes \mathcal{F}$ such that $\hat{u}(y, t) = 0$ for all $(y, t) \in X \times S$. But this latter condition would imply that u is orthogonal to the span of $\{k_y^1 \otimes k_t^2 : (y, t) \in X \times S\}$. But since the span of $\{k_y^1 : y \in X\}$ is dense in $\mathcal{H}$ and the span of $\{k_t^2 : t \in S\}$ is dense in $\mathcal{F}$, it readily follows that the span of $\{k_y^1 \otimes k_t^2 : (y, t) \in X \times S\}$ is dense in $\mathcal{H} \otimes \mathcal{F}$. Hence, if $\hat{u} = 0$, then u is orthogonal to a dense subset and so $u = 0$.

Thus, we have that the map $u \to \hat{u}$ is one-to-one from $\mathcal{H} \otimes \mathcal{F}$ onto $\mathcal{L}$ and we may use this identification to give $\mathcal{L}$ the structure of a Hilbert space. That is, for $u, v \in \mathcal{H} \otimes \mathcal{F}$, we set $\langle \hat{u}, \hat{v} \rangle_{\mathcal{L}} = \langle u, v \rangle_{\mathcal{H} \otimes \mathcal{F}}$.

Finally, since for any $(y, t) \in X \times S$, we have that $\hat{u}(y, t) = \langle \hat{u}, \widehat{k_y^1 \otimes k_t^2} \rangle$, we see that $\mathcal{L}$ is a reproducing kernel Hilbert space with kernel

$$
\begin{aligned}
K((x, s), (y, t)) &= \langle \widehat{k_y^1 \otimes k_t^2}, \widehat{k_x^1 \otimes k_s^2} \rangle_{\mathcal{L}} \\
&= \langle k_y^1 \otimes k_t^2, k_x^1 \otimes k_s^2 \rangle_{\mathcal{H} \otimes \mathcal{F}} = \langle k_y^1, k_x^1 \rangle_{\mathcal{H}} \langle k_t^2, k_s^2 \rangle_{\mathcal{F}} = K_1(x, y) K_2(s, t)
\end{aligned}
$$

and so K is a kernel function.

By the uniqueness of kernels, we have that $\mathcal{L} = \mathcal{H}(K)$ as sets of functions and as Hilbert spaces. Thus, the map $u \to \hat{u}$ is an isometric linear map from $\mathcal{H} \otimes \mathcal{F}$ onto $\mathcal{H}(K)$. $\square$

Corollary 5.25. *If X and S are sets and $K_1 : X \times X \to \mathbb{C}$ and $K_2 : S \times S \to \mathbb{C}$ are kernel functions, then $K : (X \times S) \times (X \times S) \to \mathbb{C}$ given by $K((x, s), (y, t)) = K_1(x, y)K_2(s, t)$ is positive definite.*

Definition 5.26. *We call the 4 variable function $K((x, s), (y, t)) = K_1(x, y)K_2(s, t)$ the* **tensor product** *of the kernels K_1 and K_2.*

A slightly more subtle corollary is given by the following.

Corollary 5.27. *Let X be a set and let $K_i : X \times X \to \mathbb{C}, i = 1, 2$ be kernel functions, then their product, $P : X \times X \to \mathbb{C}$ given by $P(x, y) = K_1(x, y)K_2(x, y)$ is a kernel function.*

Proof. Given any points, $\{x_1, ..., x_n\}$, set $w_i = (x_i, x_i)$, and then we have that the $n \times n$ matrix, $(P(x_i, x_j)) = (K(w_i, w_j)) \geq 0$. □

Definition 5.28. *We call the 2 variable kernel, $P(x, y) = K_1(x, y)K_2(x, y)$ the* **product** *of the kernels.*

Given $K_i : X \times X \to \mathbb{C}, i = 1, 2$ we have two kernels and two RKHS's. The first is the tensor product, $K : (X \times X) \times (X \times X) \to \mathbb{C}$ which gives a RKHS of functions on $X \times X$. The second is the product, $P : X \times X \to \mathbb{C}$ which gives a RKHS of functions on X. The relationship between these two spaces can be seen by using the results of the last subsection.

Let $\Delta : X \to X \times X$ denote the **diagonal map**, defined by $\Delta(x) = (x, x)$. Then $P(x, y) = K(\Delta(x), \Delta(y))$, that is, $P = K \circ \Delta$. Thus, $\mathcal{H}(P)$ is the pull-back of $\mathcal{H}(K) = \mathcal{H}(K_1) \otimes \mathcal{H}(K_2)$ along the diagonal map.

The last corollary can be used to prove a familiar fact from matrix theory.

Definition 5.29. *Let $A = (a_{i,j})$ and $B = (b_{i,j})$ be two $n \times n$ matrices. Then their* **Schur product** *is the $n \times n$ matrix,*

$$A * B = (a_{i,j}b_{i,j}).$$

Corollary 5.30 (Schur)**.** *Let $P = (p_{i,j})$ and $Q = (q_{i,j})$ be $n \times n$ matrices. If $P \geq 0$ and $Q \geq 0$, then $P * Q \geq 0$.*

Proof. Consider the n point set, $X = \{1, ..., n\}$. If we regard the matrices P and Q as functions on $X \times X$, then they are kernel functions. Apply the last corollary. □

Conversely, if one knows the Schur product result, then one can deduce directly that products of kernel functions are kernel functions. The following problem shows how to give a matrix theoretic proof of Schur's result.

Problem 5.31.
- (i) *Prove that an $n \times n$ matrix is rank 1 if and only if it is of the form, $(\alpha_i\overline{\beta_j})$ for some choice of scalars.*
- (ii) *Prove that an $n \times n$ rank 1 matrix is positive if and only if it is of the form $(\alpha_i\overline{\alpha_j})$, for some choice of scalars.*
- (iii) *Prove that every positive $n \times n$ rank k matrix can be written as a sum of k positive rank 1 matrices.*

(iv) *Prove that the Schur product is distributive over sums, i.e.,* $(A+B)*(C+D) = A*C + B*C + A*D + B*D$.

(v) *Deduce Schur's result.*

Problem 5.32. *Prove that the Bergman space,* $B^2(\mathbb{D})$ *is the pull-back of* $H^2(\mathbb{D}) \otimes H^2(\mathbb{D})$ *along the diagonal.*

Push-Outs of RKHS's

Given a RKHS $\mathcal{H}(K)$ on X and a function $\psi : X \to S$ we would also like to induce a RKHS on S. To carry out this construction, we first consider the subspace, $\widetilde{\mathcal{H}} = \{f \in \mathcal{H}(K) : f(x_1) = f(x_2) \text{whenever} \psi(x_1) = \psi(x_2)\}$. If $\tilde{K}(x, y)$ denotes the reproducing kernel for this subspace and we set, $\tilde{k_y}(x) = \tilde{K}(x, y)$, then it readily follows that, whenever $\psi(x_1) = \psi(x_2)$ and $\psi(y_1) = \psi(y_2)$, we have that, $\tilde{k_y}(x_1) = \tilde{k_y}(x_2)$ and $\tilde{k_{y_1}} = \tilde{k_{y_2}}$. Thus, for any such pair of points, $\tilde{K}(x_1, y_1) = \tilde{K}(x_2, y_2)$. It follows that there is a well-defined positive definite function on $K_\psi : S \times S \to \mathbb{C}$ given by $K_\psi(s, t) = \tilde{K}(\psi^{-1}(s), \psi^{-1}(t))$.

We call the RKHS, $\mathcal{H}(K_\psi)$ on S, the **push-out of $\mathcal{H}(K)$ along ψ.**

As an example, we consider the Bergman space on the disk, $B^2(\mathbb{D})$. This space has a reproducing kernel given by the formula,

$$K(z, w) = \frac{1}{(1 - \overline{w}z)^2} = \sum_{n=0}^{\infty} (n + 1)(\overline{w}z)^n.$$

If we let $\psi : \mathbb{D} \to \mathbb{D}$, be defined by $\psi(z) = z^2$, then we can pull $B^2(\mathbb{D})$ back along ψ and we can push $B^2(\mathbb{D})$ forward along ψ. We compute the kernels in each of these cases.

The kernel for the pull-back is simply,

$$K \circ \psi(z, w) = \frac{1}{(1 - \overline{w}^2 z^2)^2} = \sum_{n=0}^{\infty} (n + 1)(\overline{w}z)^{2n}.$$

For the push-out, we first compute, $\widetilde{\mathcal{H}}$. Note that $\psi(z_1) = \psi(z_2)$ if and only if $z_2 = \pm z_1$. Thus, $\widetilde{\mathcal{H}} = \{f \in B^2(\mathbb{D}) : f(z) = f(-z)\}$, i.e., the subspace of even functions. This subspace is spanned by the even powers of z, and consequently, it has kernel,

$$\tilde{K}(z, w) = \sum_{n=0}^{\infty} (2n + 1)(\overline{w}z)^{2n}.$$

Since, $\psi^{-1}(z) = \pm\sqrt{z}$, we see that the push-out of $B^2(\mathbb{D})$, is the new RKHS of functions on the disk with reproducing kernel,

$$K_\psi(z, w) = \tilde{K}(\pm\sqrt{z}, \pm\sqrt{w}) = \sum_{n=0}^{\infty} (2n + 1)(\overline{w}z)^n.$$

Thus, the pull-back and the push-out of $B^2(\mathbb{D})$ are both spaces of analytic functions on $\mathbb{D}$ spanned by powers of z and these powers are orthogonal functions with different norms. The push-out is a weighted Hardy space, but the pull-back is not a weighted Hardy space. If we generalized the definition of a weighted Hardy space, by allowing some of the weights to be 0, then the pull-back is a generalized weighted Hardy space.

Problem 5.33. *Let $\psi : \mathbb{D} \to \mathbb{D}$, be defined by $\psi(z) = z^2$, as above, and compute the pull-back and push-out of $H^2(\mathbb{D})$ along ψ.*

6. Multipliers of Reproducing Kernel Hilbert Spaces

In this section we develop the theory of functions that multiply a reproducing kernel Hilbert space back into itself.

Definition 6.1. *Let $\mathcal{H}$ be a RKHS on X with kernel function, K. A function $f : X \to \mathbb{C}$ is called a* **multiplier of** $\mathcal{H}$ *provided that $f\mathcal{H} \equiv \{fh : h \in \mathcal{H}\} \subseteq \mathcal{H}$. We let $\mathcal{M}(\mathcal{H})$ or $\mathcal{M}(K)$ denote the set of multipliers of $\mathcal{H}$.*

More generally, if $\mathcal{H}_i, i = 1, 2$ are RKHS's on X with reproducing kernels, $K_i, i = 1, 2$, then a function, $f : X \to \mathbb{C}$, such that $f\mathcal{H}_1 \subseteq \mathcal{H}_2$, is called a **multiplier of** $\mathcal{H}_1$ **into** $\mathcal{H}_2$ *and we let $\mathcal{M}(\mathcal{H}_1, \mathcal{H}_2)$ denote the set of multipliers of $\mathcal{H}_1$ into $\mathcal{H}_2$, so that $\mathcal{M}(\mathcal{H}, \mathcal{H}) = \mathcal{M}(\mathcal{H})$.*

Given a multiplier, $f \in \mathcal{M}(\mathcal{H}_1, \mathcal{H}_2)$, we let $M_f : \mathcal{H}_1 \to \mathcal{H}_2$, denote the linear map, $M_f(h) = fh$.

Clearly, the set of multipliers, $\mathcal{M}(\mathcal{H}_1, \mathcal{H}_2)$ is a vector space and the set of multipliers, $\mathcal{M}(\mathcal{H})$, is an algebra.

Proposition 6.2. *Let $\mathcal{H}$ be a RKHS on X with kernel K and let $f : X \to \mathbb{C}$ be a function, let $\mathcal{H}_0 = \{h \in \mathcal{H} : fh = 0\}$ and let $\mathcal{H}_1 = \mathcal{H}_0^{\perp}$. Set $\mathcal{H}_f = f\mathcal{H} = f\mathcal{H}_1$ and define an inner product on $\mathcal{H}_f$ by $\langle fh_1, fh_2\rangle_f = \langle h_1, h_2\rangle$ for $h_1, h_2 \in \mathcal{H}_1$. Then $\mathcal{H}_f$ is a RKHS on X with kernel, $K_f(x, y) = f(x)K(x, y)\overline{f(y)}$.*

Proof. By definition, $\mathcal{H}_f$ is a vector space of functions on X and the linear map, $h \to fh$ is a surjective, linear isometry from $\mathcal{H}_1$ onto $\mathcal{H}_f$. Thus, $\mathcal{H}_f$ is a Hilbert space.

Decomposing, $k_y = k_y^0 + k_y^1$, with $k_y^i \in \mathcal{H}_i$, we have that $K_i(x, y) = k_y^i(x), i = 1, 2$, are the kernels for $\mathcal{H}_i, i = 1, 2$.

To see that $\mathcal{H}_f$ is a RKHS, note that for any fixed $y \in X$ and $h \in \mathcal{H}_1, f(y)h(y) = f(y)\langle h, k_y\rangle = f(y)\langle fh, fk_y\rangle_f = \langle fh, \overline{f(y)}fk_y^1\rangle_f$, which shows that evaluation at y is a bounded linear functional in $\| \cdot \|_f$, and that $K_f(x, y) = \overline{f(y)}f(x)k_y^1(x)$. However, $fk_y^0 = 0$, and hence, $f(x)K_0(x, y)\overline{f(y)} = 0$, from which the result follows. ☐

We are now in a position to characterize multipliers.

Theorem 6.3. *Let $\mathcal{H}_i, i = 1, 2$ be a RKHS's on X with kernels, $K_i, i = 1, 2$ and let $f : X \to \mathbb{C}$. The following are equivalent:*

(i) $f \in \mathcal{M}(\mathcal{H}_1, \mathcal{H}_2)$,
(ii) $f \in \mathcal{M}(\mathcal{H}_1, \mathcal{H}_2)$, *and* M_f *is a bounded operator, i.e.,* $M_f \in B(\mathcal{H}_1, \mathcal{H}_2)$,
(iii) *there exists a constant,* $c \geq 0$, *such that* $f(x)K_1(x,y)\overline{f(y)} \leq c^2 K_2(x,y)$.

Moreover, in these cases, $\|M_f\|$ *is the least constant,* c *satisfying the inequality in (iii).*

Proof. Clearly, (ii) implies (i).

$(i) \Rightarrow (iii)$. By the above Proposition, $\mathcal{H}_f = f\mathcal{H}_1$ with the norm and inner product defined as above is a RKHS with kernel, $K_f(x,y) = f(x)K_1(x,y)\overline{f(y)}$. But since, $\mathcal{H}_f \subseteq \mathcal{H}_2$, by [**?**], there exists a constant $c > 0$, such that, $f(x)K_1(x,y)\overline{f(y)} = K_f(x,y) \leq c^2 K_2(x,y)$.

$(iii) \Rightarrow (ii)$. Since the kernel of the space, $\mathcal{H}_f = f\mathcal{H}_1, K_f(x,y) = f(x)K_1(x,y)\overline{f(y)} \leq c^2 K_2(x,y)$, by [**?**], we have that $f\mathcal{H}_1 \subseteq \mathcal{H}_2$. Now decompose, $\mathcal{H}_1 = \mathcal{H}_{1,0} \oplus \mathcal{H}_{1,1}$, where $f\mathcal{H}_{1,0} = (0)$, and given any $h \in \mathcal{H}_1$, write $h = h_0 + h_1$ with $h_i \in \mathcal{H}_{1,i}, i = 0, 1$. Then $\|fh\|_{\mathcal{H}_2} = \|fh_1\|_{\mathcal{H}_2} \leq c\|fh_1\|_{\mathcal{H}_f} = c\|h_1\|_{\mathcal{H}_{1,1}} \leq c\|h\|_{\mathcal{H}_1}$. Thus, M_f is bounded and $\|M_f\| \leq c$.

Finally, if $\|M_f\| = C$, then for any $h_1 \in \mathcal{H}_{1,1}, \|fh_1\|_{\mathcal{H}_2} \leq C\|h_1\|_{\mathcal{H}_1} = C\|fh_1\|_{\mathcal{H}_f}$. Thus, applying [**?**] again, we have that $f(x)K_1(x,y)\overline{f(y)} = K_f(x,y) \leq C^2 K_2(x,y)$ and the result follows. $\square$

Corollary 6.4. *Let* $\mathcal{H}_i, i = 1, 2$ *be RKHS's on* X *with reproducing kernels,* $K_i(x,y) = k_y^i(x), i = 1, 2$. *If* $f \in \mathcal{M}(\mathcal{H}_1, \mathcal{H}_2)$, *then for every* $y \in X, M_f^*(k_y^2) = \overline{f(y)}k_y^1$.

Thus, when, $K_1 = K_2$, *every kernel function is an eigenvector of* M_f^*. *Moreover, we have that*

$$\|f\|_\infty \leq \|M_f\|,$$

so that every multiplier is a bounded function on X.

Proof. For any $h \in \mathcal{H}_1$, we have that

$$\langle h, \overline{f(y)}k_y^1\rangle_1 = f(y)h(y) = \langle M_f(h), k_y^2\rangle_2 = \langle h, M_f^*(k_y^2)\rangle,$$

and hence, $\overline{f(y)}k_y^1 = M_f^*(k_y^2)$.

Now, if $K_1 = K_2 = K$, then $M_f^*(k_y) = \overline{f(y)}k_y$, and hence, $|f(y)| \leq \|M_f^*\| = \|M_f\|$, and the last inequality follows. $\square$

These problems outline an alternate proof of the above theorem, by showing instead that $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$.

Problem 6.5. *Use the closed graph theorem, to give a direct proof that (i) implies (ii).*

Problem 6.6. *Note that* M_f *is bounded if and only if* M_f^* *is bounded,* $\|M_f\| = \|M_f^*\|$ *and that, in this case,* $M_f^*(k_y^2) = \overline{f(y)}k_y^1$. *Let* $\|M_f^*\| = c$, *and for any points,* $x_1, ..., x_n \in X$ *and any choice of scalars,* $\alpha_1, ..., \alpha_n$, *compute* $\|\sum_i \overline{\alpha_i}k_{x_i}^2\|^2$ *and use it to show that* $f(x)K_1(x,y)\overline{f(y)} \leq c^2 K_2(x,y)$. *Thus, proving that (ii) implies (iii).*

Finally, note that in the above proof that (iii) implies (ii), we first proved that (iii) implies (i).

By the above results, we see that $f \in \mathcal{M}(\mathcal{H})$, then for any point $y \in X$, such that $\|k_y\| \neq 0$, we can recover the values of f by the formula,

$$f(y) = \frac{\langle k_y, M_f^*(k_y)\rangle}{K(y,y)} = \frac{\langle M_f(k_y), k_y\rangle}{K(y,y)}.$$

This motivates the following definition.

Definition 6.7. *Let $\mathcal{H}$ be a RKHS on X with kernel $K(x,y)$, and let $T \in B(\mathcal{H})$. Then the function*

$$B_T(y) = \frac{\langle T(k_y), k_y\rangle}{K(y,y)},$$

defined at any point where $K(y,y) \neq 0$, is called the **Berezin transform of T.**

We present one application of the Berezin transform concept.

