Cambridge University Press
978-1-107-10409-9 - An Introduction to the Theory of Reproducing Kernel Hilbert Spaces
Vern I . Paulsen and Mrinal Raghupathi
Excerpt
More information

# Part I
## General theory

# 1
# Introduction

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. However, to understand their structure and theory, the reader only needs a background in the theory of Hilbert spaces and real analysis.

In this chapter, we introduce the reader to the formal definition of the reproducing kernel Hilbert space and present a few of their most basic properties. This beautiful theory is filled with important examples arising in many diverse areas of mathematics and, consequently, our examples often require mathematics from many different areas. Rather than keep the reader in the dark by insisting on pedagogical purity, we have chosen to present examples that require considerably more mathematics than we are otherwise assuming. So the reader should not be discouraged if they find that they do not have all the prerequisites to fully appreciate the examples. The examples are intended to give a glimpse into the many areas where these spaces arise.

## 1.1 Definition

We will consider Hilbert spaces over the field of either real numbers, $\mathbb{R}$, or complex numbers, $\mathbb{C}$. We will use $\mathbb{F}$ to denote either $\mathbb{R}$ or $\mathbb{C}$, so that when we wish to state a definition or result that is true for either real or complex numbers, we will use $\mathbb{F}$.

Let $X$ be a set. We denote by $\mathcal{F}(X, \mathbb{F})$ the set of functions from $X$ to $\mathbb{F}$. The set $\mathcal{F}(X, \mathbb{F})$ is a vector space over the field $\mathbb{F}$ with the operations of addition, $(f + g)(x) = f(x) + g(x)$, and scalar multiplication, $(\lambda \cdot f)(x) = \lambda \cdot (f(x))$.

**Definition 1.1.** Let $X$ be a set. We will call a subset $\mathcal{H} \subseteq \mathcal{F}(X, \mathbb{F})$ a REPRODUCING KERNEL HILBERT SPACE or, more briefly, an RKHS on $X$ if

(i) $\mathcal{H}$ is a vector subspace of $\mathcal{F}(X, \mathbb{F})$;
(ii) $\mathcal{H}$ is endowed with an inner product, $\langle , \rangle$, with respect to which $\mathcal{H}$ is a Hilbert space;
(iii) for every $x \in X$, the linear EVALUATION FUNCTIONAL, $E_x : \mathcal{H} \to \mathbb{F}$, defined by $E_x(f) = f(x)$, is bounded.

If $\mathcal{H}$ is an RKHS on $X$, then an application of the Riesz representation theorem shows that the linear evaluation functional is given by the inner product with a unique vector in $\mathcal{H}$. Therefore, for each $x \in X$, there exists a unique vector, $k_x \in \mathcal{H}$, such that for every $f \in \mathcal{H}$, $f(x) = E_x(f) = \langle f, k_x \rangle$.

**Definition 1.2.** The function $k_x$ is called the REPRODUCING KERNEL FOR THE POINT $x$. The function $K : X \times X \to \mathbb{F}$ defined by

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

is called the REPRODUCING KERNEL FOR $\mathcal{H}$.

Note that we have

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

so that

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

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

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

We now look at some examples of reproducing kernel Hilbert spaces. Our examples are drawn from function theory, differential equations and statistics.

## 1.2 Basic examples

### 1.2.1 $\mathbb{C}^n$ as an RKHS

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

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

denote the usual inner product. Whenever we speak of $\mathbb{C}^n$ as a Hilbert space, this is the inner product that we shall intend. Of course, if we let

$X = \{1, \ldots, n\}$, then we could also think of a complex $n$-tuple as a function $v : X \to \mathbb{C}$, where $v(j) = v_j$. With this identification, $\mathbb{C}^n$ becomes the vector space of all functions on $X$. If we let $\{e_j\}_{j=1}^n$ denote the "canonical" orthonormal basis for $\mathbb{C}^n$, i.e. let $e_j$ be the function, $e_j(i) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$, then for every $v \in \mathbb{C}^n$ we have

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

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

Note that the reproducing kernel for $\mathbb{C}^n$ is given by

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

which can be thought of as the identity matrix.

