

Reproducing Kernels in Time Series Analysis

by

Michel St-Louis

A thesis submitted to
the Faculty of Graduate Studies and Postdoctoral Affairs
in partial fulfilment of
the requirements for the degree of
Master of Science

in

Probability and Statistics

Ottawa-Carleton Institute for
Mathematics and Statistics

Carleton University
Ottawa, Ontario, Canada

©2016, Michel St-Louis

Abstract

In this thesis, we provide a review of the theory of reproducing kernel Hilbert spaces, and a brief survey of its applications in the study of analytic functions spaces and in time series analysis. The principal aim of this thesis is to provide a theoretical basis for time series analysis that is independent of the commonly employed *stationary* hypothesis.

Contents

Acknowledgments	iii
1 Introduction	1
1.1 On the empirical nature of time series	1
1.2 On the mathematical theory of time series	1
1.3 Contents of the following chapters	2
2 Reproducing Kernel Hilbert Spaces	4
2.1 Historical overview	4
2.2 Definition and elementary properties	5
2.3 Finite dimensional spaces	15
2.4 Continuous linear operators on Hilbert spaces with reproducing kernels	17
2.5 Operations on reproducing kernels	19
3 Reproducing Kernels of Analytic Functions	24
3.1 Preliminary remarks	24
3.2 Bergman kernel	24
3.3 Bargmann-Fock kernel	32
4 Application to Time Series Analysis	37
4.1 Introduction	37
4.2 Hilbert Space representations of time series	37
4.3 Prediction and Smoothing using RKHS methods	42
4.4 Stationary time series using Hilbert space methods	47
4.5 Sampling theorems and reproducing kernels	50
5 Conclusion	53
References	54

Acknowledgments

I would like to thank my supervisor Dr. Mohamedou Ould Haye for his guidance and constant encouragement during the preparation of the thesis. Moreover, I express special thanks to Drs. Raluca Balan and Natalia Stepanova for serving on my graduate committee, and for having provided me with valuable commentary in regards to the content of this thesis.

Lastly, I would like to thank all those people that made the various digitization programs possible, such as the NUMDAM project sponsored by the French government. In a day where most academic papers are hidden behind expensive paywalls, it is a breath of fresh air to see some projects being devoted to making knowledge openly accessible to anyone.

1 Introduction

1.1 On the empirical nature of time series

A considerable portion of the investigations to which scientific researchers devote themselves can be thought, atleast in the abstract, to concern series of numerical observations ordered chronologically in time. Whether it be the meteorologist observing the daily high temperature's over a given year, the economist observing the monthly prices of a given commodity, or the physicist studying the stellar pulsation of a star for a given period, the ideal is, in the majority of cases, to establish a direct relationship between time and quantity. Although a direct relationship may be feasible for many types of quantities, it has been reasoned that often a relationship which includes some dependence on a random factor leads to a more realistic model of the phenomenon under investigation.¹ Yet, in the application of statistical methods to the explanation of natural phenomena, two immediate classifications become apparent; those phenomena whose statistical properties are invariant with respect to time, and those that vary over time. The former, which have been called *stationary*, have been the object of most developments within the theory of time series analysis, while the latter, that are appropriately called *non-stationary*, have been largely been left underdeveloped. It is the author's position, in regards to the epistemological value of our theoretical knowledge concerning time series, that the well-known observation of events in the material world exhibiting *non-stationary* properties merits, by itself, the construction of a consistent and complete theory of *non-stationary* time series. Though such a theory has yet to be fully developed, this thesis aims to clearly describe a number of first principles upon which it is assumed such a theory could be developed.

1.2 On the mathematical theory of time series

The mathematical theory of time series has its veritable origins in the elementary part of the theory of probability, itself being initially founded on the calculus of finite differences. As the mathematical methods being applied in the theory of probability gradually became more and more numerous, these methods eventually found their application in problems concerning time series, and have now become of such a common occurrence throughout the subject. Whether it be the introduction of Fourier Transforms to

¹For an intuitive explanation and justification of this assertion, consult the introduction of [31]

solve problems related to spectral analysis or limiting distributions, or the application of the theory of point sets to serve as the logical foundation for the notion of event and of observable classes of events, it is clear that the subject has continuously developed itself by using mathematical tools of an ever increasing degree of generality and power.

Having surveyed a considerable portion of the statistical literature concerning time series, the author concluded that the majority of the research being undertaken by the community at large rested a priori on the spectral representation theorem of Bochner-Khitchine, which is known to only be true provided that the time series is stationary. This assumption, which as outlined in the preceding paragraph is not obeyed by many natural phenomena, essentially reduces the consideration of the covariance function to the consideration of a function of a single variable. While this assumption, expressed in mathematical language, is incredibly simple, its necessity for all subsequent developments led the author to believe that new mathematical tools would have to be introduced to create a general theory of non-stationary time series.

In the 1950's and 1960's, various researchers began to develop a practical theory of non-stationary time series, by employing various mathematical concepts from the theory of dynamical systems, from the theory of stochastic and diffusion processes, and so forth. From this, many deep results were obtained, and it was Parzen, who was the first to arrive at the general idea from which a general theory of non-stationary time series analysis could be developed, primarily from the consideration of reproducing kernel Hilbert spaces.