For every Hilbert space, $\mathcal{H}$, there is a topology on $B(\mathcal{H})$ called the *weak operator topology.* This topology is characterized by the fact that a net $\{T_\lambda\} \subseteq B(\mathcal{H})$ converges to an operator $T \in B(\mathcal{H})$ if and only if for every pair of vectors, $h, k \in \mathcal{H}, \lim_\lambda \langle T_\lambda(h), k\rangle = \langle T(h), k\rangle$. Note that if $\{T_\lambda\}$ converges in the weak topology to T, then $\{T_\lambda^*\}$ converges in the weak topology to T^*.

Corollary 6.8. *Let $\mathcal{H}$ be a RKHS on X. Then $\{M_f : f \in \mathcal{M}(\mathcal{H})\}$ is a unital subalgebra of $B(\mathcal{H})$ that is closed in the weak operator topology.*

Proof. It is easy to see that the identity operator is the multiplier corresponding to the constant function, 1, and that products and linear combinations of multipliers are multipliers. Thus, this set is a unital subalgebra of $B(\mathcal{H})$.

To see that it is closed in the weak operator topology, we must show that the limit of a net of multipliers is again a multiplier. Let $\{M_{f_\lambda}\}$ be a net of multipliers that converges in the weak operator topology to T. Then for every point y where it is defined, $\lim_\lambda f_\lambda(y) = B_T(y)$.

Set $f(y) = B_T(y)$, whenever $k_y \neq 0$, and $f(y) = 0$, when $k_y = 0$. We claim that $T = M_f$. To see note that $\langle h, T^*(k_y)\rangle = \lim_\lambda \langle h, M_{f_\lambda}^*(k_y)\rangle = \langle h, M_f^*(k_y)\rangle$, and the result follows. □

Problem 6.9. *Show that $B_{T^*} = \overline{B_T}$.*

Problem 6.10. *Show that if $f \in \mathcal{M}(\mathcal{H})$, then $B_{M_f M_f^*} = |f|^2$.*

Problem 6.11. *Let $f(z) = a_0 + a_1 z$, show that $f \in \mathcal{M}(H^2(\mathbb{D}))$ and that $f \in \mathcal{M}(B^2(\mathbb{D}))$. Compute $B_{M_f^* M_f}$, for both these spaces. Deduce that in general, $B_{M_f^* M_f} \neq B_{M_f M_f^*}$.*

Definition 6.12. *Let $G \subseteq \mathbb{C}$ be an open connected set, then $H^\infty(G)$ denotes the functions that are analytic on G and satisfy,*

$$\|f\|_\infty \equiv sup\{|f(z)| : z \in G\} < +\infty.$$

It is not hard to see that $H^\infty(G)$ is an algebra of functions on G that is norm complete and satisfies, $\|fg\|_\infty \leq \|f\|_\infty \|g\|_\infty$, that is, $H^\infty(G)$ is a Banach algebra.

Theorem 6.13. *Let $G \subseteq \mathbb{C}$ be a bounded, open set, then $\mathcal{M}(B^2(G)) = H^\infty(G)$, and for $f \in \mathcal{M}(B^2(G)), \|M_f\| = \|f\|_\infty$.*

Proof. Since the constant function, $1 \in B^2(G)$, we have that if $f \in \mathcal{M}(B^2(G))$, then $f = f \cdot 1 \in B^2(G)$. Thus, $\mathcal{M}(B^2(G)) \subseteq B^2(G)$, and so every function in $\mathcal{M}(B^2(G))$ is analytic on G.

Moreover, by the above Corollary, $\|f\|_\infty \leq \|M_f\|$, and hence, $\mathcal{M}(B^2(G)) \subseteq H^\infty(G)$.

Conversely, if $f \in H^\infty(G)$ and $h \in B^2(G)$, then

$$\|fh\|^2_{B^2(G)} = \int\int |f(x+iy)h(x+iy)|^2 dxdy$$
$$\leq \|f\|^2_\infty \int\int |h(x+iy)|^2 dxdy = \|f\|^2_\infty \|h\|^2_{B^2(G)}.$$

Thus, $f \in \mathcal{M}(B^2(G))$ with $\|M_f\| \leq \|f\|_\infty$ and the result follows. ☐

Problem 6.14. *Let $G \subseteq \mathbb{C}$ be an open set and assume that there are enough functions in $B^2(G)$ to separate points on G. Prove that $\mathcal{M}(B^2(G)) = H^\infty(G)$, and that $\|f\|_\infty = \|M_f\|$.*

We now turn our attention to determining the multipliers of $H^2(\mathbb{D})$. For this we will need the identification of the Hardy space of the disk with the Hardy space of the unit circle. Recall that if we endow the unit circle in the complex plane, $\mathbb{T}$, with normalized arc length measure, then

$$\int_{\mathbb{T}} f(s)ds = \frac{1}{2\pi}\int_0^{2\pi} f(e^{it})dt$$

and that with respect to this measure, the functions, $e_n(e^{it}) = e^{int}, n \in \mathbb{Z}$, form an orthonormal basis for $L^2(\mathbb{T})$. Given any function in $f \in L^p(\mathbb{T}), 1 \leq p \leq +\infty$, we can define its *Fourier coefficients* by the formula,

$$\hat{f}(n) = \frac{1}{2\pi}\int_0^{2\pi} f(e^{it})e^{-int}dt.$$

The maps, $\psi_n : L^p(\mathbb{T}) \to \mathbb{C}$ defined by $\psi_n(f) = \hat{f}(n)$, are bounded linear functionals, and hence, the set

$$H^p(\mathbb{T}) \equiv \{f \in L^p(\mathbb{T}) : \hat{f}(n) = 0, \text{for all } n \leq 0\},$$

is a norm closed subspace for $1 \leq p \leq +\infty$ which is also weak*-closed in the case when $1 < p \leq +\infty$. These spaces are called the *Hardy spaces of* *the circle.* Note that $H^2(\mathbb{T})$ is the Hilbert space with orthonormal basis, $e_n, n \geq 0$.

If $f \in H^p(\mathbb{T})$, then its *Cauchy transform* defined by the formula,

$$\tilde{f}(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{f(e^{it})}{1 - ze^{-it}} dt$$

is easily seen to be an analytic function on the disk. By comparing orthonormal bases, we can see that the Cauchy transform defines a Hilbert space isomorphism, i.e., an onto isometry, between $H^2(\mathbb{T})$ and the space $H^2(\mathbb{D})$. It is somewhat more difficult to show, but the Cauchy transform also defines an isometry from $H^\infty(\mathbb{T})$ onto $H^\infty(\mathbb{D})$. See [**?**] for a proof of this fact. For these reasons the set of functions obtained as the Cauchy transforms of $H^p(\mathbb{T})$ are denoted, $H^p(\mathbb{D})$ and referred to as the *Hardy spaces of the unit disk.*

One other fact about the Hardy spaces that we shall need is that if for $f \in H^p(\mathbb{T})$ and $0 \leq r < 1$, we define $f_r(e^{it}) = \tilde{f}(re^{it})$, then $f_r \in H^p(\mathbb{T}), 1 \leq p \leq +\infty$, with $\|f_r\|_p \leq \|f\|_p$, and for $1 \leq p < +\infty, lim_{r \to 1} \|f - f_r\|_p = 0$. In the case $p = +\infty$, the functions, f_r, converge to f in the weak*-topology, but not necessarily the norm topology. In fact, they converge in the norm topology to f if and only if f is equal almost everywhere to a continuous function.

With these preliminaries out of the way we can now characterize the multipliers of $H^2(\mathbb{D})$.

Theorem 6.15. $\mathcal{M}(H^2(\mathbb{D})) = H^\infty(\mathbb{D})$, *and for* $f \in H^\infty(\mathbb{D}), \|M_f\| = \|f\|_\infty$.

Proof. Let $f \in \mathcal{M}(H^2(\mathbb{D}))$, then $f = f \cdot 1 \in H^2(\mathbb{D})$ and hence, f is analytic on $\mathbb{D}$. Moreover, since, $\|f\|_\infty \leq \|M_f\|$, we see that $f \in H^\infty(\mathbb{D})$.

Now let, $f \in H^\infty(\mathbb{D})$, so that $f = \tilde{g}$, for some function, $g \in H^\infty(\mathbb{T})$ and $\|f\|_\infty = \|g\|_\infty$, where the first norm is the supremum over the disk and the second is the essential supremum over the unit circle. Since g is essentialy bounded, we see that $M_g : H^2(\mathbb{T}) \to L^2(\mathbb{T})$ is bounded and $\|M_g\| \leq \|g\|_\infty$. By computing Fourier coefficients of ge_n, one sees that $ge_n \in H^2(\mathbb{T})$, for all $n \geq 0$, and hence that $g \cdot H^2(\mathbb{T}) \subseteq H^2(\mathbb{T})$. Also, by comparing Fourier coefficients, one sees that $\widetilde{ge_n}(z) = f(z)z^n$, and thus for any $h \in H^2(\mathbb{T}), f(z)\tilde{h}(z) = \widetilde{gh}(z) \in H^2(\mathbb{D})$. Thus, $f \in \mathcal{M}(H^2(\mathbb{D}))$ and $\|M_f\| = \|M_g\| \leq \|g\|_\infty = \|f\|_\infty$, and the result follows. □

We now take a look at some special multipliers of $H^2(\mathbb{D})$.

Definition 6.16. *A function,* $h \in L^\infty(\mathbb{T})$ *is called an* **inner function** *if* $|h(e^{it})| = 1$, *a.e. and an inner function that is in* $H^\infty(\mathbb{T})$ *is called an* **analytic inner function.** *More generally,* $g \in H^\infty(\mathbb{D})$ *is called an* **analytic inner function** *if* g *is the Cauchy transform of an analytic inner function, that is, if* $g = \tilde{h}$ *for some inner function* $h \in H^\infty(\mathbb{T})$.

Let $\alpha_1, ..., \alpha_n \in \mathbb{D}$, be distinct points, let $\varphi_{\alpha_i}(z) = \frac{z - \alpha_i}{1 - \bar{\alpha}_i z}$, denote the corresponding Mobius maps and let $m_1, ..., m_n$ be integers with $m_i \geq 1, i =$

$1, ..., n$. Then the function,

$$B(z) = \prod_{i=1}^{n} \varphi_{\alpha_i}^{m_i}(z),$$

is called the **Blaschke product** corresponding to the points, $\alpha_1, ..., \alpha_n$, and multiplicities, $m_1, ..., m_n$.

Since, $|\varphi_\alpha(e^{it})| = |e^{-it}\frac{e^{it}-\alpha}{1-\bar{\alpha}e^{it}}| = |\frac{1-e^{-it}\alpha}{1-\bar{\alpha}e^{it}}| = 1$, we see that every Mobius map and more generally, every Blaschke product is an analytic inner function.

Proposition 6.17. *Let $f \in H^\infty(\mathbb{D})$ be an inner function, then $M_f : H^2(\mathbb{D}) \to H^2(\mathbb{D})$ is an isometry and the range of M_f is the RKHS with kernel, $\frac{f(z)\overline{f(w)}}{1-\bar{w}z}$.*

Proof. Let $f = \tilde{g}$, then for any $h \in H^2(\mathbb{T})$, we have that,

$$\|gh\|^2 = \frac{1}{2\pi}\int_0^{2\pi} |g(e^{it})h(e^{it})|^2 dt = \frac{1}{2\pi}\int_0^{2\pi} |h(e^{it})|^2 dt = \|h\|^2,$$

and so M_g is an isometry. By Proposition ??, the kernel function for the range space is as given. □

We now wish to concretely identify the subspace of $H^2(\mathbb{D})$ that is the range of a Blaschke product. We first need a preliminary result.

Given an integer, $m \geq 1$, we let $f^{(m)}$ denote the $m - th$ derivative of a function f.

Proposition 6.18. *Let $w \in \mathbb{D}$ and let $m \geq 1$, be an integer. Then $z^m(1 - \bar{w}z)^{-m-1} \in H^2(\mathbb{D})$ and for $f \in H^2(\mathbb{D}), \langle f, z^m(1 - \bar{w}z)^{-m-1}\rangle = f^{(m)}(w)$. Thus, the map, $E_w^{(m)} : H^2(\mathbb{D}) \to \mathbb{C}$, defined by $E_w^{(m)}(f) = f^{(m)}(w)$, is a bounded linear functional.*

Proof. We have that

$$z^m(1-\bar{w}z)^{-m-1} = (\frac{\partial}{\partial\bar{w}})^m(1-\bar{w}z)^{-1} = (\frac{\partial}{\partial\bar{w}})^m \sum_{n=0}^{\infty}(\bar{w}z)^n$$
$$= \sum_{n=m}^{\infty} n(n-1)\cdots(n-(m-1))\bar{w}^{n-m}z^n,$$

which can be seen to be square-summable.

Given any $f(z) = \sum_{n=0}^\infty a_n z^n$, we have that $\langle f, z^m(1 - \bar{w}z)^{-m-1}\rangle = \sum_{n=m}^\infty a_n n(n-1)\cdots(n-(m-1))w^{n-m} = f^{(m)}(w)$, and the result follows. □

Theorem 6.19. *Let $\alpha_1, \ldots, \alpha_n \in \mathbb{D}$, be distinct points, let $m_i \geq 1$, be integers, and let $B(z)$ be the corresponding Blaschke product. Then*

$$B(z) \cdot H^2(\mathbb{D}) = \{f \in H^2(\mathbb{D}) : f^{(j)}(\alpha_k) = 0, 1 \leq k \leq n, 1 \leq j \leq m_k\},$$

and consequently, the reproducing kernel for this subspace is

$$\frac{B(z)\overline{B(w)}}{1-\bar{w}z}.$$

Proof. By the above result we know that the reproducing kernel for the subspace $B(z)\cdot H^2(\mathbb{D})$ is given by $\frac{B(z)\overline{B(w)}}{1-\bar{w}z}$. Thus, it remains to prove the set equality.

It is clear that $B(z)\cdot H^2(\mathbb{D})$ is contained in the set on the right. Conversely, any function in the set on the right can be written as, $f(z) = (z-\alpha_1)^{m_1}\cdots(z-\alpha_n)^{m_n}g(z)$, where g is analytic on $\mathbb{D}$. Setting, $h(z) = (1-\bar{\alpha_1}z)^{m_1}\cdots(1-\bar{\alpha_n}z)^{m_n}g(z)$, we have that h is analytic on $\mathbb{D}$ and $f(z) = B(z)h(z)$ and it remains to show that $h\in H^2(\mathbb{D})$.

But we know that there exists a function, $f_1\in H^2(\mathbb{T})$ such that for almost all $t, \lim_{r\to 1} f(re^{it}) = f_1(e^{it})$. Thus, setting $h_1(e^{it}) = \lim_{r\to 1}\overline{B(re^{it})}f(re^{it}) = \overline{B(e^{it})}f_1(e^{it})$, defines a function a.e. on the circle, with $|h_1(e^{it})| = |f_1(e^{it})|$, *a.e.*, from which it follows that $h_1\in L^2(\mathbb{T})$.

However, $\lim_{r\to 1} B(re^{it})^{-1} = \overline{B(e^{it})}$, and hence, $\lim_{r\to 1} h(re^{it}) = h_1(e^{it})$, *a.e.*, from which it follows that $h_1\in H^2(\mathbb{D})$ and that h is the Cauchy transform of h_1. Thus, $h\in H^2(\mathbb{D})$ and the theorem follows. □

Problem 6.20. *Prove that for every $w\in\mathbb{D}$, and integer $m\geq 1$, the map $E_w^{(m)}: B^2(\mathbb{D})\to\mathbb{C}$ is bounded and that it is given by the inner product with $(\frac{\partial}{\partial\bar{w}})^m(1-\bar{w}z)^2$.*

7. Negative Definite Functions

Definition 7.1. *The function $\psi: X\times X\to\mathbb{C}$ is called* **negative definite** *provided $\forall x_1,x_2,\ldots,x_n\in X$, $\forall\alpha_1,\alpha_2,\ldots,\alpha_n\in\mathbb{C}$ such that $\sum_{j=1}^n\alpha_j = 0$, we have $\sum_{i,j=1}^n\bar{\alpha}_i\alpha_j\psi(x_i,x_j)\leq 0$.*

Proposition 7.2 (Berg, Christensen, Bessel)**.** *Let $\psi: X\times X\to\mathbb{C}$, fix $x_0\in X$. Then ψ is negative definite if and only if $K(x,y) = \psi(x,x_0) + \overline{\psi(y,x_0)} - \psi(x,y) - \psi(x_0,x_0)$ is a kernel function.*

Proof. $(\Leftarrow)$: Let $\sum_{j=1}^n\alpha_j = 0$, then we get

$$0\leq\sum_{i,j=1}^n\bar{\alpha}_i\alpha_jK(x_i,x_j) = \sum_{i,j=1}^n\bar{\alpha}_i\alpha_j\psi(x_i,x_0) + \sum_{i,j=1}^n\bar{\alpha}_i\alpha_j\overline{\psi(x_j,x_0)}$$
$$-\sum_{i,j=1}^n\bar{\alpha}_i\alpha_j\psi(x_i,x_j) - \sum_{i,j=1}^n\bar{\alpha}_i\alpha_j\psi(x_0,x_0) = 0+0-\sum_{i,j=1}^n\bar{\alpha}_i\alpha_j\psi(x_i,x_j) - 0.$$

This implies $\sum_{i,j=1}^n\bar{\alpha}_i\alpha_j\psi(x_i,x_j)\leq 0$, therefore ψ is a negative function.
$(\Rightarrow)$ Given $\alpha_1,\alpha_2,\ldots,\alpha_n\in\mathbb{C}$, let $\alpha_0 = -\sum_{j=1}^n\alpha_j$, then $\sum_{j=0}^n\alpha_j = 0$. Therefore $\sum_{i,j=0}^n\bar{\alpha}_i\alpha_jK(x_i,x_j) = 0+0-\sum_{i,j=0}^n\bar{\alpha}_i\alpha_j\psi(x_i,x_j)+0$,
$\Rightarrow [K(x_i,x_j)]_{i,j=0}^n\geq 0\Rightarrow [K(x_i,x_j)]_{i,j=1}^n\geq 0$. □

Theorem 7.3 (Schoenberg, 1940). *Let (X,d) be a metric space. Then $\exists\mathcal{H}$ a Hilbert space and $\phi : X \to \mathcal{H}$ an isometry i.e. $\|\phi(x)-\phi(y)\| = d(x,y)$ if and only if $d^2(x,y)$ is negative definite.*

Proof. $(\Rightarrow)$: Let $\sum_{j=1}^n \alpha_j = 0$, then we have

$$\sum_{i,j=1}^n \bar{\alpha}_i\alpha_j d^2(x_i,x_j) = \sum_{i,j=1}^n \bar{\alpha}_i\alpha_j\|\phi(x_i)-\phi(x_j)\|^2 = \sum_{i,j=1}^n \bar{\alpha}_i\alpha_j\langle\phi(x_i){-}\phi(x_j), \phi(x_i){-}\phi(x_j)\rangle$$

$$= \sum_{i,j=1}^n \bar{\alpha}_i\alpha_j\langle\phi(x_i),\phi(x_i)\rangle - \sum_{i,j=1}^n \bar{\alpha}_i\alpha_j\langle\phi(x_i),\phi(x_j)\rangle - \sum_{i,j=1}^n \bar{\alpha}_i\alpha_j\langle\phi(x_j),\phi(x_i)\rangle$$

$$+ \sum_{i,j=1}^n \bar{\alpha}_i\alpha_j\langle\phi(x_j),\phi(x_j)\rangle = 0 - \|\sum_{i=1}^n \bar{\alpha}_i\phi(x_i)\|^2 - \|\sum_{j=1}^n \alpha_j\phi(x_j)\|^2 + 0 \le 0.$$