More generally, given any (discrete) set $X$, we set

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

Given $f, g \in \ell^2(X)$, we define $\langle f, g \rangle = \sum_{x \in X} f(x)\overline{g(x)}$. With these definitions $\ell^2(X)$ becomes a Hilbert space of functions on $X$. If for a fixed $y \in X$, we let $e_y \in \ell^2(X)$ denote the function given by $e_y(x) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$, then it is easily seen that $\{e_y\}_{y \in X}$ is an orthonormal basis for $\ell^2(X)$ and that $\langle f, e_y \rangle = f(y)$, so that these functions are also the reproducing kernels and as before

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

### 1.2.2 A non-example

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 $x \in [0, 1]$, every function in the original space $C([0, 1])$ has a value at this point, so point evaluation is well-defined on this dense subspace. Is it possible that it extends to give a bounded linear functional on $L^2[0,1]$? Not surprisingly, the answer is no. Keeping $0 < x < 1$ fixed, the sequence

$$f_n(t) = \begin{cases} (\frac{t}{x})^n & 0 \le t \le x \\ (\frac{1-t}{1-x})^n & x < t \le 1 \end{cases}$$

is in $C([0,1])$, satisfies $f_n(x) = 1$ for all $n$ and $\lim_n \|f_n\|_{L^2[0,1]} = 0$. This shows that the evaluation functional, although defined on this dense subspace, is unbounded for each $0 < x < 1$ and hence has no bounded extension to all of $L^2[0,1]$. It is easy to see how to define an analogous sequence of functions for the case $x = 0$ and $x = 1$.

Thus, there is no bounded way to extend the values of functions in $C([0,1])$ at points to regard elements in the completion $L^2[0,1]$ as having values at points. In particular, we cannot give $L^2[0,1]$ the structure of an RKHS on [0,1]. One of the remarkable successes of measure theory is showing that this completion can be regarded as equivalence classes of functions, modulo sets of measure 0, but this of course makes it difficult to talk about values at all points, since elements are not actual functions.

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

## 1.3 Examples from analysis

### 1.3.1 Sobolev spaces on [0,1]

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

A function $f : [0,1] \to \mathbb{R}$ is ABSOLUTELY CONTINUOUS provided that for every $\epsilon > 0$ there exists $\delta > 0$, so that when $(x_1, y_1), \ldots, (x_n, x_n)$ are any non-overlapping intervals contained in $[0,1]$ with $\sum_{j=1}^n |y_j - x_j| < \delta$, then $\sum_{j=1}^n |f(y_j) - f(x_j)| < \epsilon$. An important theorem from measure theory tells us that $f$ is absolutely continuous if and only if $f'(x)$ exists for almost all $x$, the derivative is integrable and, up to a constant, $f$ is equal to the integral of its derivative. Thus, absolutely continuous functions are the functions for which the first fundamental theorem of calculus applies. Let $\mathcal{H}$ denote the set of absolutely continuous functions $f : [0,1] \to \mathbb{R}$ such that $f'$ is square-integrable and satisfies $f(0) = f(1) = 0$. It is not hard to see that the set $\mathcal{H}$ is a vector space of functions on $[0,1]$.

In order to make $\mathcal{H}$ a Hilbert space we endow $\mathcal{H}$ with the nonnegative, sesquilinear form,

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

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

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

Thus, by the Cauchy-Schwartz inequality,

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

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

All that remains to show that $\mathcal{H}$ is an RKHS is to show that it is complete in the norm induced by its inner product. If $\{f_n\}$ is a Cauchy sequence in this norm, then $\{f_n'\}$ is Cauchy in $L^2[0, 1]$ and hence there exists $g \in L^2[0, 1]$ to which this sequence converges in the $L^2$-sense. 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.. Note that even though $g$ was only an equivalence class of functions, $\int_0^x g(t)dt$ was independent of the particular function chosen from the equivalence class. Hence, $f' \in L^2[0, 1]$. Finally, $f(0) = \lim_n f_n(0) = 0 = \lim_n f_n(1) = f(1)$. Thus, $\mathcal{H}$ is complete and so is an RKHS on $[0, 1]$.