1.3 Contents of the following chapters

In essence, the principal discovery of Parzen consisted in demonstrating that the various problems commonly considered in time series analysis, such as filtering, smoothing and prediction, have their natural setting in the context of reproducing kernel Hilbert spaces. This fundamental idea of Parzen has widely spread itself amongst the academic journals, to such a point now that the theory of reproducing kernel Hilbert spaces serves as a logical foundation for many subjects, such as the modern topic of Support Vector Machines.²

Thus, properly stated, the principal objective of this thesis is to describe, in rather complete detail, the theory of reproducing kernel Hilbert spaces, and its application to non-stationary time series analysis. Although certain

²The technical report written by Wahba [34] serves as a great reference clearly demonstrating the link between the two theories

details will necessarily have to be omitted, the primary focus will be placed on the exposition, and its logical order.

The organization of the thesis will be as follows.

In Chapter 2, we will study the theory of reproducing kernels in the context of abstract Hilbert spaces, by studying its most fundamental properties. Various operations on these spaces will be introduced, along with a number of propositions connected to them, both varying in nature.

In Chapter 3, we will study the particular reproducing kernels arising in certain analytic function spaces, namely in the Bergman space and in the Bargmann-Fock space. These developments will serve to demonstrate the great generality contained in the abstract theory that will be henceforth considered.

In Chapter 4, we will study the theory of reproducing kernel Hilbert spaces as applied to time series, which will naturally lead us to an interesting approach to time series analysis. General solutions to the problems of *prediction* and *smoothing* will be given, along with a clear presentation of the spectral representation theorem for stationary time series.

Though this thesis aims to be relatively self-contained, and to presuppose no particular knowledge of the subject at hand, it will be necessary to suppose the reader has some familiarity with the following subjects:

1. The theory of functional analysis, for which we refer to Conway [11], with particular importance to be placed on those chapters concerning the theory of Hilbert spaces.
2. The theory of functions of a complex variable, as it is exposed in any standard treatment of the subject, such as Conway [10].
3. The elementary parts of the modern theory of probability, as it is found in any well-known treatise of the subject, such as Billingsley [6].

It will be important to note that the author has no pretensions of originality in the following developments, and that this thesis claims to be no more than a didactical exposition of the theories contained herein.

2 Reproducing Kernel Hilbert Spaces

2.1 Historical overview

As with many other disciplines of mathematics, the theory of reproducing kernel Hilbert spaces has its origin in the theory of functions. For purposes of clarity, a brief overview of its historical evolution will be given, preceding the modern abstract treatment that will be subsequently described. The following presentation is inspired from that given in Saitoh [32].

The first investigations concerning functions having the reproducing property were undertaken by Mercer, in the context of Fredholm's integral equations [25]. The principal discovery of his research consisted in demonstrating that many curious propositions could be derived for the kernel of an integral equation, under the singular supposition that it have the reproducing property. Moore, in his highly influential *General Analysis* [26], considerably expanded upon the work of Mercer, and discovered many of the most fundamental theorems of the theory of RKHS.

Working in a surprisingly parallel direction, Salomon Bochner, in his study of Fourier Integrals [7], initiated the study of positive definite functions, and subsequently demonstrated the well-known Bochner-Khintchine theorem for functions of this type. The positive definite property ushered in by Bochner, which is of such importance in the theory of stationary time series, is a special case of the reproducing property.

Properly speaking, the first well-known occurrence of a reproducing kernel associated to a concrete Hilbert space was introduced by Bergman, in his study of holomorphic functions in \mathbb{C}^n , for which he later used in his study of the partial differential equations of mathematical physics [3].

It was at this point, that Aronszajn, following the thoughts of Moore in his seminal paper [2], saw that all the studies before him could be subsumed by the following general investigation: to any abstract Hilbert subspace of the space of all functions defined on an arbitrary set can be associated a *reproducing* kernel; and by introducing certain postulates, one can study the relations between the kernel and the space, and introduce both the arithmetical operations and the order structure.

This formulation of the theory of reproducing kernel Hilbert spaces will be that which is given, principally for reasons of both generality and simplicity. It is a remarkable fact, that a subject beginning in the theory of integral equations, eventually found its application in probability theory and time series analysis, in, for instance, the works of Loève [24] and Parzen [30].

Almost all the results contained in the following subsections are found in

Aronszjan's original papers [1] and [2], while the treatment given is heavily inspired by Saitoh [32], Berliet [5], and Istratescu [20].

2.2 Definition and elementary properties

In this chapter, we will define the concept of a reproducing kernel Hilbert space and state their principal properties, along with complete demonstrations.

Before proceeding to define the general notion of a reproducing kernel of an abstract Hilbert space, it will first be demonstrated how this notion may be generalized from the well-known theory of Fourier Transforms.