$(\Leftarrow)$: By previous proposition (BCB),

$$K(x,y) = d^2(x,x_0) + \overline{d^2(y,x_0)} - d^2(x,y) - d^2(x_0,x_0)$$
$$\Rightarrow K(x,y) = d^2(x,x_0) + d^2(y,x_0) - d^2(x,y)$$

is a kernel function. Look at $\mathcal{H}(K)$, let $\phi(x) = \frac{\sqrt{2}}{2}k_x$.
Note that $K(x,y) = K(y,x)$. Compute

$$\|\phi(x)-\phi(y)\|^2 = \frac{1}{2}\langle k_x{-}k_y, k_x{-}k_y\rangle = \frac{1}{2}\Big\{K(x,x){-}K(x,y){-}K(y,x){+}K(y,y)\Big\}$$

$$= \frac{1}{2}\Big\{[d^2(x,x_0)+d^2(x,x_0)-d^2(x_0,x_0)] - 2[d^2(x,x_0)+d^2(y,x_0)-d^2(x,y)]$$

$$+[d^2(y,x_0){+}d^2(y,x_0){-}d^2(y,y)]\Big\} = d^2(x,y) \Rightarrow \phi(x) = \frac{\sqrt{2}}{2}k_x \quad \text{is an isometry.}$$

□

Problem 7.4. *Show $\psi(x,y) = \big(sin(x-y)\big)^2$ is negative definite on $\mathbb{R}$.*

Theorem 7.5 (Schoenberg). *The function $\psi : X \times X \to \mathbb{C}$ is negative definite if and only if $e^{-t\psi}$ is a kernel function $\forall t > 0$.*

Proof. $(\Leftarrow)$: Note that $K_t(x,y) = e^{-t\psi(x,y)}$ is a kernel function $\forall t > 0$.
Then $\gamma_t(x,y) = 1 - K_t(x,y)$ is negative definite function:
Let $\sum_{i=1}^n \alpha_i = 0$, then

$$\sum_{i,j=1}^n \bar{\alpha}_i\alpha_j\gamma_t(x_i,x_j) = \sum_{i,j=1}^n \bar{\alpha}_i\alpha_j{\cdot}1 - \sum_{i,j=1}^n \bar{\alpha}_i\alpha_j K_t(x_i,x_j) = -\sum_{i,j=1}^n \bar{\alpha}_i\alpha_j K_t(x_i,x_j) \le 0$$

Now, $\gamma_t(x,y)$ being negative definite $\forall t > 0$ implies $\frac{\gamma_t(x,y)}{t}$ negative definite, i.e.

$$\lim_{t\to 0}\frac{\gamma_t(x,y)}{t} = \lim_{t\to 0}\frac{1-e^{-t\psi(x,y)}}{t} = \psi(x,y) \quad \text{is negative definite (if it exists).}$$

$(\Rightarrow)$: If ψ is negative definite, then $t\psi$ is negative definite
$\Rightarrow e^{-t\psi}$ is a kernel. Therefore, assuming ψ negative definite, it's enough to show $e^{-\psi}$ is a kernel. Hence, let ψ be negative definite, then (by BCB)
$\Rightarrow K(x,y) = \psi(x,x_0) + \overline{\psi(y,x_0)} - \psi(x,y) - \psi(x_0,x_0)$ is a kernel,
$\Rightarrow \tilde{K}(x,y) = e^{K(x,y)} = 1 + K(x,y) + \frac{K^2(x,y)}{2!} + \cdots$ is a kernel too.
Next, we have

$$\begin{aligned}\tilde{K}(x,y) &= e^{\left(\psi(x,x_0)+\overline{\psi(y,x_0)}-\psi(x,y)-\psi(x_0,x_0)\right)} \\ &= e^{\psi(x,x_0)} \cdot e^{\overline{\psi(y,x_0)}} \cdot e^{-\psi(x,y)} \cdot e^{-\psi(x_0,x_0)}\end{aligned}$$

$\Rightarrow e^{-\psi(x,y)} = e^{\psi(x_0,x_0)} \cdot e^{-\psi(x,x_0)} \cdot \tilde{K}(x,y) \cdot e^{-\overline{\psi(y,x_0)}}$
$\Rightarrow e^{-\psi(x,y)} = \alpha_0 \cdot f(x) \cdot \tilde{K}(x,y) \cdot \overline{f(y)}$, where $f(x) = e^{-\psi(x,x_0)}$,
$\Rightarrow e^{-\psi}$ is a kernel function. □

Example: It can happen that a matrix $A = (a_{ij}) \geq 0$, but $(|a_{ij}|)$ is **not** positive definite.

$$\left|\begin{pmatrix} 1 & \frac{1}{3} & \frac{z}{3} \\ \frac{1}{3} & 1 & \frac{1}{3} \\ \frac{\bar{z}}{3} & \frac{1}{3} & 1 \end{pmatrix}\right| = 1 \cdot \frac{8}{9} - \frac{1}{3}\left(\frac{1}{3} - \frac{\bar{z}}{9}\right) + \frac{z}{3}\left(\frac{1}{9} - \frac{\bar{z}}{3}\right) = \frac{7}{9} + \frac{\bar{z}}{27} + \frac{z}{27} - \frac{|z|^2}{9}$$

for $z = \sqrt{7}e^{i\theta}$.

Suppose $A \in M_n(\mathbb{C})$, set $X = \{1, \ldots, n\}$, define $K : X \times X \to \mathbb{C}$ by $K(i,j) = a_{ij}$, then K is kernel, but $|K|$ is not a kernel.
On the other hand, if $K : X \times X \to \mathbb{C}$ is a kernel, then $\bar{K}$ is kernel.
Because $\sum_{i,j=1}^n \bar{\alpha}_i \alpha_j \overline{K(x_i,x_j)} = \sum_{i,j=1}^n \bar{\alpha}_i \alpha_j K(x_j,x_i) \geq 0$,
$\Rightarrow K(x,y) \cdot \bar{K}(x,y) = |K|^2(x,y)$ is a kernel.
But, if we had a negative definite function $\psi : X \times X \to \mathbb{C}$, $K(x,y) = e^{-\psi(x,y)}$ kernel, look at $K_2(x,y) = e^{-\frac{\psi(x,y)}{2}}$ kernel $\Rightarrow K(x,y) = \big(K_2(x,y)\big)^2$.
Then $|K(x,y)| = |K_2(x,y)^2| = |K_2(x,y)|^2 \Rightarrow |K|$ is kernel.

Definition 7.6. *Let $K : X \times X \to \mathbb{C}$ be a kernel. Then K is **infinitely divisible** provided that $\forall n \in \mathbb{N}$, $\exists K_n : X \times X \to \mathbb{C}$ kernel such that $(K_n)^n = K$.*

Proposition 7.7. *Let $K : X \times X \to \mathbb{C}$ be an infinitely divisible kernel, via $K_n's$. Then $|K_n|$ is a kernel, $\forall n \in \mathbb{N}$, and $|K|$ is a kernel, too.*

Proof.

$$\begin{gathered} K = (K_n)^n = (K_{2n})^{2n} \Rightarrow |K| = |K_n|^n = |K_{2n}|^{2n} \Rightarrow |K_n| = |K_{2n}|^2 \\ \Rightarrow |K_n| \text{ is a kernel}, \forall n \text{ and } |K| = |K_n|^n \text{ is a kernel, too.} \\ \text{Or } |K| = |K_2|^2 \Rightarrow |K| \text{ is a kernel.} \end{gathered}$$

□

Theorem 7.8. *Let $K : X \times X \to \mathbb{C}$ be a kernel, $K(x,y) > 0 \quad \forall x, y$. Then K is infinitely divisible $\iff$ $\psi(x,y) = -\log(K(x,y))$ is negative definite.*

Proof. $(\Leftarrow)$: Let $\psi(x,y) = -\log(K(x,y))$ be negative definite. Then $e^{-\psi(x,y)} = K(x,y)$ is positive definite. Set $K_n(x,y) = e^{-\frac{\psi(x,y)}{n}} \Rightarrow K$ is $\infty -$ divisible.
$(\Rightarrow)$: From previous proposition, we have

$$K(x,y) = K_n(x,y)^n = |K_n(x,y)|^n.$$

We know $|K_n|$ are all kernel functions and $|K| = |K_n|^n$. Then

$$\psi(x,y) = -\log(K(x,y)) = -n\log(|K_n|(x,y)).$$

Let $t = \frac{1}{n}$, $\quad \left(e^{-\frac{1}{n}\psi(x,y)}\right)^n = K(x,y) \Rightarrow e^{-\frac{\psi(x,y)}{n}} = K(x,y)^{1/n} = |K_n|$ kernel function, $\therefore e^{-\frac{m}{n}\psi(x,y)} = \left(e^{-\frac{\psi(x,y)}{n}}\right)^m = |K_n|^m(x,y)$ is kernel.
$\Rightarrow e^{-t\psi(x,y)} > 0, \quad \forall t > 0$ (taking limits of rationals)
$\Rightarrow \psi(x,y)$ is negative definite. $\square$

8. Block or Operator Matrices

Let $\mathcal{H}, \mathcal{K}$ be any two Hilbert spaces, then $\mathcal{H} \oplus \mathcal{K} = \left\{ \begin{pmatrix} h \\ k \end{pmatrix} : h \in \mathcal{H}, k \in \mathcal{K} \right\}$.
Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in B(\mathcal{H} \oplus \mathcal{K})$,
where $A_{11} \in B(\mathcal{H}), \quad A_{12} \in B(\mathcal{K}, \mathcal{H}), \quad A_{21} \in B(\mathcal{H}, \mathcal{K}), \quad A_{22} \in B(\mathcal{K})$.

$$A \begin{pmatrix} h \\ k \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = \begin{bmatrix} A_{11}h + A_{12}k \\ A_{21}h + A_{22}k \end{bmatrix}.$$

Proposition 8.1. *Let $A \in B(\mathcal{K}, \mathcal{H})$. Then $\|A\| \le 1 \iff \begin{bmatrix} I_{\mathcal{H}} & A \\ A^* & I_{\mathcal{K}} \end{bmatrix} \ge 0$.*

Proof. $(\Leftarrow)$:

$$0 \le \langle \begin{bmatrix} I_{\mathcal{H}} & A \\ A^* & I_{\mathcal{K}} \end{bmatrix} \begin{bmatrix} -Ak \\ k \end{bmatrix}, \begin{bmatrix} -Ak \\ k \end{bmatrix} \rangle = \langle \begin{bmatrix} -Ak + Ak \\ -A^*Ak + k \end{bmatrix}, \begin{bmatrix} -Ak \\ k \end{bmatrix} \rangle$$

$$= \langle -A^*Ak + k, k \rangle = -\langle Ak, Ak \rangle + \langle k, k \rangle = -\|Ak\|^2 + \|k\|^2.$$

So, we get $0 \le -\|Ak\|^2 + \|k\|^2 \Rightarrow \|Ak\|^2 \le \|k\|^2 \quad \forall k \Rightarrow \|A\| \le 1$.

$(\Rightarrow)$:

$$\left\langle \begin{bmatrix} I_{\mathcal{H}} & A \\ A^* & I_{\mathcal{K}} \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}, \begin{bmatrix} h \\ k \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} h + Ak \\ A^*h + k \end{bmatrix}, \begin{bmatrix} h \\ k \end{bmatrix} \right\rangle$$

$$= \langle h, h\rangle + \langle Ak, h\rangle + \langle A^*h, k\rangle + \langle k, k\rangle \geq \|h\|^2 + \|k\|^2 - 2\|Ak\|\|h\|$$

$$= \big(\|h\| - \|k\|\big)^2 + 2\|h\| \underbrace{\big(\|k\| - \|Ak\|\big)}_{\geq 0,\text{ A contractive}} \geq 0.$$

□

Example: Let $\mathcal{U}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, $\quad 0 \leq \theta \leq \frac{\pi}{2}$, $\quad \|\mathcal{U}_\theta\| = 1$,
let $\mathcal{B} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \cos\theta \cdot I + \sin\theta \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$,
$\sigma(\mathcal{B}) = \cos\theta \pm \sin\theta$, $\quad \|\mathcal{B}\| = \cos\theta + \sin\theta > 1$, $\quad 0 < \theta < \frac{\pi}{2}$.
If $0 \leq \theta \leq \frac{\pi}{4}$, then $\|\mathcal{B}\| = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = \sqrt{2} > 1$, $\quad \theta = \frac{\pi}{4}$.

Note that from a previous example, we have $P = (p_{ij}) \geq 0$,
but $(|p_{ij}|)$ not positive, as it is demonstrated below:

$$\begin{pmatrix} 1 & 0 & \cos\theta & -\sin\theta \\ 0 & 1 & \sin\theta & \cos\theta \\ \cos\theta & \sin\theta & 1 & 0 \\ -\sin\theta & \cos\theta & 0 & 1 \end{pmatrix} \geq 0, \quad \text{but} \begin{pmatrix} 1 & 0 & \cos\theta & \sin\theta \\ 0 & 1 & \sin\theta & \cos\theta \\ \cos\theta & \sin\theta & 1 & 0 \\ \sin\theta & \cos\theta & 0 & 1 \end{pmatrix} \ngeq 0.$$

9. RKHS and Cholesky's Algorithm

Lemma 9.1 (Cholesky). $P = (p_{ij})_{i,j=0}^n \geq 0 \iff R = P - \left(\dfrac{\bar{p}_{0i} p_{0j}}{p_{00}}\right) \geq 0.$

Note: Look at $(0,i)^{th}$ entry of R, we get $r_{0j} = p_{0j} - \dfrac{\bar{p}_{00} p_{0j}}{p_{00}} = 0$,
$r_{i0} = p_{i0} - \dfrac{\bar{p}_{0i} p_{00}}{p_{00}} = 0$, then R looks as $R = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \bar{P}_{n\times n} & \\ 0 & & \end{pmatrix}$.

Cholesky algorithm is really used for factoring positive square matrices $P \geq 0$ as $P = U^*U$, where U is an upper triangular matrix. Here is how it works:

Example: Deduce whether $P = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 6 & 7 \\ 3 & 7 & 12 \end{pmatrix} \geq 0$. If yes, factorize it.

Solution: Calculate

$$R = P - \frac{1}{1}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} \Rightarrow \tilde{P} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}.$$

Then

$$\tilde{R} = \tilde{P} - \frac{1}{2}\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 5/2 \end{pmatrix} \Rightarrow P \geq 0.$$

Therefore

$$P = \underbrace{\frac{1}{1}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}}_{\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}} + \underbrace{\frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix}}_{\begin{pmatrix} \frac{2}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \frac{2}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 5/2 \end{pmatrix}.$$

i.e.

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 2 & \frac{2}{\sqrt{2}} & 0 \\ 3 & \frac{1}{\sqrt{2}} & \sqrt{\frac{5}{2}} \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 0 & \frac{2}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \sqrt{\frac{5}{2}} \end{pmatrix} = U^*U.$$

Recall: Let $\mathcal{H}$ be an RKHS and K its kernel, $x_0 \in X$, $\mathcal{M} = \{f \in \mathcal{H} : f(x_0) = 0\}$, then $K_{\mathcal{M}} = K_{\mathcal{H}} - K_{\mathcal{M}^\perp} = K(x,y) - \dfrac{k_{x_0}(x)\overline{k_{x_0}(y)}}{K(x_0,x_0)} \geq 0$, since $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$, $\quad \mathcal{M}^\perp = span\{k_{x_0}\}$, $\quad K_{\mathcal{M}^\perp} = \dfrac{k_{x_0}(x)\overline{k_{x_0}(y)}}{\|k_{x_0}\|^2}$.

Cholesky's lemma's proof: $(\Rightarrow)$: Let $P = (p_{ij}) \geq 0$, and $X = \{0, 1, \ldots, n\}$. Denote $K(i,j) = p_{ij}$, i.e. K is a kernel function.

Then $K_{\mathcal{M}}(i,j) = K(i,j) - \dfrac{K(i,0)\overline{K(j,0)}}{K(0,0)}$

$\Rightarrow (K_{\mathcal{M}}(i,j)) = (p_{ij}) - \left(\dfrac{p_{i0}\bar{p}_{0j}}{p_{00}}\right) \geq 0, \quad p_{i0} = \bar{p}_{0i}.$

$(\Leftarrow)$: Assume $R = P - \left(\dfrac{p_{0i}\bar{p}_{0j}}{p_{00}}\right) \geq 0$

$\Rightarrow P = R + \left(\dfrac{p_{0i}\bar{p}_{0j}}{p_{00}}\right) \geq 0$ as a sum of 2 positive matrices. $\square$

Recall bijective **Moebius maps** $\varphi_\alpha : \mathbb{D} \to \mathbb{D}, \quad \alpha \in \mathbb{D}, \quad \varphi_\alpha(z) = \dfrac{z-\alpha}{1-\bar{\alpha}z}$.
Note that $\varphi_\alpha(\alpha) = 0, \quad \varphi_{-\alpha} = {\varphi_\alpha}^{-1}$.
Also, if $f : \mathbb{D} \to \mathbb{D}, \quad \alpha, \beta \in \mathbb{D}, \quad f(z_i) = \lambda_i, \quad \varphi_\alpha(\lambda_i) = \mu_i, \quad \varphi_\beta(z_i) = \zeta_i$, then $h = \varphi_\alpha \circ f \circ \varphi_{-\beta} : \mathbb{D} \to \mathbb{D}$ gives $h(\zeta_i) = \mu_i$.