We now wish to find the kernel function. We know that $f(x) = \int_0^1 f'(t)\chi_{[0,x]}(t)dt$. Thus, if we could solve the boundary-value problem,

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

then $g \in \mathcal{H}$ with $f(x) = \langle f, g\rangle$ and so $g = k_x$.

Unfortunately, this boundary-value problem has no continuous solution! Yet we know the function $k_x(t)$ exists and is continuous.

Instead, to find the kernel function we formally derive a different boundary-value problem. Then we will show that the function we obtain by this formal solution belongs to $\mathcal{H}$ and we will verify that it is the kernel function.

To find $k_x(t)$, we first apply integration by parts. We have

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

If we let $\delta_x$ denote the formal Dirac-delta function, then $f(x) = \int_0^1 f(t)\delta_x(t)dt$. Thus, it appears that we need to solve the boundary-value problem,

$$-k_x''(t) = \delta_x(t), k_x(0) = k_x(1) = 0.$$

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

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

After formally obtaining this solution, it can now be easily seen that

$$k_x'(t) = \begin{cases} (1-x) & t < x \\ -x & t > x \end{cases},$$

so $k_x$ is differentiable except at $x$ (thus almost everywhere), is equal to the integral of its derivative (and so is absolutely continuous), $k_x(0) = k_x(1)$ and $k_x'$ is square-integrable. Hence, $k_x \in \mathcal{H}$. Finally, for $f \in \mathcal{H}$,

$$\begin{aligned} \langle f, k_x \rangle = \int_0^1 f'(t)k_x'(t)dt &= \int_0^x f'(t)(1-x)dt + \int_x^1 f'(t)(-x)dt \\ =&(f(x) - f(0))(1-x) - x(f(1) - f(x)) = f(x) \end{aligned}$$

Thus, $k_x(t)$ satisfies all the conditions to be the reproducing kernel for $\mathcal{H}$.

Note that $\|E_x\|^2 = \|k_x\|^2 = K(x,x) = x(1-x)$, which considerably improves our original estimate of $\|E_x\|$. The fact that we are able to obtain the precise value of $\|E_x\|$ from the fact that we had an RKHS gives some indication of the utility of this theory. It takes a pretty sharp analyst to see that the original estimate of $|f(x)| \le \|f\|\sqrt{x}$ can be improved, but given the theory we now know that $|f(x)| \le \|f\|\sqrt{x(1-x)}$ and that this is the best possible inequality.

If we change $\mathcal{H}$ slightly by considering instead the space $\mathcal{H}_1$ of absolutely continuous functions $f : [0,1] \to \mathbb{R}$ such that $f' \in L^2[0,1]$ and $f(0) = 0$, but retain the same inner product, then $\mathcal{H}_1$ will still be an RKHS on $[0,1]$ and will contain $\mathcal{H}$ as a codimension one subspace. In fact, every function in $\mathcal{H}_1$ can be expressed uniquely as a function in $\mathcal{H}$ plus a scalar multiple of the function $f(x) = x$.

We leave it to the exercises to verify this claim and to compute the kernel function of $\mathcal{H}_1$. Not too surprisingly, one finds that the kernel function for $\mathcal{H}_1$

is determined by a first-order boundary-value problem, instead of the second-order problem needed when two boundary conditions were specified for $\mathcal{H}$. In later chapters we will return to this example.

### 1.3.2 The Paley-Wiener spaces: an example from signal processing

If we imagine that a function $f : \mathbb{R} \to \mathbb{R}$ represents the amplitude $f(t)$ of a sound wave at time $t$, then the *Fourier transform* of $f$,

$$\widehat{f}(x) = \int_{-\infty}^{\infty} f(t)e^{-2\pi ixt}\,dt,$$

is an attempt to decompose $f$ into "frequencies" with the number $\widehat{f}(x)$ representing the magnitude of the contribution of $e^{2\pi ixt}$. Of course, the above integral is not defined for every function $f$ and various different sets of hypotheses can be made, depending on the situation of interest, to make this rigorous. These hypotheses often lead one to consider various reproducing kernel Hilbert spaces. One family of such examples is the Paley-Wiener spaces.