In essence, the theory of Fourier transforms of a real valued function $f(x)$ consists in the study of the function $\phi(y)$, defined by the integral below, assuming of course that f is integrable in the sense of Lebesgue :

$$\phi(y) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi ixy} dx$$

Two generalizations of the equation above are immediate, namely by letting $f(x)$ take values in an arbitrary set X , and replacing the exponential term by an arbitrary function $K(x, y)$. Doing so, we derive the following equation:

$$\phi(y) = \int_X f(x)K(x, y)dx$$

By replacing the characteristic ϕ of the function $\phi(y)$ in the equation above by the letter f , we would be lead to the study of a rather interesting equation, having the fundamental reproducing property, viz.

$$f(y) = \int_X f(x)K(x, y)dx$$

Generalizing even further, we may assume the functions $f(x)$ belong to a class of functions H forming a Hilbert space, and replace the symbol of integration by the corresponding inner product, that is

$$f(y) = \langle f(x), K(x, y) \rangle_H$$

Thus, by continuously generalizing a simple Fourier transform, we are lead to the study of the equation above, which possesses the so-called *reproducing* property that is of primary interest to us. The first question that needs to be adressed is clearly the following: under what condition imposed on $K(x, y)$ will the preceding equation hold for all functions $f(x)$ in H ?

An conclusive answer to this question will given by Theorem 2.2.4., and the required condition will be demonstrated to be that of *positive definiteness*, which we shall now define.

Definition 2.2.1. A complex valued function $K(x, y)$ defined on the Cartesian product $X \times X$ of an arbitrary set X is said to be positive definite provided that, for any finite set of points $\{x_j\}_{j=1}^n$ of X and any corresponding complex numbers $\{\xi_j\}_{j=1}^n$,

$$\sum_{j=1}^n \sum_{k=1}^n \xi_j \overline{\xi_k} K(x_j, x_k) \geq 0.$$

To anyone who has extensively studied the theory of matrices, the definition given above is clearly a generalization of that given for a positive semi-definite matrix. In our case, we employ the term *positive definite*, as opposed to *positive semi-definite* for reasons of brevity.

Before presenting the relation between positive definite functions and reproducing kernel Hilbert spaces, it will be necessary to state a few definitions and some elementary propositions.

Definition 2.2.2. Let H denote a Hilbert space of functions $f(x)$ defined on a set X , with its inner product denoted by $\langle f, g \rangle_H$ and norm $\|f\|_H$. A complex valued function $K(x, y)$ defined on $X \times X$ is said to be a *reproducing kernel* of H if the following conditions hold:

1. For any $y \in X$, $K_y(x) = K(x, y)$ is an element of H , regarded as a function of x .
2. For any $y \in X$ and for any $f \in H$, $f(y) = \langle f(x), K(x, y) \rangle_H$.

Definition 2.2.3. A Hilbert space H on functions on a set X is called a reproducing kernel Hilbert space if there exists a reproducing kernel $K(x, y)$ of H .

Theorem 2.2.1. If $K(x, y)$ is a reproducing kernel on the Hilbert space H , then the following conditions hold:

1. $K(x, x) \geq 0$
2. $K(x, y) = \overline{K(y, x)}$
3. $|K(x, y)|^2 \leq K(x, x)K(y, y)$

Proof. The first two assertions are immediate consequences of the properties of $K(x, y)$, viz.

$$\begin{aligned} K(x, x) &= \langle K(x, x), K(x, x) \rangle_H \geq 0 \\ K(x, y) &= \langle K(x, x), K(x, y) \rangle_H = \overline{\langle K(x, y), K(x, x) \rangle_H} = \overline{K(y, x)}. \end{aligned}$$

While the third is an application of the Cauchy-Schwarz inequality, i.e.

$$\begin{aligned} |K(x, y)|^2 &= |\langle K(x, x), K(x, y) \rangle_H|^2 \\ &\leq \langle K(x, x), K(x, x) \rangle_H \langle K(y, y), K(y, y) \rangle_H \\ &= K(x, x)K(y, y). \end{aligned}$$

Theorem 2.2.2. The reproducing kernel $K(x, y)$ associated to a Hilbert space H , assuming it exists, is unique.

Proof. Assume $K^*(x, y)$ is another reproducing kernel of H . Consider, for all $y \in X$, the square of the norm of their difference in x , viz.

$$\begin{aligned} \|K(x, \cdot) - K^*(x, \cdot)\|_H^2 &= \langle K(x, \cdot) - K^*(x, \cdot), K(x, \cdot) - K^*(x, \cdot) \rangle_H \\ &= \langle K(x, \cdot) - K^*(x, \cdot), K(x, \cdot) \rangle_H \\ &\quad - \langle K(x, \cdot) - K^*(x, \cdot), K^*(x, \cdot) \rangle_H \\ &= \{K(x, \cdot) - K^*(x, \cdot)\} - \{K(x, \cdot) - K^*(x, \cdot)\} = 0, \end{aligned}$$

where the last equality is derived from the assumption that both of the functions $K(x, y)$ and $K^*(x, y)$ are reproducing kernels. Therefore, it necessarily follows that $K(x, y) = K^*(x, y)$. \square

Theorem 2.2.3. Any reproducing kernel $K(x, y)$ on a Hilbert space H is a positive definite function, cf. Definition 2.2.1.