Lemma 9.2. $\left(\dfrac{1-\lambda_i\bar{\lambda}_j}{1-z_i\bar{z}_j}\right) \geq 0 \quad \Rightarrow \quad \left(\dfrac{1-\varphi_\alpha(\lambda_i)\overline{\varphi_\alpha(\lambda_j)}}{1-\varphi_\beta(z_i)\overline{\varphi_\beta(z_j)}}\right) \geq 0.$

Proof. $\left(\dfrac{1-\varphi_\alpha(\lambda_i)\overline{\varphi_\alpha(\lambda_j)}}{1-z_i\bar{z}_j}\right) = \left(\dfrac{1-\left(\dfrac{\lambda_i-\alpha}{1-\bar{\alpha}\lambda_i}\right)\left(\dfrac{\bar{\lambda}_j-\bar{\alpha}}{1-\alpha\bar{\lambda}_j}\right)}{1-z_i\bar{z}_j}\right)$

$$= \left(\frac{1}{1-\bar{\alpha}\lambda_i} \cdot \frac{(1-|\alpha|^2)(1-\lambda_i\bar{\lambda}_j)}{1-z_i\bar{z}_j} \cdot \frac{1}{1-\alpha\bar{\lambda}_j}\right)$$

$$= (1-|\alpha|^2) \cdot \underbrace{\left(\frac{1}{1-\bar{\alpha}\lambda_i}\right)}_{D} \cdot \underbrace{\left(\frac{1-\lambda_i\bar{\lambda}_j}{1-z_i\bar{z}_j}\right)}_{\geq 0} \cdot \underbrace{\left(\frac{1}{1-\alpha\bar{\lambda}_j}\right)}_{D^*}$$

$$= (1-|\alpha|^2) \cdot D \cdot \left(\frac{1-\lambda_i\bar{\lambda}_j}{1-z_i\bar{z}_j}\right) \cdot D^* \geq 0.$$

Similarly, we get

$$\left(\frac{1-\lambda_i\bar{\lambda}_j}{1-\varphi_\beta(z_i)\overline{\varphi_\beta(z_j)}}\right) = \frac{1}{1-|\beta|^2} \cdot \underbrace{(1-\bar{\beta}z_i)}_{\tilde{D}} \cdot \underbrace{\left(\frac{1-\lambda_i\bar{\lambda}_j}{1-z_i\bar{z}_j}\right)}_{\geq 0} \cdot \underbrace{(1-\beta\bar{z}_j)}_{\tilde{D}^*} \geq 0.$$

□

Theorem 9.3 (Pick's theorem)**.** *Given $z_1, z_2, \ldots, z_n \in \mathbb{D}$ such that $z_i \neq z_j \quad \forall i \neq j$, $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{D}$, then $\exists f \in \mathbb{H}^\infty(\mathbb{D}), \quad \|f\|_\infty \leq 1$ such that $f(z_i) = \lambda_i, \quad 1 \leq i \leq n \iff \left(\dfrac{1-\lambda_i\bar{\lambda}_j}{1-z_i\bar{z}_j}\right) \geq 0$.*

Proof. $(\Rightarrow) : \|f\|_\infty = \|M_f\|_{\mathbb{H}^2(\mathbb{D})} \leq 1 \Rightarrow \left(1-f(z)\overline{f(w)}\right) K_{\mathbb{H}^2}(z,w)$ is a kernel function. Thus, evaluating at $z_1, z_2, \ldots, z_n$, we get $\left(\dfrac{1-f(z_i)\overline{f(z_j)}}{1-z_i\bar{z}_j}\right) \geq 0$.
$(\Leftarrow)$: Assume $\left(\dfrac{1-\lambda_i\bar{\lambda}_j}{1-z_i\bar{z}_j}\right) \geq 0$. We will do by induction on n:
Case for $n = 1$: We have $\dfrac{1-|\lambda_1|^2}{1-|z_1|^2} \geq 0 \Rightarrow |\lambda_1| \leq 1$. Set $f(z) = \lambda_1$.
Assume true for n.
Then, for given $z_0, z_1, z_2, \ldots, z_n \in \mathbb{D}, \quad \lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{D}$

such that $\left(\dfrac{1-\lambda_i\bar{\lambda}_j}{1-z_i\bar{z}_j}\right)_{(n+1)\times(n+1)} \geq 0,$
apply lemma where $\alpha = \lambda_0, \quad \beta = z_0, \quad \mu_i = \varphi_\alpha(\lambda_i), \quad \zeta_i = \varphi_\beta(z_i) \quad 0 \leq i \leq n.$
Note that $\mu_0 = \varphi_{\lambda_0}(\lambda_0), \quad \zeta_0 = \varphi_{z_0}(z_0)$. Then,

$$0 \leq \left(\frac{1-\mu_i\bar{\mu}_j}{1-z_i\bar{z}_j}\right) = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \left(\dfrac{1-\mu_i\bar{\mu}_j}{1-z_i\bar{z}_j}\right)_{n\times n} & \\ 1 & & \end{pmatrix}. \text{ Use Cholesky:}$$

$$\Rightarrow 0 \leq \left(\frac{1-\mu_i\bar{\mu}_j}{1-z_i\bar{z}_j} - 1\right) = \left(\frac{z_i\bar{z}_j - \mu_i\bar{\mu}_j}{1-z_i\bar{z}_j}\right)_{n\times n}.$$

$$\text{Conjugate by } D = \begin{pmatrix} \frac{1}{z_1} & & \\ & \ddots & \\ & & \frac{1}{z_n} \end{pmatrix} \Rightarrow 0 \leq \left(\frac{1-\left(\frac{\mu_i}{z_i}\right)\overline{\left(\frac{\mu_j}{z_j}\right)}}{1-z_i\bar{z}_j}\right)_{n\times n}.$$

$\therefore \exists h \in \mathbb{H}^\infty(\mathbb{D})$ such that $\|h\|_\infty \leq 1$ and $h(\zeta_i) = {\zeta_i}^{-1} \cdot \mu_i = \frac{\mu_i}{\zeta_i}, 1 \leq i \leq n.$

Let $g(z) = z \cdot h(z), \quad g \in \mathbb{H}^\infty(\mathbb{D}), \quad \|g\|_\infty = \|h\|_\infty \leq 1.$

Then $g(\zeta_i) = \mu_i, \quad 1 \leq i \leq n, \quad g(\zeta_0) = g(0) = 0 = \mu_0.$

Let $f = \varphi_{-\alpha} \circ g \circ \varphi_\beta$,
then $f(z_i) = \varphi_{-\alpha}(g(\varphi_\beta(z_i))) = \varphi_{-\alpha}(g(\zeta_i)) = \varphi_{-\alpha}(\mu_i) = \lambda_i, \quad 0 \leq i \leq n.$

$\square$

There are deeper results achieved from D. Marshall, who showed that given $z_1, z_2, \ldots, z_n \in \mathbb{D}, \quad \lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{D}, \quad \exists a_1, a_2, \ldots, a_n \in \mathbb{D}$ and $|c| \leq 1$ such that

$$f(z) = c\prod_{i=1}^{n} \varphi_{a_i}(z),$$

which is called the **Blaschke Product**, $\|f\|_\infty \leq 1, \quad f(z_i) = \lambda_i, 1 \leq i \leq n.$

Now, let's consider 2-Point Pick's theorem:
Let $z_1, z_2, \lambda_1, \lambda_2 \in \mathbb{D}, \quad \exists f(z_i) = \lambda_i \iff \left(\dfrac{1-\lambda_i\bar{\lambda}_j}{1-z_i\bar{z}_j}\right) \geq 0$

$$\iff \left(\frac{1-\varphi_{\lambda_1}(\lambda_i)\overline{\varphi_{\lambda_1}(\lambda_j)}}{1-\varphi_{z_1}(z_i)\overline{\varphi_{z_1}(z_j)}}\right) \geq 0 \iff \alpha = \lambda_1, \quad \beta = z_1.$$

$$\begin{pmatrix} 1 & 1 \\ 1 & \frac{1-|\varphi_{\lambda_1}(\lambda_2)|^2}{1-|\varphi_{z_1}(z_2)|^2} \end{pmatrix} \geq 0 \iff \frac{|\varphi_{z_1}(z_2)|^2 - |\varphi_{\lambda_1}(\lambda_2)|^2}{1-|\varphi_{z_1}(z_2)|^2} \geq 0$$

$$\iff |\varphi_{\lambda_1}(\lambda_2)|^2 \leq |\varphi_{z_1}(z_2)|^2.$$

Hence,

$$\{\exists f, f(z_i) = \lambda_i, i = 1,2 \text{ with} \|f\|_\infty \leq 1\} \iff |\varphi_{\lambda_1}(\lambda_2)| \leq |\varphi_{z_1}(z_2)|.$$

Definition 9.4. $d_{\mathcal{H}}(z_1, z_2) = |\varphi_{z_2}(z_1)|$ *is called pseudo-hyperbolic metric.*

Theorem 9.5 (A generalization of Schwartz's Lemma by Pick)**.** *If* $f : \mathbb{D} \to \mathbb{D}$*, then* $d_{\mathcal{H}}(f(z_1), f(z_2)) \leq d_{\mathcal{H}}(z_1, z_2)$.

Analogues of Pick's theorem

Abrahamse: Let $G \subset \mathbb{C}$ be a "nice g-holed domain", i.e. G is a bounded open region where δG consists of $g+1$ disjoint smooth curves. Denote $\mathbb{T}^g = \{\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_g) : |\lambda_i| = 1, i = 1, \ldots, g\}$.
Let $\{K_\lambda(z, w)\}$ be a family of kernels on G satisfying:

1) $\forall \lambda, K_\lambda(z, w)$ is analytic in z, co-analytic in w,
2) $\forall$ fixed z, w, the map $\lambda \to K_\lambda(z, w)$ is continuous,
3) Given $z_1, \ldots, z_n \in G, \mu_1, \ldots, \mu_n \in \mathbb{D}, \exists f \in \mathbb{H}^\infty(G), f(z_i) = \mu_i, \forall i$ and $\|f\|_\infty \leq 1 \iff ((1 - \mu_i \bar{\mu}_j) K_\lambda(z_i, z_j)) \geq 0, \forall \lambda \in \mathbb{T}^g$.

Example: Fix $0 < q < 1$, let $\mathbb{A}_q = \{z : q < |z| < 1\}$ and
$K_t(z, w) = \sum_{-\infty}^{\infty} \frac{(z\bar{w})^n}{1 + q^{2n+1} \cdot q^{-2t}}$. Then, we have

$$\begin{aligned} K_{t+1}(z, w) &= \sum_{-\infty}^{\infty} \frac{(z\bar{w})^n}{1 + q^{2n+1} \cdot q^{-2(t+1)}} = \sum_{-\infty}^{\infty} \frac{(z\bar{w})^n}{1 + q^{2(n-1)+1} \cdot q^{-2t}} \\ &= \sum_{m=-\infty}^{\infty} \frac{(z\bar{w})^{m+1}}{1 + q^{2m+1} \cdot q^{-2t}} = z\bar{w} K_t(z, w) = z K_t(z, w) \bar{w}. \end{aligned}$$

Note: $((1 - \mu_i \bar{\mu}_j) K_t(z_i, z_j)) \geq 0 \iff ((1 - \mu_i \bar{\mu}_j) z_i K_t(z_i, z_j) \bar{z}_j) \geq 0 \iff ((1 - \mu_i \bar{\mu}_j) K_{t+1}(z_i, z_j)) \geq 0$.

Theorem 9.6 (Abrahamse, 1979)**.** *Given* $z_1, \ldots, z_n \in \mathbb{A}_q, \mu_1, \ldots, \mu_n \in \mathbb{D}, \exists f \in \mathbb{H}^\infty(\mathbb{A}_q)$ *such that* $f(z_i) = \mu_i$, $\|f\|_\infty \leq 1 \iff ((1 - \mu_i \bar{\mu}_j) K_t(z_i, z_j)) \geq 0, \forall 0 \leq t \leq 1$.

Theorem 9.7 (Fedorov-Visinikov / McCullugh, 1990)**.** *Given* $z_1, \ldots, z_n \in \mathbb{A}_q, \mu_1, \ldots, \mu_n \in \mathbb{D}, \exists f \in \mathbb{H}^\infty(\mathbb{A}_q)$ *such that* $f(z_i) = \mu_i$, $\|f\|_\infty \leq 1 \iff \exists t_1, t_2$ *such that* $((1 - \mu_i \bar{\mu}_j) K_{t_k}(z_i, z_j)) \geq 0, k = 1, 2$.

Open Problem: Let G be any $g-$holed domain, fix $z_1, \dots, z_n \in G$. Does there exist a finite subset $\Lambda_z \subseteq \mathbb{T}^g$, enough to check Abrahamse for $\lambda \in \Lambda_z$???

Theorem 9.8 (Agler)**.** *Let* $z_i \in \mathbb{D}^2$, $z_i = (\alpha_i, \beta_i)$, $\mu_1, \dots, \mu_n \in \mathbb{D}$ *then* $\exists f \in \mathbb{H}^\infty(\mathbb{D}^2)$ *such that* $f(z_i) = f((\alpha_i, \beta_i)) = \mu_i, \forall i$, $\|f\|_\infty \leq 1$
$\iff \exists P, Q \in M_n(\mathbb{C}), P = (p_{ij}) \geq 0, Q = (q_{ij}) \geq 0$
such that $(1 - \mu_i \bar{\mu}_j) = ((1 - \alpha_i \bar{\alpha}_j) p_{ij}) + \big((1 - \beta_i \bar{\beta}_j) q_{ij}\big)$.

Theorem 9.9 (Pick's version)**.** $0 \leq \left(\dfrac{1 - \mu_i \bar{\mu}_j}{1 - z_i \bar{z}_j} \right) = P = (p_{ij}) \iff (1 - \mu_i \bar{\mu}_j) = ((1 - z_i \bar{z}_j) p_{ij})$.

Theorem 9.10 (Dual Version)**.** *Let* $z_i = (\alpha_i, \beta_i) \in \mathbb{D}^2$, $\mu_1, \dots, \mu_n \in \mathbb{D}$, *then* $\exists f \in \mathbb{H}^\infty(\mathbb{D}^2)$ *such that* $f(z_i) = f((\alpha_i, \beta_i)) = \mu_i$, $\|f\|_\infty \leq 1 \iff ((1 - \mu_i \bar{\mu}_j) = (r_{ij})) \geq 0, \forall (r_{ij}) \in M_n(\mathbb{C})$ *such that* $((1 - \alpha_i \bar{\alpha}_j) r_{ij}) \geq 0$ *and* $\big((1 - \beta_i \bar{\beta}_j) r_{ij}\big) \geq 0$.

Another Direction of Generalizations

Given $K : X \times X \to \mathbb{C}$ a kernel function, recall if $f \in \mathcal{M}(\mathcal{H}_K)$, then $\|f\|_\infty \leq \|M_f\|$ and $\|M_f\| \leq 1 \iff (1 - f(x)\overline{f(y)})K(x,y)$ is a kernel. If, for given $x_1, \dots, x_n \in X, \mu_1, \dots, \mu_n \in \mathbb{D}, \exists f \in \mathcal{M}(\mathcal{H}_K)$ with $\|M_f\| \leq 1, f(x_i) = \mu_i$, then $((1 - \mu_i \bar{\mu}_j) K(x_i, x_j)) \geq 0$.

Definition 9.11. *K is called a* **Pick kernel**, *if* $\forall n$, $\forall x_1, \dots, x_n \in X$, $\forall \mu_1, \dots, \mu_n \in \mathbb{D}$, $((1 - \mu_i \bar{\mu}_j) K(x_i, x_j)) \geq 0 \Rightarrow \exists f \in \mathcal{M}(\mathcal{H}_K)$ *with* $\|f\|_\infty \leq \|M_f\| \leq 1$.

The above definition is still not very well-understood, but there exists a definition of a **complete Pick Kernel**.

Theorem 9.12 (McCullugh - Quiggen)**.** *The function* $K : X \times X \to \mathbb{C}$ *is a* **complete Pick Kernel** *if and only if there exists a Hilbert space* $\mathcal{L}, \mathcal{B} = ball(\mathcal{L}), \varphi : X \to \mathcal{B}$ *such that* $K(x,y) = \dfrac{1}{1 - \langle \varphi(x), \varphi(y) \rangle}$.

10. Vector-Valued RKHS's

Definition 10.1. *Let $\mathcal{C}$ be a Hilbert space, and X be a set such that the collection of all functions from X to $\mathcal{C}$ is a vector space under pointwise sum. $\mathcal{H}$ is a called a RKHS of $\mathcal{C}$-valued functions provided:*

i) *$\mathcal{H}$ is a Hilbert space,*
ii) *$\mathcal{H}$ is a vector space of $\mathcal{C}$-valued functions ,*
iii) *$\forall y \in X$, the linear map $E_y : \mathcal{H} \to \mathcal{C}, E_y(f) = f(y)$ is bounded.*

Example: Let $\mathcal{H}$ be a RKHS on X. Then $\mathcal{H}^{(n)} = \{ \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} : f_i \in \mathcal{H}\}$ and for any $f \in \mathcal{H}, \|f\|^2 = \|f_1\|^2 + \dots + \|f_n\|^2, \langle f, g \rangle_{\mathcal{H}^{(n)}} = \sum_{i=1}^n \langle f_i, g_i \rangle$. So,

$\mathcal{H}^{(n)}$ is a Hilbert space. Given $f \in \mathcal{H}^{(n)}$, regard f as a function $f : X \to \mathbb{C}^n$ as $f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}$. Clearly, $\mathcal{H}^{(n)}$ satisfies $i), ii)$ of the definition.

Need to show, $\forall y \in X, E_y : \mathcal{H}^{(n)} \to \mathbb{C}^n$ is bounded:

Since $\mathcal{H}$ is a RKHS, then $\exists k_y \in \mathcal{H}$ such that $f_i(y) = \langle f_i, k_y \rangle_{\mathcal{H}}$, which implies

$$f(y) = \begin{pmatrix} \langle f_1, k_y \rangle \\ \vdots \\ \langle f_n, k_y \rangle \end{pmatrix} \Rightarrow \|f(y)\|^2_{\mathbb{C}^n} = \sum_{i=1}^n |f_i(y)|^2 = \sum_{i=1}^n |\langle f_i, k_y \rangle|^2$$

$$. \qquad \leq \sum_{i=1}^n \|f_i\|^2_{\mathcal{H}} \|k_y\|^2_{\mathcal{H}} = \|k_y\|^2_{\mathcal{H}} \cdot \|f\|^2_{\mathcal{H}^{(n)}}$$

$\Rightarrow E_y$ is a bounded linear map, and $\|E_y\| \leq \|k_y\|_{\mathcal{H}}$.

Looking at $f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} \in \mathcal{H}^{(n)}$, we see $\|E_y\| = \|k_y\|$, therefore $\mathcal{H}^{(n)}$ is a RKHS of $\mathbb{C}^n$-valued functions.

Definition 10.2. *Given an RKHS $\mathcal{H}$ of $\mathcal{C}$-valued functions on X, set $K : X \times X \to B(\mathcal{C})$ by $K(x,y) = E_x \circ E_y^* \in B(\mathcal{C})$, then K is called* **operator-valued reproducing kernel** *of $\mathcal{H}$.*

10.1. **Matrices of Operators.** .

Fix $\mathcal{C}$ a Hilbert space, consider $T_{ij} \in B(\mathcal{C}), 1 \leq i \leq m, 1 \leq j \leq n$. Then we regard $T = (T_{ij})_{m \times n} : \mathcal{C}^n \to \mathcal{C}^m$ by $T \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n T_{1j} v_j \\ \vdots \\ \sum_{j=1}^n T_{mj} v_j \end{pmatrix}$ and T is bounded, $\|T\|^2 = \sum_{i,j=1}^{n,m} \|T_{ij}\|^2 \Rightarrow T \in B(\mathcal{C}^n, \mathcal{C}^m)$.

Let $P_{ij} \in B(\mathcal{C}), 1 \leq i,j \leq n, P = (P_{ij}) \in B(\mathcal{C}^n)$. Then $P \geq 0 \iff \langle Pv, v \rangle \geq 0, \forall v \in \mathcal{C}^n$. Having $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, then $\langle Pv, v \rangle_{\mathcal{C}^n} = \sum_{i,j=1}^n \langle P_{ij} v_j, v_i \rangle_{\mathcal{C}}$.