The Paley-Wiener spaces play an important role in Fourier analysis and sampling theory, especially when one wishes to study Fourier transforms of functions that are compactly supported in either the time or frequency domain.

Suppose that we wish to consider only functions that are nonzero for a finite time, say $[-A, +A]$ for some $A > 0$. If we assume that $f$ is square- integrable on $[-A, +A]$, then the Fourier transform integral exists and we define the *Paley-Wiener space* $PW_A$ to be the set of functions that are Fourier transforms of $L^2$ functions whose support is contained in the interval $[-A, +A]$. That is,

$$PW_A = \{\widehat{f} : f \in L^2[-A, +A]\}.$$

Note that even though an element $f \in L^2[-A, +A]$ is really only an equivalence class of functions, the number $\widehat{f}(x)$ is well-defined, and independent of the function chosen from the equivalence class. Thus, even though we have seen that $L^2$ functions can not form a reproducing kernel Hilbert space, their Fourier transforms, i.e. the space $PW_A$, is a well-defined set of concrete functions on $\mathbb{R}$.

We claim that, endowed with a suitable norm, the space $PW_A$ is a reproducing kernel Hilbert space on $\mathbb{R}$. Let $F \in PW_A$, then there exists a function $f \in L^2[-A, A]$ such that

$$F(x) = \int_{-A}^{A} f(t)e^{-2\pi ixt}\,dt.$$

In fact, up to almost everywhere equality, there is a unique such function. That is, the linear map

$$\widehat{\ }: L^2[-A,+A] \to PW_A,$$

is one-to-one. To see this we use the well-known fact that the functions $e^{2\pi int/A}$, $n \in \mathbb{Z}$ are an orthonormal basis for $L^2[-A,+A]$. Thus, $\widehat{f}(n/A) = F(n/A) = 0$ for every $n \in \mathbb{Z}$ implies $f = 0$, a.e.. Since $\widehat{\ }: L^2[-A,+A] \to PW_A$ is linear, one-to-one and onto, if we define a norm on $PW_A$ by setting

$$\|\hat{f}\| = \|f\|_2,$$

then $PW_A$ will be a Hilbert space and $\widehat{\ }$ will define a Hilbert space isomorphism. With respect to this norm we have that for any $x \in \mathbb{R}$ and $F = \widehat{f} \in PW_A$,

$$|F(x)| = \left| \int_{-A}^{A} f(t) e^{-2\pi ixt}\, dt \right| \le \|f\|_2 \|e^{2\pi ixt}\|_2 = \sqrt{2A}\|F\|.$$

Thus, $PW_A$ is an RKHS on $\mathbb{R}$.

To compute the kernel function for $PW_A$, we note that since we are identifying $PW_A$ with the Hilbert space $L^2[-A,+A]$ via the map $\widehat{\ }$, we have that when $F = \widehat{f}$ and $G = \widehat{g}$, then

$$\langle F, G\rangle_{PW_A} = \langle f, g\rangle_{L^2} = \int_{-A}^{+A} f(t)\overline{g(t)}dt.$$

Now since

$$\langle F, k_y\rangle_{PW_A} = F(y) = \langle f, e^{2\pi iyt}\rangle_{L^2},$$

we obtain that

$$k_y(x) = \widehat{e^{2\pi iyt}}(x) = \int_{-A}^{+A} e^{2\pi i(y-x)t} dt.$$

Evaluating this integral yields

$$K(x,y) = \begin{cases} \dfrac{1}{\pi}\dfrac{\sin(2\pi A(x-y))}{x-y} & \text{if } x \neq y \\ 2A & \text{if } x = y \end{cases}.$$

Although this approach gives us the reproducing kernel for $PW_A$ in an efficient manner, we have not given an explicit formula for the inner product in $PW_A$. To obtain the inner product one needs to apply the Fourier inversion theorem and this is more detail than we wish to go into.

It is known that the Fourier transform of an $L^1$ function is a continuous function on $\mathbb{R}$. Since $L^2[-A,A] \subseteq L^1[-A,A]$, and is a closed subspace of $L^2(\mathbb{R})$ we see that $PW_A$ is a Hilbert space of continuous functions on $\mathbb{R}$. In