Proof. Let $\{x_j\}_{j=1}^n$ be an arbitrary finite set of points in X , and $\{\xi_j\}_{j=1}^n$ be complex numbers. The results easily follows by considering the appropriate squared norm of the following quadratic form, viz.

$$\begin{aligned} \left\| \sum_{j=1}^n \xi_j K(x, x_j) \right\|_H^2 &= \sum_{j=1}^n \sum_{k=1}^n \xi_j \bar{\xi}_k \langle K(x, x_j), K(x, x_k) \rangle_H \\ &= \sum_{j=1}^n \sum_{k=1}^n \xi_j \bar{\xi}_k K(x_j, x_k)_H. \end{aligned}$$

\square

The fundamental result of the theory of reproducing kernel Hilbert spaces, demonstrated by both Moore [26] and Aronszajn [1], is the following theorem, showing that any positive definite function generates a reproducing kernel Hilbert space.

Theorem 2.2.4. To any positive definite function $K(x, y)$, there corresponds a unique Hilbert space H having the reproducing kernel $K(x, y)$. The subspace H_* of H spanned by the set of functions $\{K(\cdot, y) \mid y \in X\}$ is dense in H and H is the completion of H_* with respect to the inner product

$$\langle f, g \rangle_{H_*} = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \overline{\beta_k} K(y_j, y_k)$$

where $f(x) = \sum_{j=1}^n \alpha_j K(x, y_j)$ and $g(x) = \sum_{k=1}^n \overline{\beta_k} K(x, y_k)$.

Proof. The fundamental idea in the proof given below consists in forming a class of functions H_* obtained by certain linear combinations of $K(x, y)$, and finally demonstrating that the closure of this class of functions, denoted by H , forms a Hilbert space.

Formally speaking, define the class H_* to consist of all functions $f(x)$ defined on E , such that for a finite set of points $\{y_j\}_{j=1}^n$ taken in X and a set of complex numbers $\{\alpha_j\}_{j=1}^n$,

$$f(x) = \alpha_1 K(x, y_1) + \alpha_2 K(x, y_2) + \dots + \alpha_n K(x, y_n).$$

For any function $f(x) \in H$, the representation must exist but is not necessarily unique, and in general, there exists many different representations for the same function.

Let $f(x) = \sum_{j=1}^n \alpha_j K(x, y_j)$ and $g(x) = \sum_{k=1}^n \overline{\beta_k} K(x, y_k)$ be arbitrary elements taken in H_* , and define a function $\langle f, g \rangle_{H_*}$ by the formula

$$\langle f, g \rangle_{H_*} = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \overline{\beta_k} K(y_j, y_k).$$

Since $\langle f, g \rangle_{H_*} = \sum_{j=1}^n \alpha_j g(y_j) = \sum_{k=1}^n \overline{\beta_k} f(y_k)$, it is plain that the function $\langle f, g \rangle_{H_*}$ is independent of their particular given representations.

Moreover, it is clear that the function $\langle f, g \rangle_{H_*}$ is an inner product, having the three requisite properties

1. $\langle f, g \rangle_{H_*} = \overline{\langle g, f \rangle_{H_*}}$
2. $\langle f + g, h \rangle_{H_*} = \langle f, h \rangle_{H_*} + \langle g, h \rangle_{H_*}$
3. $\langle \alpha f, h \rangle_{H_*} = \alpha \langle f, h \rangle_{H_*}$

For any $f(x) \in H_*$, we will also consider the Hermitian form defined by the inner product, which is necessarily positive definite:

$$\langle f(x), f(x) \rangle_{H_*} = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \overline{\alpha_k} K(y_j, y_k) \geq 0$$

By setting $\|f\|_{H_*} = \sqrt{\langle f, f \rangle_{H_*}}$, it is plain that $\|f\|_{H_*}$ is a semi-norm with the following three properties

1. $|\langle f, g \rangle_{H_*}| \leq \|f\|_{H_*} \|g\|_{H_*}$
2. $\|f + g\|_{H_*} \leq \|f\|_{H_*} + \|g\|_{H_*}$
3. $\|\alpha f\|_{H_*} = |\alpha| \|f\|_{H_*}$

By the definition of H_* , it is immediate that, for any $y \in X$, the function $K(x, y)$ belongs to H_* , and satisfies, for all $f(x) \in H_*$,

$$\begin{aligned} \|K(x, \cdot)\|_{H_*}^2 &= K(x, x) \\ \langle f(x), K(x, y) \rangle_{H_*} &= f(y) \end{aligned}$$

Whence, from applying the first property of $\|f\|_{H_*}$, we obtain

$$|f(x)| \leq K(x, x) \|f(x)\|_{H_*}$$

From the equation above, it immediately follows that the only zero element of H_* is the zero function, therefore $\|f\|_{H_*}$ is a norm, generated by the inner product $\langle f, g \rangle_{H_*}$.