Proposition 10.3. *Let $\mathcal{H}$ be an RKHS of $\mathcal{C}$-valued functions, $K(x,y) = E_x E_y^* \in B(\mathcal{C})$ kernel, let $x_1, \ldots, x_n \in X$. Then $(K(x_i, x_j)) \in B(\mathcal{C}^n)$ is positive (semi-definite).*

Proof. Let $v \in \mathcal{C}^n$, then we have

$$\langle (K(x_i, x_j))v, v \rangle = \sum_{i,j=1}^n \langle K(x_i, x_j) v_j, v_i \rangle = \sum_{i,j=1}^n \langle E_{x_i} E_{x_j}^* v_j, v_i \rangle$$

$$= \sum_{i,j=1}^n \langle E_{x_j}^* v_j, E_{x_i}^* v_i \rangle = \| \sum_{j=1}^n E_{x_j}^* v_j \|^2 \geq 0$$

□

Proposition 10.4. *Let $\mathcal{H}$ be an RKHS of $\mathcal{C}$-valued functions, $K(x,y) = E_x E_y^* \in B(\mathcal{C})$ kernel,et $x_1, \ldots, x_n \in X, v_1, \ldots, v_n \in \mathcal{C}$. Then the function $g(x) = \sum_{i=1}^n K(x,x_i)v_i \in \mathcal{H}$ for any $g : X \to \mathcal{C}$.*

Proof.

$$g(x) = \sum_{i=1}^n K(x,x_i)v_i = \sum_{i=1}^n E_x E_{x_i}^* v_i = E_x \underbrace{(\sum_{i=1}^n E_{x_i}^* v_i)}_{\in \mathcal{H}, E_{x_i}^* : \mathcal{C} \to \mathcal{H}},$$

$$\Rightarrow g = \sum_{i=1}^n E_{x_i}^* v_i \in \mathcal{H}.$$

□

Note: $K(x,y) = E_x E_y^* \Rightarrow K(x,y)^* = (E_x E_y^*)^* = E_y E_x^* = K(y,x)$.
Take $f \in \mathcal{H}, v \in \mathcal{C}$, then $\langle f(x), v\rangle_{\mathcal{C}} = \langle E_x(f), v\rangle = \langle f, E_x^*(v)\rangle_{\mathcal{H}}$.
Let $x_1, x_2, \ldots, x_n \in X$, $v_1, v_2, \ldots, v_n \in \mathcal{C}$, $g = \sum_{j=1}^n E_{x_j}^* v_j \in \mathcal{H}$.
Then $\|g\|^2 = \|\sum_{j=1}^n E_{x_j}^* v_j\|^2 = \langle \sum_{j=1}^n E_{x_j}^* v_j, \sum_{i=1}^n E_{x_i}^* v_i\rangle$
$= \sum_{i,j=1}^n \langle E_{x_i} E_{x_j}^* v_j, v_i\rangle = \langle (K(x_i,x_j)) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}\rangle$.

Proposition 10.5. *Let $x_1, x_2, \ldots, x_n \in X$, $v_1, v_2, \ldots, v_n \in \mathcal{C}$. Then the functions $g = \sum_{j=1}^n E_{x_j}^* v_j$ are dense in $\mathcal{H}$.*

Proof. Let $f \in \mathcal{H}$ be orthogonal to the *span*$\{g : g = \sum_{j=1}^n E_{x_j}^* v_j\}$.
Then we have $\forall x \in X, \forall v \in \mathcal{C}, f \perp E_x^* \iff \langle f, E_x^* v\rangle_{\mathcal{H}} = 0 \iff \langle E_x(f), v\rangle_{\mathcal{C}} = \langle f(x), v\rangle_{\mathcal{C}} = 0, \forall v \in \mathcal{C}$. This implies $f(x) = 0, \forall x \in X$, therefore $f = 0$. So $g = \sum_{j=1}^n E_{x_j}^* v_j$ are dense in $\mathcal{H}$. □

Proposition 10.6. *Let $\mathcal{H}, \mathcal{H}'$ be two RKHS of $\mathcal{C}$-valued functions on X, and $K(x,y), K'(x,y)$ be their kernels respectively. If $K = K'$, then $\mathcal{H} = \mathcal{H}'$ and $\|f\|_{\mathcal{H}} = \|f\|_{\mathcal{H}'}$.*

Proof. By the definition, $g(\cdot) = \sum_{j=1} K(\cdot, x_j)v_j = \sum_{j=1} K'(\cdot, x_j)v_j$, which implies $g \in \mathcal{H} \cap \mathcal{H}'$ and such g's are dense in both spaces. Then $\|g\|_{\mathcal{H}}^2 = \langle (K(x_i,x_j)) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}\rangle = \langle (K'(x_i,x_j)) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}\rangle = \|g\|_{\mathcal{H}'}^2$.

□

Example: Let $\mathcal{H}$ be the ordinary RKHS on X, define $\mathcal{H}^{(n)} = \{\vec{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} : f_i \in \mathcal{H}\}$, and $\|\vec{f}\|_{\mathcal{H}^{(n)}}^2 = \sum_{i=1}^n \|f_i\|_{\mathcal{H}}^2$ which gives RKHS of $\mathbb{C}^n$-valued functions.

Let $K(x,y) \in B(\mathbb{C})$, $K_n(x,y) \in B(\mathbb{C}^n) = M_n$ be the kernels of $\mathcal{H}$, $\mathcal{H}^{(n)}$.
We know $E_y : \mathcal{H}^{(n)} \to \mathbb{C}^n$, $E_y(\vec{(f)}) = \begin{pmatrix} f_1(y) \\ \vdots \\ f_n(y) \end{pmatrix} = \begin{pmatrix} \langle f_1, k_y\rangle \\ \vdots \\ \langle f_n, k_y\rangle \end{pmatrix}$, and
$E_y^* : \mathbb{C}^n \to \mathcal{H}^{(n)}$ with $\langle E_y(f), \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}\rangle_{\mathbb{C}^n} = \langle \begin{pmatrix} f_1(y) \\ \vdots \\ f_n(y) \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}\rangle_{\mathbb{C}^n} =$
$\sum_{j=1}^n \bar{\lambda}_j f_j(y) = \sum_{j=1}^n \bar{\lambda}_j \langle f_j, k_y\rangle_{\mathcal{H}} = \sum_{j=1}^n \langle f_j, \lambda_j k_y\rangle_{\mathcal{H}} = \langle \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}, \begin{pmatrix} \lambda_1 k_y \\ \vdots \\ \lambda_n k_y \end{pmatrix}\rangle_{\mathcal{H}^{(n)}}$.
Therefore $E_y^*(\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}) = \begin{pmatrix} \lambda_1 k_y \\ \vdots \\ \lambda_n k_y \end{pmatrix}$.
Now consider $\langle K_n(x,y) \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}, \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}\rangle_{\mathbb{C}^n} = \langle E_x E_y^* \underbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}}_{\lambda}, \underbrace{\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}}_{\mu}\rangle =$
$\langle E_y^*(\lambda), E_x^*(\mu)\rangle_{\mathcal{H}^{(n)}} = \langle \lambda k_y, \mu k_x\rangle_{\mathcal{H}^{(n)}} = \lambda\bar{\mu}K(x,y) = \sum_{j=1}^n \lambda_j \bar{\mu}_i K(x,y)$.
So, $K_n(x,y) = \begin{bmatrix} K(x,y) & & 0 \\ & \ddots & \\ 0 & & K(x,y) \end{bmatrix}$, i.e. $K_n(x,y) = K(x,y) \cdot I_n$.

Let $\mathcal{H}$ be an RKHS of functions, let $\mathcal{C}$ be some arbitrary Hilbert space. Write $\tilde{E}_y : \mathcal{H} \to \mathbb{C}$ such that $\tilde{E}_y(f) = \langle f, k_y\rangle$. Recall **the Hilbert Space Tensor Product**, then $\mathcal{H} \otimes \mathcal{C} = \{\sum f_i \otimes v_i : f_i \in \mathcal{H}, v_i \in \mathcal{C}\}^-$.
Recall $\mathcal{H}_1, \mathcal{C}_1$ given Hilbert spaces, $T \in B(\mathcal{H}, \mathcal{H}_1)$, $R \in B(\mathcal{C}, \mathcal{C}_1)$, then there exists $T \otimes R : \mathcal{H} \otimes \mathcal{C} \to \mathcal{H}_1 \otimes \mathcal{C}_1$ bounded, and $\|T \otimes R\| = \|T\| \cdot \|R\|$. Consider $\tilde{E}_y : \mathcal{H} \to \mathbb{C}$ and $I_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}$, then there exists a unique bounded operator $\tilde{E}_y \otimes I_{\mathcal{C}} : \mathcal{H} \otimes \mathcal{C} \to \mathbb{C} \otimes \mathcal{C} \cong \mathcal{C}$. So $(\tilde{E}_y \otimes I_{\mathcal{C}})(\sum_i f_i \otimes v_i) = \sum_i \tilde{E}_y(f_i) \otimes v_i = \sum_i f_i(y) v_i$ and $\|\tilde{E}_y \otimes I_{\mathcal{C}}\| = \|\tilde{E}_y\|$. Define $\mathcal{K} = \{(\tilde{E}_y \otimes I)(u) : u \in \mathcal{H} \otimes \mathcal{C}\}$, i.e $\mathcal{K} = \{\hat{u}(y) : u \in \mathcal{H} \otimes \mathcal{C}\}$ is a set of $\mathcal{C}$-valued functions.
Claim: For any $u_1, u_2 \in \mathcal{H} \otimes \mathcal{C}\}$, we have $\hat{u}_1 = \hat{u}_2 \iff u_1 = u_2$:
Pick a basis $\{e_\alpha\}_{\alpha \in A}$ for $\mathcal{C}$. Every $u \in \mathcal{H} \otimes \mathcal{C}\}$ has a unique expansion as $u = \sum f_\alpha \otimes e_\alpha$, where $f_\alpha \in \mathcal{H}$, and $\|u\|^2_{\mathcal{H}\otimes\mathcal{C}} = \sum_\alpha \|f_\alpha\|^2_{\mathcal{H}}$. Therefore $u_1 = \sum f^1_\alpha \otimes e_\alpha$ and $u_2 = \sum f^2_\alpha \otimes e_\alpha$. Then $\hat{u}_1 = \hat{u}_2 \iff \sum f^1_\alpha(y) e_\alpha = \sum f^2_\alpha(y) e_\alpha, \forall y, \iff f^1_\alpha(y) = f^2_\alpha(y) \iff f^1_\alpha = f^2_\alpha, \forall \alpha \in A \iff u_1 = u_2$.
Using the claim, then as vector spaces $\mathcal{K} \cong \mathcal{H} \otimes \mathcal{C}$, since $u_1 \hat{+} u_2 = \hat{u}_1 +$

$\hat{u}_2$, $\alpha\hat{u}_1 = \alpha\hat{u}_1$. If we set $\|\hat{u}\|_{\mathcal{K}} = \|u\|_{\mathcal{H}\otimes\mathcal{C}}$, then $\mathcal{K}$ is a Hilbert space of $\mathcal{C}$-valued functions. For $y \in X, \hat{u}(y) = (\tilde{E}_y \otimes I_{\mathcal{C}})(u)$ and $\|\hat{u}(y)\|_{\mathcal{C}} \leq \|\tilde{E}_y \otimes I_{\mathcal{C}}\| \cdot \|u\|_{\mathcal{H}\otimes\mathcal{C}} = \|\tilde{E}_y\| \cdot \|\hat{u}\|_{\mathcal{K}}$. Therefore, $\mathcal{K}$ is an RKHS of $\mathcal{C}$-valued functions with $E_y(\hat{u}) = \hat{u}(y) = (\tilde{E}_y \otimes I_{\mathcal{C}})(u)$. Hence, $\mathcal{K} \cong \mathcal{H} \otimes \mathcal{C}$ is an RKHS of $\mathcal{C}$-valued functions if $u = \sum_{i=1}^n f_i \otimes v_i$ and $\hat{u}(x) = \sum_{i=1}^n f_i(x)v_i$.

Note: Let $\mathcal{C} = \mathbb{C}^n$, and $\mathcal{H}^{(n)} = \{\vec{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} : f_i \in \mathcal{H}\}$. Then identify

$\vec{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \underbrace{f_1 \otimes e_1 + \cdots + f_n \otimes e_n}_{\mathcal{H}\otimes\mathbb{C}^n} = \sum_{i=1}^n f_i \otimes e_i$.

Proposition 10.7. *Let $\mathcal{H}$ be an RKHS of functions with kernel $K(x,y)$, let $\mathcal{C}$ be any Hilbert space. If we regard $\mathcal{H}\otimes\mathcal{C}$ as an RKHS of $\mathcal{C}$-valued functions as above, then* **the kernel of** *$\mathcal{H}\otimes\mathcal{C}$ is defined as $K_{\mathcal{C}}(x,y) = K(x,y) \cdot I_{\mathcal{C}}$.*

Proof. Look at $E_x : \mathcal{H}\otimes\mathcal{C} \to \mathcal{C}$ and $E_x^* : \mathcal{C} \to \mathcal{H}\otimes\mathcal{C}$. Then,
for any $v, w \in \mathcal{C}$, $f \in \mathcal{H}$, we have $E_x^*(v) = k_x \otimes v$ and
$\langle f \otimes w, E_x^*(v)\rangle_{\mathcal{H}\otimes\mathcal{C}} = \langle E_x(f\otimes w), v\rangle_{\mathcal{C}} = \langle f(x)w, v\rangle_{\mathcal{C}} = f(x)\langle w, v\rangle_{\mathcal{C}}$
$= \langle f, k_x\rangle_{\mathcal{H}}\langle w, v\rangle_{\mathcal{C}} = \langle f\otimes w, k_x \otimes v\rangle_{\mathcal{H}\otimes\mathcal{C}}$.
So $K_{\mathcal{C}}(x,y) = E_x E_y^* \in B(\mathcal{C})$, therefore $\langle K_{\mathcal{C}}(x,y)v, w\rangle_{\mathcal{C}} = \langle E_y^* v, E_x^* w\rangle_{\mathcal{H}\otimes\mathcal{C}}$
$= \langle k_y \otimes v, k_x \otimes w\rangle = \langle k_y, k_x\rangle_{\mathcal{H}}\langle v, w\rangle_{\mathcal{C}} = \langle K(x,y)\cdot I_{\mathcal{C}} v, w\rangle$.
Hence, $K_{\mathcal{C}}(x,y) = K(x,y) \cdot I_{\mathcal{C}}$. □

Corollary 10.8.

$$\mathcal{H}\otimes\mathbb{C}^n = \{\vec{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} : f_i \in \mathcal{H}, \|\vec{f}\|^2 = \sum_{i=1}^n \|f_i\|^2\}.$$

Proof. Earlier, we showed that the right hand side of the above is a $\mathbb{C}^n$-valued RKHS with kernel $k_y(x)I_{\mathbb{C}^n}$, which is a kernel for the left hand side. □

Recall from the scalar case the bounded linear functionals $f \to f(x)$, with $|f(x)| = |\langle f, k_x\rangle| \leq \|f\| \cdot \|k_x\| \Rightarrow |f(x)| \leq \|f\|K(x,x)^{1/2}$.

Proposition 10.9. *Let $\mathcal{H}$ be an RKHS of $\mathcal{C}$-valued functions. Then $\forall f \in \mathcal{H}$, $\|f(x)\|_{\mathcal{C}} \leq \|f\| \cdot \|K(x,x)^{1/2}\|$.*

Proof. Note that $\|K(x,x)^{1/2}\|^2 = \|(K(x,x)^{1/2})^* K(x,x)^{1/2}\| = \|K(x,x)\|$.

$$\|f(x)\|_{\mathcal{C}}^2 = \langle f(x), f(x)\rangle_{\mathcal{C}} = \langle E_x(f), E_x(f)\rangle = \langle E_x^* E_x(f), f\rangle \leq \|E_x^* E_x\| \cdot \|f\|^2$$
$$= \|E_x E_x^*\| \cdot \|f\|^2 = \|K(x,x)\| \cdot \|f\|^2 \Rightarrow \|f(x)\|_{\mathcal{C}} \leq \|f\| \cdot \|K(x,x)^{1/2}\|.$$

□

Definition 10.10. *The map $K : X \times X \to B(\mathcal{C})$ is* **positive***, provided $\forall x_1, \dots, x_n \in X, \forall n$, the matrix $(K(x_i, x_j)) \in M_n(B(\mathcal{C})) = B(\mathcal{C}^n)$ is positive-definite, i.e. $(K(x_i, x_j)) \geq 0$.*

Theorem 10.11 (Extension of Moore)**.** *Let $K : X \times X \to B(\mathcal{C})$ be a positive function. Then there exists a unique RKHS of $\mathcal{C}$-valued functions, $\mathcal{H}$, such that K is the kernel function for $\mathcal{H}$.*

Proof. Let $W = \{g(\cdot) = \sum_{j=1}^n K(\cdot, x_j)v_j : \forall x_j \in X, \forall v_j \in \mathcal{C}\}$ be the set of those $\mathcal{C}$-valued functions. Define $B : W \times W \to \mathbb{C}$ by $B((\sum_{j=1}^n K(\cdot, x_j)v_j, \sum_{i=1}^n K(\cdot, x_i)w_i)) = \sum_{i,j=1}^n \langle K(x_i, x_j)v_j, w_i\rangle$.
B is well-defined: Suppose $f(\cdot) = \sum_{j=1}^n K(\cdot, x_j)v_j$ and $f(x) = 0, \forall x$. Want to show $B(f, h) = 0, B(h, f) = 0$. It is enough to take $h(\cdot) = K(\cdot, y)w$ by linearity. Then $B(f, h) = \sum_{j=1}^n \langle K(y, x_j)v_j, w\rangle = \langle f(y), w\rangle = 0$. Similarly, $B(h, f) = 0$, therefore B is well-defined.
B **is an inner-product on** W**:** Let $f(\cdot) = K(\cdot, x)v$, and $h(\cdot) = K(\cdot, y)w$.

- Clearly, we have $B(f_1 + f_2, h) = B(f_1, h) + B(f_2, h)$,
  and $B(f, h_1 + h_2) = B(f, h_1) + B(f, h_2)$,
  and $B(\lambda f, h) = \lambda B(f, h)$.
- $B(f, h) = \langle K(y, x)v, w\rangle_{\mathcal{C}} = \langle v, K(y, x)^* w\rangle$
  $= \langle v, K(x, y)w\rangle_{\mathcal{C}} = \overline{\langle K(x, y)w, v\rangle} = \overline{B(h, f)}$.
- $f(\cdot) = \sum_{j=1}^n K(\cdot, x_j)v_j \Rightarrow B(f, f) = \sum_{i,j=1}^n \langle K(x_i, x_j)v_j, v_i\rangle \geq 0$
  $B(f, f) = 0 \iff \sum_{i,j=1}^n \langle K(x_i, x_j)v_j, v_i\rangle = 0$.
  Consider $B(f + \lambda th, f + \lambda th) = B(f, f) + \bar{\lambda} tB(f, h) + \lambda tB(h, f) + |\lambda|^2 t^2 B(h, h) \geq 0$, then $\forall \lambda, t$, pick λ such that $-2t|B(f, h)| + t^2 B(h, h) \geq 0, \forall t$,
  $\Rightarrow -|B(f, h)| \geq 0 \Rightarrow B(f, h) = 0, \forall h$.
  Let $h(\cdot) = K(\cdot, y)w$, then $0 = B(f, h) = \sum_{j=1}^n \langle K(y, x_j)v_j, w\rangle_{\mathcal{C}} = \langle f(y), w\rangle = 0, \forall y, w$. Therefore $f = 0$.