By considering the completion H of H_* by adding all limits of Cauchy sequences of H_* , it will be shown that H forms a Hilbert space. Let us first remark that for any Cauchy sequence $\{f_k(y)\}_{k=1}^\infty$ of elements of H_* , the limit exists, for

$$\begin{aligned} |f_m(y) - f_n(y)| &= |\langle f_m(x), K(x, y) \rangle_{H_*} - \langle f_n(x), K(x, y) \rangle_{H_*}| \\ &= |\langle f_m(x) - f_n(x), K(x, y) \rangle_{H_*}| \\ &\leq \|f_m(x) - f_n(x)\|_{H_*} K(y, y)^{1/2} \end{aligned}$$

Thus, $\lim_{k \rightarrow \infty} f_k(x) = f(x)$ necessarily exists, and thus we can define the norm and inner product in H as such

$$\begin{aligned} \|f(x)\|_H &= \lim_{n \rightarrow \infty} \|f_n(x)\|_{H_*} \\ \langle f(x), g(x) \rangle_H &= \lim_{n, m \rightarrow \infty} \langle f_n(x), g_m(x) \rangle_{H_*} \end{aligned}$$

To show that H is complete, let $(f_k)_{k=1}^\infty$ be a Cauchy sequence of functions in H_* converging pointwise to $f \in H$. For $\lim_{k \rightarrow \infty} \langle f_k, f_k \rangle_{H_*} = \langle f, f \rangle_H$, we deduce that

$$\begin{aligned} \lim_{h \rightarrow \infty} \langle f_k, f_h \rangle_{H_*} &= \langle f_k, f \rangle_H \\ \lim_{k \rightarrow \infty} \lim_{k \rightarrow \infty} \langle f_k, f_k \rangle_{H_*} &= \langle f, f \rangle_H \end{aligned}$$

And, because

$$\|f - f_k\|_H^2 = \langle f, f \rangle_H - \langle f, f_k \rangle_H - \langle f_k, f \rangle_H + \langle f_k, f_k \rangle_H$$

it follows that

$$\lim_{k \rightarrow \infty} \|f - f_k\|_H = 0$$

Hence, $\{f_k(y)\}_{k=1}^\infty$ converges in norm to f , and thus H_* is dense in H , H forms a Hilbert space with reproducing kernel $K(x, y)$, and the unicity is a consequence of the following theorem. \square

Theorem 2.2.5. Let H_1 and H_2 be two Hilbert spaces of functions defined on a set X , admitting the reproducing kernels $K_1(x, y)$ and $K_2(x, y)$. If $K_1(x, y) = K_2(x, y)$ for all $x, y \in X$, then $H_1 = H_2$ and $\|f\|_1 = \|f\|_2$ for every function f .

Proof. Let H_*^i , for $i \in \{1, 2\}$, denote the class of all functions of H_i that are linear combinations of $K_x(y) = K(x, y)$. By the result above, H_*^i is dense in H_i . For any function $f \in H_*^i$, it is clear that, since $K_1(x, y) = K_2(x, y)$,

$$f(x) = \sum_{j=1}^n \alpha_j K_1(x, y_j) = \sum_{j=1}^n \alpha_j K_2(x, y_j)$$

Hence, its symbolic representation is independent of whether we consider it as an element of H_*^1 or H_*^2 , thus $H_*^1 = H_*^2$.

Moreover, for any function f ,

$$\|f\|_{1,*} = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \overline{\alpha_k} K_1(y_j, y_k) = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \overline{\alpha_k} K_2(y_j, y_k) = \|f\|_{2,*}$$

Now, if $f \in H_1$, there exists, by the discussions given above, a sequence of functions $\{f_k\}_{k=1}^\infty$ in H_*^1 such that $\lim_{k \rightarrow \infty} \|f - f_k\|_{1,*} = 0$. Since the norms are equal, $\{f_k\}_{k=1}^\infty$ is also a Cauchy sequence in H_*^2 , and thus there exists a $g \in H_2$ such that $\lim_{k \rightarrow \infty} \|g - f_k\| = 0$. But, clearly, $f = \lim_{k \rightarrow \infty} f_k = g$, and hence $H_1 \subset H_2$. By the same argument, we can demonstrate $H_2 \subset H_1$, hence $H_1 = H_2$.

Finally, since H_*^1 and H_*^2 are dense subspaces of H_1 and H_2 , respectively, and $\|f\|_{1,*} = \|f\|_{2,*}$, we deduce that $\|f\|_1 = \|f\|_2$. \square

For our purposes, the Hilbert space under consideration always consists of functions defined on an arbitrary set. Hence, a natural question arises: Under what conditions imposed on the functions themselves will there exist a reproducing kernel for the corresponding Hilbert space? The answer to this question is found in the following theorem.

Theorem 2.2.6. For any Hilbert space H whose elements are functions $f(x)$ on X , a necessary and sufficient condition that H have a reproducing kernel $K(x, y)$ is that for every $y \in X$, the linear operator L_y defined by

$$L_y\{f\} = f(y)$$

be bounded, i.e.

$$|L_y\{f\}| \leq M \|f\|_H$$

for some positive constant M and for all $f \in H$.

Proof. Assume first that the reproducing kernel $K(x, y)$ exists. Then, by definition, for any $y \in X$, the following formula holds true

$$L_y\{f\} = f(y) = \langle f(x), K(x, y) \rangle_H.$$

By a straightforward application of the Cauchy-Schwarz inequality, we derive that

$$\begin{aligned} |f(y)|^2 &\leq \langle f(x), f(x) \rangle_H \langle K(x, y) K(x, y) \rangle_H \\ &= \|f\|_H^2 K(y, y). \end{aligned}$$

Hence, by taking $M = \sqrt{K(y, y)}$, the result clearly follows

Now, if $L_y\{f\} = f(y)$ is a bounded linear operator on X , then, by the well-known Riesz representation theorem, there exists a function $h(x, \cdot)$ belonging to H such that

$$f(y) = \langle f(x), h(x, y) \rangle_H.$$

Hence, by setting $K(x, y) = h(x, y)$, the results follows. \square

Theorem 2.2.7. If a Hilbert space H of functions $f(x)$ defined on X admits a reproducible kernel $K(x, y)$, then any closed linear subspace M of H admits a reproducible kernel $M(x, y)$, and if M^\perp denotes the orthogonal complement of M with reproducing kernel $M^\perp(x, y)$, then

$$K(x, y) = M(x, y) + M^\perp(x, y).$$

Proof. The first assertion made in the theorem is an obvious consequence of Theorem 2.2.6, for any functional $L_y\{f\} = f(y)$ defined on H does not cease to be continuous when f is restricted to a subspace of H .

To prove the second assertion, let $f(x) \in H$ have the orthogonal decomposition $f(x) = f_M(x) + f_{M^\perp}(x)$. Then,

$$\begin{aligned} \langle f(x), M(x, y) + M^\perp(x, y) \rangle_H &= \langle f_M(x) + f_{M^\perp}(x), M(x, y) + M^\perp(x, y) \rangle_H \\ &= \langle f_M(x), M(x, y) \rangle_H + \langle f_M(x), M^\perp(x, y) \rangle_H \\ &\quad + \langle f_{M^\perp}(x), M(x, y) \rangle_H + \langle f_{M^\perp}(x), M^\perp(x, y) \rangle_H \\ &= \langle f_M(x), M(x, y) \rangle_H + \langle f_{M^\perp}(x), M^\perp(x, y) \rangle_H \\ &= f_M(y) + f_{M^\perp}(y) \\ &= f(y). \end{aligned}$$

Hence, $M(x, y) + M^\perp(x, y)$ is a reproducing kernel for the Hilbert space H , and by unicity, it follows that $K(x, y) = M(x, y) + M^\perp(x, y)$. \square

Theorem 2.2.8. Let M be a closed linear subspace of a Hilbert space H . Then, if M admits a reproducing kernel $M(x, y)$, the orthogonal projection P_M of H onto M is obtained by the formula:

$$f_M(y) = P_M\{f(x)\} = \langle f(x), M(x, y) \rangle$$

Proof. Let f be an arbitrary function in H , and let $f = f_M + f_{M^\perp}$ be its orthogonal decomposition with respect to M , such $f_M \in M$ and $f_{M^\perp} \in M^\perp$. Since f_{M^\perp} is orthogonal to $M(x, y)$ considered as a function of x , we obtain

$$\langle f, M(x, y) \rangle = \langle f_M + f_{M^\perp}, M(x, y) \rangle = \langle f_M, M(x, y) \rangle = f_M(y).$$

\square

The following theorem is taken from Fortet [14], and is of great utility in both determining if a function belongs to a given reproducing kernel Hilbert space and in the approximative computation of the norm of a function in a reproducing kernel Hilbert space.

Theorem 2.2.9. Let H be a Hilbert space of functions defined on a set X admitting a reproducing kernel $K(x, y)$. Let $f(x)$ be an arbitrary function on X . Then, for $f(x)$ to belong to H , a necessary and sufficient condition is that the ratio:

$$\frac{\|\sum_{h=1}^n \alpha_h f(x_h)\|^2}{\sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l K(x_k, x_l)} = \frac{\sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l \overline{f(x_k)} f(x_l)}{\sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l K(x_k, x_l)}$$

be always bounded above, for any values of n, α_k or x_h .

If this condition holds, then, for any $f(x) \in H$, we have

$$\sup_{n, x_h, \alpha_h} \frac{\|\sum_{h=1}^n \alpha_h f(x_h)\|_H^2}{\sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l K(x_k, x_l)} = \|f\|_H^2.$$

Proof. To prove the necessity, let $f(x)$ be an arbitrary function in H , and consider the expression,

$$\begin{aligned} \sum_{h=1}^n \alpha_h \overline{f(x_h)} &= \sum_{h=1}^n \alpha_h \overline{\langle f(x_h), K(x, x_h) \rangle_H} \\ &= \sum_{h=1}^n \alpha_h \langle K(x, x_h), f(x_h) \rangle_H \\ &= \left\langle \sum_{h=1}^n \alpha_h K(x, x_h), f(x) \right\rangle_H \end{aligned}$$

Define $\phi(x) = \sum_{h=1}^n \alpha_h K(x, x_h) \in H$, which satisfies

$$\|\phi(x)\|_H^2 = \sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l K(x_k, x_l)$$