Hence, B is an inner product on W.
Now, let $\mathcal{H}$ be the Hilbert space obtained by completing W. Then this $\mathcal{H}$ is a space of $\mathcal{C}$-valued functions: Idea of showing it, consists of taking $\{f_n\} \subseteq W$, a Cauchy sequence. Then $\forall x, \{f_n(x)\}$ is Cauchy in $\mathcal{C}$, which implies $\exists f : X \to \mathcal{C}$ defined by $f(x) = \lim_{n\to\infty} f_n(x)$. And the rest is similar... ☐

Theorem 10.12 (Sums of Kernels)**.** *Let $\mathcal{H}_1, \mathcal{H}_2$ be two RKHS's of $\mathcal{C}$-valued functions on X with kernels K_1, K_2 respectively. Then $\mathcal{H} = \{f = f_1 + f_2 : f_i \in \mathcal{H}_i, i = 1, 2\}$ with norm $\|f\|^2_{\mathcal{H}} = \|f_1 + f_2\|^2_{\mathcal{H}} = \inf\{\|f_1\|^2_{\mathcal{H}_1} + \|f_2\|^2_{\mathcal{H}_2}\}$ is an RKHS of $\mathcal{C}$-valued functions on X with kernel $K(x, y) = K_1(x, y) + K_2(x, y)$.*

Proof. Look at $\mathcal{K} = \mathcal{H}_1 \oplus \mathcal{H}_2 = \{(f_1, f_2) : \|(f_1, f_2)\|^2 = \|f_1\|^2 + \|f_2\|^2\}$. Define $\Gamma : \mathcal{K} \to \mathcal{H}$ by $\Gamma((f_1, f_2)) = f_1 + f_2$, which is **onto**. Let $\mathcal{N} = Ker\Gamma$.
$\mathcal{N}$ **is closed:** Let $\{(f_1^n, f_2^n)\} \subseteq \mathcal{N}$ be a sequence such that $(f_1^n, f_2^n) \to (f_1, f_2)$. Then $f_1(x) = \lim_n f_1^n(x) = \lim_n(-f_2^n(x)) = -f_2(x) \Rightarrow f_1(x) + f_2(x) = 0 \forall x$, i.e. $f_1 + f_2 = 0$. Hence, $(f_1, f_2) \in \mathcal{N} \Rightarrow \mathcal{N}$ is closed.

Therefore, $\Gamma : \mathcal{N}^{\perp} \to \mathcal{H}$ is an **one-to-one** map. So, Γ is an isometry on $\mathcal{N}^{\perp}$, and $\mathcal{H}$ with this norm is a Hilbert space.
Now, let $f \in \mathcal{H}$, $f = f_1 + f_2$, where $(f_1, f_2) \in \mathcal{N}^{\perp}$. Then

$$\begin{aligned}\|f(x)\| &\leq \|f_1(x)\| + \|f_2(x)\| \leq \|E_x^1\| \cdot \|f_1\|_{\mathcal{H}_1} + \|E_x^2\| \cdot \|f_2\|_{\mathcal{H}_2} \\ &\leq (\|E_x^1\|^2 + \|E_x^2\|^2)^{1/2}(\|f_1\|^2_{\mathcal{H}_1} + \|f_2\|^2_{\mathcal{H}_2})^{1/2} \\ &\Rightarrow \|f(x)\| \leq \sqrt{\|E_x^1\|^2 + \|E_x^2\|^2} \cdot \|f\|_{\mathcal{H}} \Rightarrow \mathcal{H} \text{ is an RKHS.}\end{aligned}$$

Let $E_x : \mathcal{H} \to \mathcal{C}$ and $K(x,y) = E_x E_y^*$. To show that $K(x,y) = K_1(x,y) + K_2(x,y)$, it is enough to show that for any $v \in \mathcal{C}$, $E_y^* v = {E^1}_y^* v + {E^2}_y^* v$, i.e. $E_y^* v = K_1(\cdot, y)v + K_2(\cdot, y)v$. Since $K(x,y)v = E_x(E_y^* v) = K_1(x,y)v + K_2(x,y)v$, then done.
Claim: $(K_1(\cdot,y)v, K_2(\cdot,y)v) \in \mathcal{N}^{\perp}$: Let $(f_1, f_2) \in \mathcal{N}$, then

$$\begin{aligned}\langle (f_1,f_2), (K_1(\cdot,y)v, K_2(\cdot,y)v)\rangle_{\mathcal{H}_1 \oplus \mathcal{H}_2} &= \langle f_1, {E^1}_y^* v\rangle_{\mathcal{H}_1} + \langle f_2, {E^2}_y^* v\rangle_{\mathcal{H}_2} \\ &= \langle E_y^1(f_1), v\rangle_{\mathcal{C}} + \langle E_y^2(f_2), v\rangle_{\mathcal{C}} = \langle f_1(y) + f_2(y), v\rangle_{\mathcal{C}} = 0\end{aligned}$$

Let $f_1 + f_2 \in \mathcal{H}$, where $(f_1, f_2) \in \mathcal{N}^{\perp}$, then

$$\begin{aligned}\langle (f_1 + f_2), ({E^1}_y^* v + {E^2}_y^* v)\rangle_{\mathcal{H}} &= \langle (f_1, f_2), ({E^1}_y^* v, {E^2}_y^* v)\rangle_{\mathcal{H}_1 \oplus \mathcal{H}_2} \\ &= \langle f_1, {E^1}_y^* v\rangle_{\mathcal{H}_1} + \langle f_2, {E^2}_y^* v\rangle_{\mathcal{H}_2} = \langle f_1(y), v\rangle_{\mathcal{C}} + \langle f_2(y), v\rangle_{\mathcal{C}} \\ &= \langle f_1(y) + f_2(y), v\rangle_{\mathcal{C}} = \langle E_y(f_1 + f_2), v\rangle_{\mathcal{C}} = \langle f_1 + f_2, E_y^* v\rangle_{\mathcal{H}}.\end{aligned}$$

Therefore, $E_y^* v = {E^1}_y^* v + {E^2}_y^* v$. □

Theorem 10.13 (Douglas' Factorization Theorem). *Let $\mathcal{H}_1.\mathcal{H}_2, \mathcal{K}$ be three given Hilbert spaces, $B \in B(\mathcal{H}_1, \mathcal{K})$, $F \in B(\mathcal{H}_2, \mathcal{K})$. Then t.f.a.e. :*

1) *$Range(F) \subseteq Range(B)$,*
2) *There exists $m > 0$, such that $FF^* \leq m^2 BB^*$,*
3) *There exists $X \in B(\mathcal{H}_2, \mathcal{H}_1)$, such that $F = BX$.*

Proof. (3) $\Rightarrow$ (1) : It is obvious.
(1) $\Rightarrow$ (3) : Let $\mathcal{N}_1 = Ker(B)$, then $B : \mathcal{N}_1^{\perp} \to Range(B)$ is one-to-one map. For each $h_2 \in \mathcal{H}_2$, there exists a unique $h_1 \in clN_1^{\perp}$ such that $F(h_2) = B(h_1)$, since $Range(F) \subseteq Range(B)$. Define $X(h_2) = h_1$.

- X is linear: By the above, for any $h_2, h_2' \in \mathcal{H}_2$, we can have $F(h_2) = B(h_1), F(h_2') = B(h_1')$, which implies $F(h_2 + h_2') = B(h_1 + h_1')$. Therefore $X(h_2 + h_2') = h_1 + h_1' = X(h_2) + X(h_2')$. Similarly, we can show $X(\alpha h_2) = \alpha X(h_2)$.
- X is bounded: Let $h_2^n \to h_2$, and $X(h_2^n) \to h_1$. Need to show $X(h_2) = h_1$. Let $h_1^n = X(h_2^n)$, then $F(h_2^n) = B(h_1^n)$, where $F(h_2^n) \to F(h_2)$ and $B(h_1^n) \to B(h_1)$, therefore $F(h_2) = B(h_1)$. Hence, by Closed Graph theorem, we have $X(h_2) = h_1$.

(3) $\Rightarrow$ (2) : Let $F = BX$, let $\|X\| = m$.
Then $XX^* \leq m^2 I \Rightarrow B(XX^*)B^* \leq B(m^2 I)B^* \Rightarrow FF^* \leq m^2 BB^*$.

$(2) \Rightarrow (3)$: Having $FF^* \leq m^2 BB^*$, then $\|F^*k\|^2 = \langle FF^*k, k\rangle \leq m^2\langle BB^*k, k\rangle = m^2\|B^*k\|^2$. Define $Y : \underbrace{Range(B^*)}_{\subseteq \mathcal{H}_1} \to \underbrace{Range(F^*)}_{\subseteq \mathcal{H}_2}$ by $Y(B^*k) = F^*k$.

- Y is well-defined: Suppose $B^*k_1 = B^*k_2 \Rightarrow B^*(k_1 - k_2) = 0 \Rightarrow F^*(k_1 - k_2) = 0 \Rightarrow F^*k_1 = F^*k_2$.
- Y is linear: It is obvious by the definition.
- Y is bounded by m: $\|Y(B^*k)\| = \|F^*k\| \leq m\|B^*k\|$.

We extend Y to $\tilde{Y}$ by continuity as $\tilde{Y} : \overline{Range(B^*)} \to \mathcal{H}_2$.
Let $P : \mathcal{H}_1 \to \overline{Range(B^*)}$, then $\tilde{Y}P = Z : \mathcal{H}_1 \to \mathcal{H}_2$ and Z is bounded by m, and $ZB^*k = YB^*k = F^*k$, which implies $ZB^* = F^* \Rightarrow F = BZ^*$. Let $X = Z^*$. Done. □

Lemma 10.14 (1). *Let $\mathcal{H}$ be an RKHS of $\mathcal{C}$-valued functions on X with kernel $K(x, y)$, let $v_1, \ldots, v_n \in \mathcal{C}$. Then $\exists\, g \in \mathcal{H}$ such that $g(x_i) = v_i, \forall i$ if and only if* $\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in Range((K(x_i, x_j))^{1/2})$.

Proof. Define $B \in B(\mathcal{H}, \mathcal{C}^n)$ by $B(f) = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} E_{x_1}(f) \\ \vdots \\ E_{x_n}(f) \end{bmatrix}$. Consider

$$\langle B(f), \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}\rangle_{\mathcal{C}^n} = \langle \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}, \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}\rangle_{\mathcal{C}^n} = \sum_{i=1}^n \langle f(x_i), v_i\rangle = \langle f, \sum_{i=1}^n E^*_{x_i} v_i\rangle.$$

Hence, $B^* : \mathcal{C}^n \to \mathcal{H}$ can be defined as $B^* = \begin{bmatrix} E^*_{x_1} \\ \vdots \\ E^*_{x_n} \end{bmatrix}$, and we have

$$B^*(\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}) = \sum_{i=1}^n E^*_{x_i} v_i = \sum_{i=1}^n K(\cdot, x_i) v_i.$$

Next, $BB^* : \mathcal{C}^n \to \mathcal{C}^n$ and $BB^* = (E^*_{x_i} E_{x_j}) = K(x_i, x_j) = ((K(x_i, x_j))^{1/2})^2$, which implies $Range((K(x_i, x_j))^{1/2}) = Range(B)$, and
$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in Range(B) \iff \exists\, f \in \mathcal{H}$ such that $\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}$. □

Lemma 10.15 (2). *Let $\mathcal{H}$ be an RKHS of $\mathcal{C}$-valued functions on X with kernel $K(x, y)$, let $f \in \mathcal{H}$, and for each finite subset $F \subseteq X$, set $g_F = P_F(f)$, where P_F is the orthogonal projection onto $\overline{span\{K(\cdot, x)v : x \in F, v \in \mathcal{C}\}}$. Then, the net $\{g_F : F\ finite\}$ converges in norm to f.*

Sketch of the proof: 1) Show $g_F(x) = f(x), \forall x \in F$.
2) Rest is the same as in scalar case(done before).

Definition 10.16. *Let $v, w \in \mathcal{C}$, let $R_{v,w} = v \otimes \bar{w}$ denote the operator defined by $R_{v,w}(z) = \langle z, w \rangle v$.*

Some important properties about $R_{v,w}$:

(1) $R_{v,w}$ is rank-1 operator.
(2) Let $R : \mathcal{C} \to \mathcal{C}$ be some rank-1 operator, choose any $v \in Range(R), v \neq 0$. Then $R(z) = \lambda(z)v$, where λ is some scalar function. By linearity of R, we have $\lambda(z_1 + z_2)v = R(z_1 + z_2) = R(z_1) + R(z_2) = \lambda(z_1)v + \lambda(z_2)v$, i.e. $\lambda(z_1 + z_2) = \lambda(z_1) + \lambda(z_2)$. Similarly, $\lambda(\alpha z) = \alpha\lambda(z)$. Hence, λ is linear.
Look at $|\lambda(z)| \cdot \|v\| = \|R(z)\| \leq \|R\| \cdot \|z\| \Rightarrow |\lambda(z)| \leq (\frac{\|R\|}{\|v\|})\|z\|$, which implies λ is a bounded linear functional. Therefore, $\exists\, w \in \mathcal{C}$ such that $\lambda(z) = \langle z, w \rangle \Rightarrow R(z) = \langle z, w \rangle v$, i.e. $R = R_{v,w}$.
(3) Similarly, if $R : \mathcal{C} \to \mathcal{C}$ is of rank-n, choose an orthonormal basis $\{v_1, \ldots, v_n\}$ for $Range(R)$, and get $R = \sum_{i=1}^n R_{v_i,w_i}$.
(4) $R_{\alpha v,w} = (\alpha v) \otimes \bar{w} = \alpha(v \otimes \bar{w}) = \alpha R_{v,w}$,
and $R_{v,\alpha w} = v \otimes \overline{(\alpha w)} = \bar{\alpha}(v \otimes \bar{w}) = \bar{\alpha} R_{v,w}$.
(5) Let $\mathcal{C} = \mathbb{C}^n$, let $v = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$, and $w = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$, then $R_{v,w} = \begin{bmatrix} \alpha_1\bar{\beta}_1 & \cdots & \alpha_1\bar{\beta}_n \\ \vdots & & \vdots \\ \alpha_n\bar{\beta}_1 & \cdots & \alpha_n\bar{\beta}_n \end{bmatrix}$,
i.e. $R_{v,w} = (\alpha_i\bar{\beta}_j)$.

Theorem 10.17 (3)**.** *Let $\mathcal{H}$ be an RKHS of $\mathcal{C}$-valued functions on X with kernel $K(x, y)$, let $f : X \to \mathcal{C}$. Then, t.f.a.e.:*

i) *$f \in \mathcal{H}$,*
ii) *$\exists\, m \geq 0$, such that $\forall x_1, \ldots, x_n \in X, \quad \exists\, h \in \mathcal{H}$ with $\|h\| \leq m$ and $h(x_i) = f(x_i), \forall 1 \leq i \leq m$.*
iii) *$\exists\, m \geq 0$, such that $f(x) \otimes \overline{f(y)} \leq m^2 K(x, y)$, i.e. $m^2(K(x_i, x_j)) - (f(x_i) \otimes \overline{f(x_j)}) \geq 0$.*

Moreover, the least such m that works is $m = \|f\|$.

Proof. $(i) \Rightarrow (iii)$: Fix $x_1, \ldots, x_n \in X$, $v_1, \ldots, v_n \in \mathcal{C}$,
let $g(\cdot) = \sum_{j=1}^n K(\cdot, x_j)v_j = \sum_{j=1}^n E_{x_j}^* v_j$. Then

$$\|g\|^2 = \sum_{i,j=1}^n \langle E_{x_j}^* v_j, E_{x_i}^* v_i \rangle = \sum_{i,j=1}^n \langle K(x_i, x_j)v_j, v_i \rangle = \langle (K(x_i, x_j)) \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \rangle.$$

Therefore, letting $m = \|f\|$, we have

$$m^2 \sum_{i,j=1}^{n} \langle K(x_i, x_j)v_j, v_i\rangle - \sum_{i,j=1}^{n} \langle f(x_i) \otimes \overline{f(x_j)} v_j, v_i\rangle$$

$$= m^2 \cdot \|g\|^2 - \sum_{i,j=1}^{n} \langle\langle v_j, f(x_j)\rangle f(x_i), v_i\rangle = m^2 \cdot \|g\|^2 - \sum_{i,j=1}^{n} \langle v_j, f(x_j)\rangle \overline{\langle f(x_i), v_i\rangle}$$

$$= m^2 \cdot \|g\|^2 - |\sum_{j=1}^{n} \langle v_j, f(x_j)\rangle|^2 = \|f\|^2 \cdot \|g\|^2 - |\sum_{j=1}^{n} \langle E_{x_j}^* v_j, f\rangle|^2$$

$$= \|f\|^2 \cdot \|g\|^2 - |\langle g, f\rangle|^2 \geq 0 \text{ by Cauchy-Schwarz.}$$

$(iii) \Rightarrow (ii)$: Fix $x_1, \dots, x_n \in X$, $v_1, \dots, v_n \in \mathcal{C}$. Define $T \in B(\mathbb{C}, \mathcal{C}^n)$ by $T(\lambda) = \lambda \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$, then $T^* : \mathcal{C}^n \to \mathbb{C}$, $\quad T^*(\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}) = \sum_{j=1}^n \langle z_j, v_j\rangle$. Substitute $v_1, \dots, v_n \in \mathcal{C}$ with $f(x_1), \dots, f(x_n) \in \mathcal{C}$, then $T^*(\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}) = \sum_{j=1}^n \langle z_j, f(x_j)\rangle$.
Look at $TT^* : \mathcal{C}^n \to \mathcal{C}^n$, then

$$\langle (TT^*) \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}\rangle = \langle T^* \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, T^* \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}\rangle$$

$$= \langle \sum_{j=1}^{n} \langle z_j, f(x_j)\rangle, \sum_{i=1}^{n} \langle w_i, f(x_i)\rangle\rangle_{\mathbb{C}} = \sum_{i,j=1}^{n} \langle z_j, f(x_j)\rangle \overline{\langle w_i, f(x_i)\rangle}$$

$$= \sum_{i,j=1}^{n} \langle z_j, f(x_j)\rangle \langle f(x_i), w_i\rangle = \langle (f(x_i) \otimes \overline{f(x_j)}) \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}\rangle.$$

Hence, $TT^* = (f(x_i) \otimes \overline{f(x_j)})$.
Let $B \in B(\mathcal{H}, \mathcal{C}^n)$, $B(g) = \begin{bmatrix} g(x_1) \\ \vdots \\ g(x_n) \end{bmatrix}$, then $BB^* = (K(x_i, x_j))$ and $m^2 BB^* \geq TT^*$ by (iii), which implies $\exists\, X : \mathbb{C} \to \mathcal{H}$ such that $T = BX$, let $h = X(1)$. Then $\|h\| \leq \|X\| \cdot 1 \leq m$. Therefore,

$$\begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} = T(1) = BX(1) = B(h) = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_n) \end{bmatrix}.$$