Applying the Cauchy-Schwarz inequality, we derive that

$$\left\| \sum_{h=1}^n \alpha_h \overline{f(x_h)} \right\|_H^2 = \|\langle \phi, f \rangle\|_H^2 \leq \|f\|_H^2 \|\phi\|_H^2.$$

Hence, by formally dividing the latter two terms by $\|\phi\|_H^2$ and choosing n, t_h, α_h such that $\|f - \phi\|_H^2$ is smaller than any prescribed number, we obtain the result

$$\sup_{n, x_h, \alpha_h} \frac{\|\sum_{h=1}^n \alpha_h f(x_h)\|_H^2}{\sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l K(x_k, x_l)} = \frac{\sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l \overline{f(x_k)} f(x_l)}{\sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l K(x_k, x_l)} = \|f\|_H^2.$$

To prove the sufficiency, let $f(x) \in H$, such that there exists a positive constant M such that

$$\begin{aligned} \left\| \sum_{h=1}^n \alpha_h \overline{f(x_h)} \right\|_H^2 &= \sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l \overline{f(x_k)} f(x_l) \\ &\leq M \sum_{k=1}^n \sum_{l=1}^n \alpha_k \bar{\alpha}_l K(x_k, x_l) \end{aligned}$$

The case where $M = 0$ is trivial, hence we suppose M is necessarily positive.

For any function $\phi(x) = \sum_{j=1}^n \alpha_j K(x, y_j) \in H_*$, as defined in Theorem 2.2.5, define the complex-valued linear operator $T\{\phi(x)\}$ by the formula

$$T\{\phi(x)\} = \sum_{h=1}^n \alpha_h \overline{f(x_h)}$$

Yet, by applying the hypothesis, we observe that

$$|T\{\phi(x)\}|^2 \leq M \sum_{k=1}^n \sum_{l=1}^n \alpha_k \overline{\alpha_l} K(x_k, x_l) \leq M \|\phi\|_{H_*}^2 .$$

Thus, for T is a continuous linear functional on H_* , there exists a unique operator W on H , whose restriction on H_* is T , and by the Riesz representation theorem, there exists an element $\theta(x) \in H$ such that

$$W\{\phi(x)\} = \langle \phi(x), \theta(x) \rangle_H$$

If we take $\phi(x) = K(x, y)$ and substitute it into the equation above, we obtain

$$\overline{f(x)} = \langle K(x, y), \theta(x) \rangle_H = \overline{\langle \theta(x), K(x, y) \rangle_H} = \overline{\theta(x)}$$

From which $f(x) = \theta(x)$ necessarily belongs to H . □

The following theorem will be useful in the formulation of the Bergman kernel discussed in the next chapter.

Theorem 2.2.10. Let H be a reproducing kernel Hilbert space of functions $f(x)$ on X with kernel $K(x, y)$. If $\{\phi_s : s \in S\}$ is an orthonormal basis for H , then the reproducing kernel will have the pointwise convergent representation

$$K(x, y) = \sum_{s \in S} \phi_s(x) \overline{\phi_s(y)}$$

Proof. Since $\{\phi_s : s \in S\}$ is an orthonormal basis for H , $K(x, y)$ will have the following representation, convergent in the norm:

$$K(x, y) = \sum_{s \in S} \alpha_s(y) \phi_s(x)$$

where

$$\alpha_s(y) = \langle K(x, y), \phi_s(x) \rangle_H = \overline{\langle \phi_s(x), K(x, y) \rangle_H} = \overline{\phi_s(y)}$$

Thus, we obtain

$$K(x, y) = \sum_{s \in S} \phi_s(x) \overline{\phi_s(y)}$$

which necessarily is pointwise convergent for it converges in norm. \square

2.3 Finite dimensional spaces

Amongst all those Hilbert spaces that may be considered, those that are of noteworthy interest, in view of either their applications or simplicity, are those that are *finite-dimensional*.

Definition 2.3.1. A Hilbert space H of complex-valued functions defined on a set X is said to be n dimensional if there exists a set $e_1(x), e_2(x), \dots, e_n(x)$ of n linearly independent functions of H having the property that to any function $f(x)$ of H there corresponds a unique set $\{c_i\}_{i=1}^n$ of complex numbers such that

$$f(x) = \sum_{i=1}^n c_i e_i(x)$$

Thus, in a finite-dimensional Hilbert space H , any element $f(x) \in H$ may be uniquely expressed as a linear combination of the basis elements $e_1(x), e_2(x), \dots, e_n(x)$.

Throughout the rest of this chapter, it will always be assumed that the Hilbert space H under consideration is finite-dimensional, and admits the elements e_1, e_2, \dots, e_n as a basis.