$(ii) \Rightarrow (i)$: Let $\mathcal{F}$ be the set of all finite subsets of X. Then $\forall F \in \mathcal{F}$, $\exists\, h_F \in \mathcal{H}$ such that $h_F(x) = f(x)$, $x \in F$, $\|h_F\| \leq m$. Look at the projection $P_F : \mathcal{H} \to \mathcal{H}_F = \{g : g(x) = 0, \forall x \in F\}^{\perp}$. Then, for any $x \in F$, $(P_F h_F)(x) = h_F(x) = f(x)$ and $\|P_F h_F\| \leq \|h_F\|$.
WLOG, let $h_F \in \mathcal{H}_F$, **we claim** $\{h_F\}_{F\in\mathcal{F}}$ **is a Cauchy net:**
Given $\epsilon \geq 0$, let $M = \sup\{\|h_F\| : F \in \mathcal{F}\} \leq m$. Pick F_o such that $M^2 - \dfrac{\epsilon^2}{4} \leq \|h_{F_o}\|^2 \leq M^2$. Choose F containing F_o, then $(h_F - h_{F_o}) \in \mathcal{H}_{F_0}^{\perp}$, therefore $\|h_F\|^2 = \|h_{F_o}\|^2 + \|h_F - h_{F_o}\|^2 \leq M^2$ by Pythagorea theorem. Then $\|h_F - h_{F_o}\|^2 \leq \dfrac{\epsilon^2}{4} \Rightarrow \|h_F - h_{F_o}\| \leq \dfrac{\epsilon}{2}$. Therefore, for any F_1, F_2 containing F_o, we have $\|h_{F_1} - h_{F_2}\| \leq \|h_{F_1} - h_{F_o}\| + \|h_{F_2} - h_{F_o}\| < \epsilon \Rightarrow \{h_F\}$ is Cauchy. Hence, $\exists\, h \in \mathcal{H}$ such that $h = \lim_F h_F$. Given $x_o \in X, F_o = \{x_o\}$, then $\forall F$ containing F_o, $x_o \in F$, we have $h_F(x_o) = f(x_o)$.
Therefore, $h(x_o) = \lim_F h_F(x_o) = f(x_o), \forall x_o \in X \Rightarrow h = f \Rightarrow \|f\| \leq m$. □

Proposition 10.18 (1)**.** *Let $\mathcal{H}_1, \mathcal{H}_2$ be two RKHS's of $\mathcal{C}$-valued functions on X with kernels $K_1(x,y), K_2(x,y)$ and norms $\|\cdot\|_1, \|\cdot\|_2$ respectively. If there exists $m \geq 0$ such that $K_2(x,y) = m^2 K_1(x,y), \forall x, y \in X$, then $\mathcal{H}_1 = \mathcal{H}_2$ and $\|f\|_1 = m\|f\|_2$.*

Proof. By property (3), $f \in \mathcal{H}_1 \iff f \in \mathcal{H}_2$. Norms work by "moreover" part, i.e. $\|f\|_1 = m\|f\|_2$. □

Proposition 10.19 (2)**.** *Let $\mathcal{H}_1, \mathcal{H}_2$ be two RKHS's of $\mathcal{C}$-valued functions on X with kernels $K_1(x,y), K_2(x,y)$ respectively. If $\mathcal{H}_1 \subseteq \mathcal{H}_2$ and $I_1 : \mathcal{H}_1 \to \mathcal{H}_2$ the inclusion map, then I_1 a bounded linear map, and $I_1^*(K_2(\cdot,y)v) = K_1(\cdot,y)v$.*

Proof. Suppose $f_n \to f$ in $\mathcal{H}_1$, and $I_1(f) \to g$ in $\mathcal{H}_2$. Need to show $I_1(f) = g$ in $\mathcal{H}_2$: We have $f(x) = \lim_n f_n(x) = \lim_n I_1(f_n)(x) = g(x) \Rightarrow f = g$.
Let $f \in \mathcal{H}_1$, then

$$\begin{aligned}\langle I_1^*(K_2(\cdot,y)v), f\rangle_1 &= \langle K_2(\cdot,y)v, I_1(f)\rangle_2 = \langle {E^2}^*_y(v), f\rangle_2 = \langle v, E_y^2(f)\rangle_{\mathcal{C}} \\ &= \langle v, f(y)\rangle_{\mathcal{C}} = \langle v, E_y^1(f)\rangle_{\mathcal{C}} = \langle {E^1}^*_y(v), f\rangle_1 = \langle K_1(\cdot,y)v, f\rangle_1.\end{aligned}$$

Hence, $I_1^*(K_2(\cdot,y)v) = K_1(\cdot,y)v$. □

Theorem 10.20 (3)**.** *Let $\mathcal{H}_1, \mathcal{H}_2$ be two RKHS's of $\mathcal{C}$-valued functions on X with kernels $K_1(x,y), K_2(x,y)$ respectively. If there exists $m > 0$ such that $K_1(x,y) \leq m^2 K_2(x,y), \forall x, y \in X$, then $\mathcal{H}_1 \subseteq \mathcal{H}_2$ and $\|f\|_2 \leq m\|f\|_1$. Conversely, if $\mathcal{H}_1 \subseteq \mathcal{H}_2$, then $\exists\, m$ such that $K_1(x,y) \leq m^2 K_2(x,y)$ and $m = \|I_1\|$, where $I_1 : \mathcal{H}_1 \to \mathcal{H}_2$ is the inclusion operator.*

Proof. Assume $K_1(x,y) \leq m^2 K_2(x,y)$. Let $f \in \mathcal{H}_1$, then $f(x) \otimes \overline{f(y)} \leq \|f\|_1^2 K_1(x,y) \leq \|f\|_1^2 m^2 K_2(x,y) \Rightarrow f \in \mathcal{H}_2$ and $\|f\|_2 \leq m\|f\|_1$.
Conversely, if $\mathcal{H}_1 \subseteq \mathcal{H}_2$, then $\|I_1\| = m < \infty$ and $I_1^*(K_2(\cdot,y)v) = K_1(\cdot,y)v$.

Fix $x_1, \ldots, x_n \in X$, $v_1, \ldots, v_n \in \mathcal{C}$, then

$$\sum_{i,j=1}^{n} \langle K_1(x_i, x_j)v_j, v_i\rangle_{\mathcal{C}} = \|\sum_{j=1}^{n} K_1(\cdot, x_j)v_j\|^2_{\mathcal{H}_1} = \|I_1^*(\sum_{j=1}^{n} K_2(\cdot, x_j)v_j)\|^2_{\mathcal{H}_1}$$

$$\leq \|I_1^*\|^2 \cdot \|\sum_{j=1}^{n} K_2(\cdot, x_j)v_j\|^2_{\mathcal{H}_2} = \|I_1^*\|^2 \cdot \sum_{i,j=1}^{n} \langle K_2(x_i, x_j)v_j, v_i\rangle_{\mathcal{C}}.$$

Therefore, $K_1(x, y) \leq m^2 K_2(x, y)$. $\square$

Lemma 10.21. *Let $A, B \in B(\mathcal{C})$, $\quad v, w \in \mathcal{C}$,then $A(v \otimes \bar{w})B^* = Av \otimes \overline{Bw}$.*

Proof. $(A(v\otimes\bar{w})B^*)(z) = A(v\otimes\bar{w})B^*(z) = A[\langle B^*(z), w\rangle v] = \langle B^*(z), w\rangle Av = \langle z, Bw\rangle Av = [Av \otimes \overline{Bw}](z), \forall z$. Therefore $A(v \otimes \bar{w})B^* = Av \otimes \overline{Bw}$. $\square$

Theorem 10.22. *Let $\mathcal{H}_1, \mathcal{H}_2$ be two RKHS's of $\mathcal{C}$-valued functions on X with kernels $K_1(x, y), K_2(x, y)$ respectively, let $F : X \to B(\mathcal{C})$. Then t.f.a.e.:*

1) *$F \cdot \mathcal{H}_1 \subseteq \mathcal{H}_2$,*
2) *$M_F : \mathcal{H}_1 \to \mathcal{H}_2$, $\quad M_F(h_1) = F \cdot h_1$ is bounded,*
3) *$\exists\, c > 0$ such that $F(x)K_1(x, y)F(y)^* \leq c^2 K_2(x, y)$.The least such c is $\|M_F\| = c$.*

Proof. 1) $\Rightarrow$ 2) : Use Closed Graph Theorem.
2) $\Rightarrow$ 3) : Let $c = \|M_F\|$, then for any $h \in \mathcal{H}_1$, we have
$E^2_x(M_F h) = F(x) \cdot h(x) = F(x) \cdot E^1_x(h) \Rightarrow E^2_x M_F = F(x)E^1_x, \forall x$.
And $M_F^* E_y^{2^*} = E_y^{1^*} F(y)^*, \forall y, \qquad M_F \cdot M_F^* \leq c^2 \cdot I$.
So, $F(x)K_1(x, y)F(y)^* = F(x)E^1_x E_y^{1^*} F(y)^* = E^2_x M_F \cdot M_F^* E_y^{2^*}$
$\leq E^2_x(c^2 \cdot I)E_y^{2^*} = c^2 K_2(x, y)$.
3) $\Rightarrow$ 1) : There exists $c > 0$ such that $F(x)K_1(x, y)F(y)^* \leq c^2 K_2(x, y)$.
Let $h \in \mathcal{H}_1$, then $h(x) \otimes \overline{h(y)} \leq \|h\|_1^2 K_1(x, y)$.
$(F(x)h(x)) \otimes \overline{(F(y)h(y))} = F(x)[h(x) \otimes \overline{h(y)}]F(y)^*$
$\leq \|h\|_1^2 F(x)K_!(x, y)F(y)^* \leq c^2\|h\|_1^2 K_2(x, y)$
$\Rightarrow F(\cdot)h(\cdot) \in \mathcal{H}_2$ and $\|F(\cdot)h(\cdot)\|_2 \leq c\|h\|_1$,
i.e. $F \cdot \mathcal{H}_1 \subseteq \mathcal{H}_2$, moreover $\|M_F\| \leq c$. $\square$

Recall $M_f : \mathcal{H}_1 \to \mathcal{H}_2$, in scalar case we proved $\|f\|_\infty \leq \|M_f\|$.
What about in operator spaces?

Corollary 10.23. *If $M_F : \mathcal{H}_1 \to \mathcal{H}_2$ is bounded, then $M_F^*(E_y^* v) = E_y^* F(y)^* v$, i.e. $M_F^*(K(\cdot, y)v) = K(\cdot, y)F(y)^* v$.*

We have $\|M_F^*(E_y^* v)\|^2 \leq \|M_F\|^2 \|E_y^* v\|^2$.

$\|M_F^*(E_y^* v)\|^2 = \|E_y^* F(y)^* v\|^2 = \langle E_y^* F(y)^* v, E_y^* F(y)^* v\rangle = \langle F(y)K(y, y)F(y)^* v, v\rangle$,

and

$$\|M_F\|^2 \|E_y^* v\|^2 = \|M_F\|^2 \langle E_y^* v, E_y^* v\rangle = \|M_F\|^2 \langle K(y, y)v, v\rangle.$$

Therefore,

$$F(y)K(y, y)F(y)^* \leq \|M_F\|^2 K(y, y).$$

Assume $K(y,y)$ is invertible, then $K(y,y) = P^2$ with $P > 0$ invertible. Then $F(y)P^2F(y)^* \le \|M_F\|^2P^2 \Rightarrow (P^{-1}F(y)P)(PF(y)^*P^{-1}) \le \|M_F\|^2 \cdot I$ $\Rightarrow \|P^{-1}F(y)P\| \le \|M_F\|$. **Facts:** Given $T \in B(\mathcal{C})$, $\sigma(T) = \{\lambda \in \mathbb{C} :$

$(T - \lambda I)$ not invertible$\}$, then

1) $\sigma(T)$ is compact set,
2) $r(T) = \{|\lambda| : \lambda \in \sigma(T)\} \le \|T\|$, the spectral radius of T.
3) $r(T) = \liminf_n \|T^n\|^{1/n}$,
4) When $K(y,y)$ has no kernel, then $r(F(y)) \le \|M_F\|$.

Example: Look at $H^2(\mathbb{D})^m = \{\vec{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} : f_i \in H^2(\mathbb{D}),\ \|f\|^2 = \sum_{i=1}^m \|f_i\|^2_{H^2(\mathbb{D})}\}$, and $K(z,w) = \dfrac{1}{1 - z\bar{w}} \cdot I_m$, let $F : \mathbb{D} \to M_m$. Then by the theorem, $M_F : H^2(\mathbb{D})^m \to H^2(\mathbb{D})^m$ is bounded, and say

$$\|M_F\| = 1 \iff F(z)K(z,w)F(w)^* \le K(z,w),$$

i.e.

$$F(z)F(w)^* \frac{1}{1 - z\bar{w}} \cdot I_m \le \frac{1}{1 - z\bar{w}} \cdot I_m \iff \left(\frac{I - F(z)F(w)^*}{1 - z\bar{w}} \right) \ge 0.$$

Take $z = w = z_1$, then $\left(\dfrac{I - F(z)F(z)^*}{1 - |z|^2} \right) \ge 0$, which implies $I - F(z)F(z)^* \ge 0$, i.e. $I \ge F(z)F(z)^*$, so $\|F(z)\|_{M_m} \le 1$.
Therefore, $\sup\{\|F(z)\|_{M_m} : z \in \mathbb{D}\} \le \|M_F\|$. Assume $\sup\{\|F(z)\|_{M_m} : z \in$

$\mathbb{D}\} = C$, and let $h = \begin{pmatrix} h_1 \\ \vdots \\ h_m \end{pmatrix} \in H^2(\mathbb{D})^m$, then we have

$$\|M_F \cdot h\|^2 = \| \begin{pmatrix} \sum_{j=1}^m f_{1j}h_j \\ \vdots \\ \sum_{j=1}^n f_{mj}h_j \end{pmatrix} \|^2 = \sum_{i=1}^m \| \sum_{j=1}^m f_{ij}h_j\|^2_{H^2(\mathbb{D})}$$

$$= \lim_{r\to 1} \sum_{i=1}^m \frac{1}{2\pi} \int_0^{2\pi} |\sum_{j=1}^m f_{ij}(re^{i\theta})h_j(re^{i\theta})|^2 d\theta$$

$$= \lim_{r\to 1} \frac{1}{2\pi} \int_0^{2\pi} \| \begin{pmatrix} \sum_{j=1}^m f_{1j}(re^{i\theta})h_j(re^{i\theta}) \\ \vdots \\ \sum_{j=1}^m f_{mj}(re^{i\theta})h_j(re^{i\theta}) \end{pmatrix} \|^2_{\mathbb{C}^m} d\theta$$

$$= \lim_{r\to 1} \frac{1}{2\pi} \int_0^{2\pi} \|F(re^{i\theta})h(re^{i\theta}\|^2 d\theta \le \lim_{r\to 1} \frac{1}{2\pi} \cdot C^2 \int_0^{2\pi} \|h(re^{i\theta}\|^2 d\theta$$

$$\le C^2 \left[\|h_1\|^2 + \cdots + \|h_m\|^2\right] = C^2 \cdot \|h\|^2 \Rightarrow \|M_F\| \le C.$$

Therefore, $\|M_F\| = \sup\{\|F(z)\|_{M_m} : z \in \mathbb{D}\}$.

11. Linear Fractional Maps

Consider $\varphi_U(z) = \dfrac{az+b}{cz+d}$, $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and $\varphi_{\alpha\cdot U} = \varphi_u(z)$.
WLOG, assume $\det u = 1$, i.e. $ad - bc = 1$.
Let $J = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, which is called **a J-contraction** if and only if $UJU^* \le J$, and U is **J-unitary** if and only if $UJU^* = J$.

Proposition 11.1. *If U is J-unitary, then U^{-1} is J-unitary, too.*

Proof. $UJU^* = J \Rightarrow U^{-1}JU^{-1*} = U^{-1}(UJU^*)U^{-1*} = J$. □

Proposition 11.2. $\varphi_V \circ \varphi_U = \varphi_{V\cdot U}$

Proof. Let $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $V = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$, then $V \cdot U = \begin{bmatrix} ae+cf & be+df \\ ag+ch & bg+dh \end{bmatrix}$. Then,

$$\varphi_V(\varphi_U(z)) = \varphi_V\left(\frac{az+b}{cz+d}\right) = \frac{e\left[\frac{az+b}{cz+d}\right] + f}{g\left[\frac{az+b}{cz+d}\right] + h}$$

$$= \frac{(ae+cf)z + (be+df)}{(ag+ch)z + (bg+dh)} = \varphi_{V\cdot U}(z).$$

□

Theorem 11.3. *The map $\varphi_U : \mathbb{D} \to \mathbb{D}$ is an automorphism if and only if U is a J-contraction; and $\varphi_U : \mathbb{D} \to \mathbb{D}$ is an onto map if and only if U is J-unitary.*

Proof. Assume $\varphi_U(\mathbb{D}) \subseteq \mathbb{D}$, which implies $|\varphi_U(z)|^2 \le 1 \iff |az+b|^2 \le |cz+d|^2$. This implies, $|a|^2|z|^2 + a\bar{b}z + \bar{a}b\bar{z} + |b|^2 \le |c|^2|z|^2 + c\bar{d}z + \bar{c}d\bar{z} + |d|^2 \Rightarrow 0 \le (|c|^2 - |a|^2)|z|^2 + (c\bar{d} - a\bar{b})z + (\bar{c}d - \bar{a}b)\bar{z} + (|d|^2 - |b|^2)$, **(*)**.
Then, $\forall t \in [0,1]$,we have $0 \le (|c|^2 - |a|^2)t^2 - 2(c\bar{d} - a\bar{b})t + (|d|^2 - |b|^2)$. This implies

$$0 \le \langle \begin{bmatrix} |c|^2 - |a|^2 & |c\bar{d} - a\bar{b}| \\ |c\bar{d} - a\bar{b}| & |d|^2 - |b|^2 \end{bmatrix} \begin{pmatrix} t \\ -1 \end{pmatrix}, \begin{pmatrix} t \\ -1 \end{pmatrix} \rangle,$$

or $\forall z \in \mathbb{D}$, say $t = \bar{z}, |z| \le 1$, then

$$0 \le \langle \begin{bmatrix} |c|^2 - |a|^2 & c\bar{d} - a\bar{b} \\ c\bar{d} - a\bar{b} & |d|^2 - |b|^2 \end{bmatrix} \begin{pmatrix} \bar{z} \\ -1 \end{pmatrix}, \begin{pmatrix} \bar{z} \\ -1 \end{pmatrix} \rangle.$$

If U is J-contraction, then $U^* \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} U \le \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Then,

$$\begin{bmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \bar{a} & -\bar{c} \\ \bar{b} & -\bar{d} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} |a|^2 - |c|^2 & \bar{a}b - \bar{c}d \\ a\bar{b} - c\bar{d} & |b|^2 - |d|^2 \end{bmatrix} \le \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

This implies $A = \begin{bmatrix} 1 + |c|^2 - |a|^2 & \bar{c}d - \bar{a}b \\ c\bar{d} - a\bar{b} & |d|^2 - |b|^2 - 1 \end{bmatrix} \ge 0$.

Hence, $0 \le \langle A \begin{pmatrix} \bar{z} \\ -1 \end{pmatrix}, \begin{pmatrix} \bar{z} \\ -1 \end{pmatrix} \rangle = (*) + |z|^2 - 1 \le (*)$. Therefore, **(*)** is positive. So, U, U^{-1} J-contractive implies $\varphi_U : \mathbb{D} \to \mathbb{D}$ is an automorphism. □

12. Linear Fractional Operator-Valued Maps

Definition 12.1. *Let $\mathcal{H}$ be a Hilbert space, $A, B, C, D \in B(\mathcal{H})$, let $U = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, $J = \begin{bmatrix} I_{\mathcal{H}} & 0 \\ 0 & -I_{\mathcal{H}} \end{bmatrix}$. Then U is called a J-contraction if $U^*JU \le J$, and U is called J-unitary if $U^*JU = J$.*

Proposition 12.2. *If U is J-unitary, then U is left-invertible.*

Proof. $J^{-1}(U^*JU) = J^{-1}J = I \Rightarrow \underbrace{(J^{-1}U^*J)}_{U^{-1}} = I$. □

- In finite dimensions, U is invertible and $U^{-1} = JU^*J$.
- Let $\mathcal{H} = l^2$, let S be the forward shift operator, i.e. $Se_j = e_{j+1}$, let $U = \begin{bmatrix} S & 0 \\ 0 & S \end{bmatrix}$. Then $U^*JU = J$, so U is J-unitary, but not invertible (it is not onto).

Proposition 12.3. *If U is J-unitary and invertible, then U^{-1} is J-unitary, too.*

Proof. $U^{-1*}JU^{-1} = U^{-1*}[U^*JU]U^{-1} = J$. □

Proposition 12.4. *Let U be an invertible operator(i.e. $dim\mathcal{H} < \infty$), then U is J-unitary if and only if U^* is J-unitary.*

Proof. Assume U is J-unitary and invertible, then $U^{-1} = JU^*J$ is J-unitary by previous propositon, which implies $(JU^*J)^*J(JU^*J) = J$. But, $(JU^*J)^*J(JU^*J) = JUJJJU^*J = JUJU^*J$. Hence, $(U^*)^*JU^* = UJU^* = J(JUJU^*J)J = JJJ = J$, i.e. U^* is J-unitary. The converse follows by reversing roles of U and U^*. □

Example: Let $\|B\| \le 1$, and $U = \begin{bmatrix} (1-BB^*)^{-1/2} & (1-BB^*)^{-1/2}B \\ (1-BB^*)^{-1/2}B^* & (1-BB^*)^{-1/2} \end{bmatrix}$
$= \begin{bmatrix} (1-BB^*)^{-1/2} & 0 \\ 0 & (1-BB^*)^{-1/2} \end{bmatrix} \begin{bmatrix} 1 & B \\ B^* & 1 \end{bmatrix}$. Then, we have
$\begin{bmatrix} 1 & B \\ B^* & 1 \end{bmatrix} J \begin{bmatrix} 1 & B \\ B^* & 1 \end{bmatrix} = \begin{bmatrix} 1 & B \\ B^* & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & B \\ B^* & 1 \end{bmatrix} = \begin{bmatrix} 1 & -B \\ B^* & -1 \end{bmatrix} \begin{bmatrix} 1 & B \\ B^* & 1 \end{bmatrix}$
$= \begin{bmatrix} 1-BB^* & 0 \\ 0 & B^*B-1 \end{bmatrix} = \begin{bmatrix} 1-BB^* & 0 \\ 0 & -(1-B^*B) \end{bmatrix}$
$= \begin{bmatrix} (1-BB^*)^{1/2} & 0 \\ 0 & (1-B^*B)^{1/2} \end{bmatrix} J \begin{bmatrix} (1-BB^*)^{1/2} & 0 \\ 0 & (1-B^*B)^{1/2} \end{bmatrix}$.
Therefore, $UJU^* = J \Rightarrow U$ is J-unitary..
Let $U^* = \begin{bmatrix} (1-BB^*)^{-1/2} & (1-BB^*)^{-1/2}B \\ (1-BB^*)^{-1/2}B^* & (1-BB^*)^{-1/2} \end{bmatrix}$.

Proposition 12.5. *Let $\|B\| \le 1$, then $U_B = \begin{bmatrix} (1-BB^*)^{-1/2} & -B(1-B^*B)^{-1/2} \\ -B^*(1-BB^*)^{-1/2} & (1-B^*B)^{-1/2} \end{bmatrix}$ is J -unitary and invertible.(which implies U_B^* is J-unitary.)*

Proof. $\begin{bmatrix} 1 & -B \\ -B^* & 1 \end{bmatrix} J \begin{bmatrix} 1 & -B \\ -B^* & 1 \end{bmatrix} = \begin{bmatrix} 1 & -B \\ -B^* & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & -B \\ -B^* & 1 \end{bmatrix} =$
$\begin{bmatrix} 1-BB^* & 0 \\ 0 & B^*B-1 \end{bmatrix}$. Note that $U_B = \begin{bmatrix} 1 & -B \\ -B^* & 1 \end{bmatrix} \begin{bmatrix} (1-BB^*)^{-1/2} & 0 \\ 0 & (1-B^*B)^{-1/2} \end{bmatrix}$.
It is easy to check that $U_B^*JU_B = J$, i.e. U_B is J-unitary.
To see that U_B is invertible, it's enough to show that $\begin{bmatrix} 1 & -B \\ -B^* & 1 \end{bmatrix}$ is invertible. But, using Cholesky, we have $\begin{bmatrix} 1 & -B \\ -B^* & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -B^* & R \end{bmatrix} \begin{bmatrix} 1 & -B \\ 0 & R \end{bmatrix}$,
where $R = \sqrt{1-B^*B}$ is positive and invertible. Hence, $\begin{bmatrix} 1 & -B \\ 0 & R \end{bmatrix}^{-1} =$
$\begin{bmatrix} 1 & BR^{-1} \\ 0 & R^{-1} \end{bmatrix}$, and $\begin{bmatrix} 1 & o \\ -B^* & R \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ R^{-1}B^* & R^{-1} \end{bmatrix}$. We are done. □

Definition 12.6. *Let* $U = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, *then set* $\psi_U(X) = (AX+B)(CX+D)^{-1}$ *for any* X *where the inverse exists.*

Theorem 12.7. *Let* $U = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ *be a* J*-contraction,* $\|X\| \leq 1$. *Then* $(CX+D)$ *is invertible and* $\|\psi_U(X)\| < 1$, *i.e.* $\psi_U : \mathcal{B} \to \mathcal{B}$, *where* $\mathcal{B}$ *denotes the unit ball of* $B(\mathcal{H})$, *when* $dim\mathcal{H} < \infty$.

Proof. $U^*JU \leq J \Rightarrow \begin{bmatrix} X^* & I \end{bmatrix} [U^*JU] \begin{bmatrix} X \\ I \end{bmatrix} \leq \begin{bmatrix} X^* & I \end{bmatrix} J \begin{bmatrix} X \\ I \end{bmatrix}$. Then,
$(AX+B)^*(AX+B) - (CX+D)^*(CX+D) = \begin{bmatrix} X^* & I \end{bmatrix} [U^*JU] \begin{bmatrix} X \\ I \end{bmatrix}$
$\leq \begin{bmatrix} X^* & I \end{bmatrix} J \begin{bmatrix} X \\ I \end{bmatrix} = X^*X - I$, which implies

$$(AX+B)^*(AX+B) + I \leq X^*X + (CX+D)^*(CX+D),$$

therefore

$$0 < (1 - \|X\|^2)I < I - X^*X \leq (CX+D)^*(CX+D),$$

which gives

$$\delta \cdot I \leq (CX+D)^*(CX+D) = M.$$

Hence, there exists R such that $RM = [R(CX+D)^*](CX+D) = I$ when $dim\mathcal{H} < \infty$, so $(CX+D)$ is invertible.
We have $(AX+B)^*(AX+B) \leq (CX+D)^*(CX+D) + \underbrace{(X^*X - I)}_{<0} <$ $(CX+D)^*(CX+D) \Rightarrow (CX+D)^{*-1}(AX+B)^*(AX+B)(CX+D)^{-1} < I$, which gives to us the wanted result

$$\psi_U^*(X) \cdot \psi_U(X) < I \Rightarrow \|\psi_U(X)\| < 1.$$

$\square$

Proposition 12.8. *Let* $U = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ *and* $V = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$ *be a* J*-contractions. Then* $\psi_V \circ \psi_U(X) = \psi_{VU}(X)$.

Proof.

$$\begin{aligned}
\psi_V \circ \psi_U(X) &= \psi_V((AX+B)(CX+D)^{-1}) \\
&= [E(AX+B)(CX+D)^{-1} + F][G(AX+B)(CX+D)^{-1} + H]^{-1} \\
&= [(EAX+EB+FCX+FD)(CX+D)^{-1}][(GAX+GB+HCX+HD)(CX+D)^{-1}]^{-1} \\
&= [(EA+FC)X+(EB+FD)][(GA+HC)X+(GB+HD)]^{-1} = \psi_{VU}(X).
\end{aligned}$$

$\square$

13. Matrix-Valued Pick Interpolation

Theorem 13.1. *Let $z_1, \dots, z_n \in \mathbb{D}$, $X_1, \dots, X_n \in M_m$, $\alpha \in \mathbb{D}$, and let $U = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ be J-unitary. Then*

$$\left(\frac{I - X_i X_j^*}{1 - z_i z_j^*} \right) \geq 0 \iff \left(\frac{I - \psi_U(X_i^*)^* \psi_U(X_j^*)}{1 - \varphi_\alpha(z_i)\overline{\varphi_\alpha(z_j)}} \right).$$

Proof. Since $\psi_U^{-1} = \psi_{U^{-1}}$, where U^{-1} is J-unitary, and $\varphi_\alpha^{-1} = \varphi_{\alpha^{-1}}$, then it will be enough to prove $' \Rightarrow '$ direction. It is also sufficient to prove:

$$(1) \qquad \left(\frac{I - X_i X_j^*}{1 - z_i z_j^*} \right) \geq 0 \Rightarrow \left(\frac{I - X_i X_j^*}{1 - \varphi_\alpha(z_i)\overline{\varphi_\alpha(z_j)}} \right) \geq 0,$$

and

$$(2) \qquad \left(\frac{I - X_i X_j^*}{1 - z_i z_j^*} \right) \geq 0 \Rightarrow \left(\frac{I - \psi_U(X_i^*)^* \psi_U(X_j^*)}{1 - z_i z_j^*} \right) \geq 0.$$

The proof of (1) is identical to the scalar case, so we will focus on (2). $\begin{bmatrix} X_i & I \end{bmatrix} [U^* J U] \begin{bmatrix} X_j^* \\ I \end{bmatrix} = (AX_i^* + B)^*(AX_j^* + B) - (CX_i^* + D)^*(CX_j^* + D)$, and since U is J-unitary, i.e. $U^* J U = J$, then $\begin{bmatrix} X_i & I \end{bmatrix} [U^* J U] \begin{bmatrix} X_j^* \\ I \end{bmatrix} = \begin{bmatrix} X_i & I \end{bmatrix} J \begin{bmatrix} X_j^* \\ I \end{bmatrix} = X_i X_j^* - I$, therefore

$$(AX_i^* + B)^*(AX_j^* + B) - (CX_i^* + D)^*(CX_j^* + D) = X_i X_j^* - I.$$

Note that,

$$\begin{gathered} I - X_i X_j^* = (CX_i^* + D)^*(CX_j^* + D) - (AX_i^* + B)^*(AX_j^* + B) \\ = (CX_i^* + D)^*[I - (CX_i^* + D)^{*-1}(AX_i^* + B)^*(AX_j^* + B)(CX_j^* + D)^{-1}](CX_j^* + D) \\ = (CX_i^* + D)^*[I - \psi_U(X_i^*)^* \psi_U(X_j^*)](CX_j^* + D) \end{gathered}$$

$$\Rightarrow \left(\frac{I - \psi_U(X_i^*)^* \psi_U(X_j^*)}{1 - z_i z_j^*} \right) = \left((CX_i^* + D)^{*-1} \left(\frac{I - X_i X_j^*}{1 - z_i z_j^*} \right) (CX_j^* + D)^{-1} \right).$$

Denoting $\tilde{D} = \begin{bmatrix} (CX_1^* + D)^{*-1} & & 0 \\ & \ddots & \\ 0 & & (CX_n^* + D)^{*-1} \end{bmatrix}$, a diagonal matrix, we get the wanted result as follows

$$\left(\frac{I - \psi_U(X_i^*)^* \psi_U(X_j^*)}{1 - z_i z_j^*} \right) = \tilde{D} \left(\frac{I - X_i X_j^*}{1 - z_i z_j^*} \right) \tilde{D}^* \geq 0.$$

□

Proposition 13.2. *Let $\|X\| \leq 1$, then $(I - X^* X)^{\pm 1/2} X^* = X^*(I - XX^*)^{\pm 1/2}$.*

Proof. Let $P = I - X^*X \geq 0$, $Q = I - XX^* \geq 0$.
Note $PX^* = X^* - X^*XX^* = X^*(I - XX^*) = X^*Q$. Hence,
$P^2X^* = P(PX^*) = PX^*Q = X^*Q^2$, and by induction $P^nX^* = X^*Q^n$.
Letting p be a polynomial, we get $p(P)X^* = X^*p(Q)$.
Now, let $\{p_n(t)\}$ be a sequence of polynomials,
that converges uniformly to the function $\sqrt{t}$ on $0 \leq t \leq 1$.
Then $P^{1/2}X^* = \lim_n p_n(P)X^* = \lim_n X^*p_n(Q) = X^*Q^{1/2}$, and
$P^{-1/2}X^* = P^{-1/2}(X^*Q^{1/2})Q^{-1/2} = P^{-1/2}(P^{1/2}X^*)Q^{-1/2} = X^*Q^{-1/2}$. □

Theorem 13.3 (Matrix-Valued Version of Pick)**.** *Let $z_1, \ldots, z_n \in \mathbb{D}$, $X_1, \ldots, X_n \in M_m$. Then there exists $F = (f_{ij}) : \mathbb{D} \to M_m$ with $f_{ij} \in H^\infty$, $\|F\|_\infty \leq 1$ and* $F(z_i) = X_i \iff \left(\dfrac{I - X_iX_j^*}{1 - z_i\bar{z}_j}\right) \geq 0.$

Proof. $(\Rightarrow)$: Look at $M_F : H^2(\mathbb{D})^{(m)} \to H^2(\mathbb{D})^{(m)}$, $\|M_F\| = \|F\|_\infty \leq 1$.
$K(z,w) = \dfrac{1}{1 - z\bar{w}} \cdot I_m \Rightarrow 0 \leq K(z,w) - F(z)K(z,w)F(w)^*$,
i.e. $\left(\dfrac{I}{1 - z_i\bar{z}_j}\right) - \left(\dfrac{F(z_i)F(z_j)^*}{1 - z_i\bar{z}_j}\right) \geq 0 \Rightarrow \left(\dfrac{I - X_iX_j^*}{1 - z_i\bar{z}_j}\right) \geq 0.$
$(\Leftarrow)$: We will do it by induction on n:
It is clear for $n = 1$, assume $F(z_1) = X$ constant, done.
Assume it is true for all sets of n points in $\mathbb{D}$, and n matrices X's in M_m.
Given $z_0, z_1, \ldots, z_n \in \mathbb{D}$, $X_0, X_1, \ldots, X_n \in M_m$ with $\left(\dfrac{I - X_iX_j^*}{1 - z_i\bar{z}_j}\right)^n_{i,j\geq 1} \geq 0.$
Let $\alpha = z_0$ and $U = U_{X_0} = \begin{bmatrix} (1 - X_0X_0^*)^{-1/2} & -X_0(1 - X_0^*X_0)^{-1/2} \\ -X_0^*(1 - X_0X_0^*)^{-1/2} & (1 - X_0^*X_0)^{-1/2} \end{bmatrix}$.
Then,
$\psi_{U_{X_0}}(X_0^*) = \left((I - X_0^*X_0)^{-1/2}X_0^* - X_0^*(I - X_0X_0^*)^{-1/2}\right) \cdot$
$\cdot \left(-X_0(I - X_0^*X_0)^{-1/2}X_0^* + (I - X_0X_0^*)^{-1/2}\right)^{-1}$.
But, by previous proposition, $(I - X_0^*X_0)^{-1/2}X_0^* = X_0^*(I - X_0X_0^*)^{-1/2}$,
since $\|X_0\| \leq 1$, therefore $\psi_{U_{X_0}}(X_0^*) = 0$.
Next, by last theorem $\left(\dfrac{I - \psi_{U_{X_0}}(X_i^*)^*\psi_{U_{X_0}}(X_j^*)}{1 - \varphi_\alpha(z_i)\overline{\varphi_\alpha(z_j)}}\right) \geq 0 \quad (*)$.
Let $W_i = \psi_{U_{X_0}}(X_i^*)^* \Rightarrow W_0 = 0$, and $\lambda_i = \varphi_\alpha(z_i) \Rightarrow \lambda_0 = 0$, but $\lambda_i \neq 0, i = 1, \ldots, n$, then using $(*)$, we get

$$\begin{bmatrix} I & \cdots & I \\ \vdots & \left(\dfrac{I - W_iW_j^*}{1 - \lambda_i\bar{\lambda}_j}\right) & \\ I & & \end{bmatrix} \geq 0 \Rightarrow \underbrace{\left(\frac{I - W_iW_j^*}{1 - \lambda_i\bar{\lambda}_j} - I\right) \geq 0}_{\text{by} Cholesky's lemma}.$$

i.e. we get

$$\left(\frac{\lambda_i\bar{\lambda}_j - W_iW_j^*}{1-\lambda_i\bar{\lambda}_j}\right) \geq 0 \Rightarrow \left[\frac{I - \left(\frac{W_i}{\lambda_i}\right)\left(\frac{W_j}{\lambda_j}\right)^*}{1-\lambda_i\bar{\lambda}_j}\right]_{n\times n} \geq 0.$$

Hence, by the inductive hypothesis, $\exists H : \mathbb{D} \to M_n$, with $\|H\|_\infty \leq 1$ such that $H(\lambda_i) = \frac{W_i}{\lambda_i}$, $i = 1, 2, \ldots, n$, and $G(z) = zH(z)$, $\|G\|_\infty \leq 1$, analytic, which gives $G(\lambda_i) = \lambda_i F(\lambda_i) = W_i$, $i = 0, 1, 2, \ldots, n$.
Let $F(z) = \psi_{U_{X_0}^{-1}}(G(\varphi_{-\alpha}(z))^*)^*$, which is analytic.
Then $F(z_i) = \psi_{U_{X_0}^{-1}}(G(\lambda_i)^*)^* = \psi_{U_{X_0}^{-1}}(W_i^*)^* = (X_i^*)^* = X_i$. $\square$

References

[1] J. Agler and J. McCarthy, text
[2] Aronszajn, ..
[3] John B. Conway, *A Course in Functional Analysis,* Graduate Texts in Mathematics 96, Springer-Verlag, New York, 1990.
[4] L. de Branges
[5] L. de Branges and J. Rovnyak,

Department of Mathematics, University of Houston, Houston, Texas 77204-3476, U.S.A.

E-mail address: vern@math.uh.edu