By putting $\varepsilon_{ij} = \langle e_i, e_j \rangle_H$, we can define the norm and inner product of any two functions $f(x)$ and $g(x)$ in H as follows:

Definition 2.3.2. The inner product of any two functions $f(x) = \sum_{i=1}^n c_i e_i(x)$ and $g(x) = \sum_{j=1}^n d_j e_j(x)$ in H is defined by the formula

$$\langle f, g \rangle_H = \sum_{i=1}^n \sum_{j=1}^n \varepsilon_{ij} c_i \overline{d_j}$$

Definition 2.3.3. The norm of any function $f(x) = \sum_{i=1}^n c_i e_i(x)$ belonging to H will be defined by the general positive definite norm Hermitian form

$$\|f\|_H^2 = \sum_{i=1}^n \sum_{j=1}^n \varepsilon_{ij} c_i \overline{c_j}$$

Theorem 2.3.1. If $E = (\varepsilon_{ij})$ denotes the Gram matrix associated to the basis e_1, e_2, \dots, e_n and $F = (f_{ij})$ denotes the conjugate transpose of its inverse, then the function

$$K(x, y) = \sum_{j=1}^n \sum_{k=1}^n f_{jk} e_j(x) \overline{e_k(y)}$$

is the reproducing kernel for the space H with the inner product as defined in Theorem 2.2.4.

Proof. Since E is a Hermitian matrix, it is a well-known result that its inverse $E^{-1} = (\varepsilon_{ij}^{-1})$ is also Hermitian. Therefore, it follows from the definition of F that $\overline{f_{ji}} = \varepsilon_{ij}^{-1}$, and thus

$$\sum_{i=1}^n \varepsilon_{ij} \overline{f_{jk}} = \delta_{ik}.$$

It is plain that it satisfies the first property of a reproducing kernel, for any term in the expansion of the double summation will be a linear multiple of one of the basis elements e_1, e_2, \dots, e_n .

To show it satisfies the reproducing property, we note that

$$\begin{aligned} \langle f, K(x, y) \rangle_H &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \langle c_i e_i(x), f_{jk} e_j(x) \overline{e_k(y)} \rangle_H \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n c_i e_k(y) \overline{f_{jk}} \langle e_i(x), e_j(x) \rangle_H \\ &= \sum_{i=1}^n \sum_{k=1}^n \delta_{ik} c_i e_k(y) \\ &= \sum_{i=1}^n c_i e_i(y) = f(y). \end{aligned}$$

□

Thus, from the discussions given above, we immediately deduce the following theorem, due to Saitoh [32], which completely characterizes finite-dimensional reproducing kernel Hilbert spaces.

Theorem 2.3.2. A function $K(x, y)$ is the reproducing kernel of a finite-dimensional Hilbert space of functions $f(x)$ admitting the basis functions $e_1(x), e_2(x), \dots, e_n(x)$ if and only if it is of the form

$$K(x, y) = \sum_{j=1}^n \sum_{k=1}^n f_{jk} e_j(x) \overline{e_k(y)},$$

where the matrix $F = (f_{jk})$ is positive definite.

2.4 Continuous linear operators on Hilbert spaces with reproducing kernels

The study of operators in Hilbert space is a well-known subject that has been, and continues to be, extensively studied, and whose treatment can be found in any standard textbook of functional analysis.

The object of this chapter will be to demonstrate the surprising fact that the introduction of a reproducible kernel within the study of continuous linear operators defined on a Hilbert space leads to a multitude of interesting results.

Let H be a Hilbert space of complex-valued functions $f(x)$ defined on a set X admitting a reproducing kernel $K(x, y)$ defined on $X \times X$. If $F : H \rightarrow \mathbb{C}$ is a continuous linear functional, then, by the well-known Riesz representation theorem, there exists $g(x) \in H$ such that

$$F\{f\} = \langle f(x), g(x) \rangle_H$$

The following two theorems will demonstrate how $g(x)$ and its appropriate generalization may be explicitly computed by using the reproducing kernel $K(x, y)$.

Theorem 2.4.1. If $F : H \rightarrow \mathbb{C}$ and $g(x) \in H$ are as defined above, then

$$g(y) = \overline{F\{K(x, y)\}}.$$

Proof. The proof essentially is an application of the second property from Theorem 2.2.1, i.e.

$$g(y) = \langle g(x), K(x, y) \rangle_H = \overline{\langle K(x, y), g(x) \rangle_H} = \overline{F\{K(x, y)\}}$$

□

Theorem 2.4.2. Let H be a Hilbert space of functions $f(x)$ with reproducing kernel $K(x, y)$. If $T : H \rightarrow H$ is a bounded linear functional, then it will have the following representation:

$$(Tf)(y) = \langle f(x), G_y \rangle_H$$

where $G_y = T^*K(x, y)$ denotes the adjoint of the operator T .

Proof. The proof is essentially an application of the reproducing property, i.e.

$$(Tf)(y) = \langle (Tf)(x), K(x, y) \rangle_H = \langle f(x), T^*K(x, y) \rangle_H = \langle f(x), G_y \rangle_H$$

□

Throughout the rest of this chapter, the subscript x in the symbolic expression $T_x\{K(x, y)\}_u$ will designate that the operator T is operating on the function $K(x, y)$ as a function of x , while the subscript u will designate that the function corresponding to $T_x\{K(x, y)\}$ is evaluated at u .

The following very interesting theorem of Chalmers [9] expresses the reproducing kernel of one Hilbert space in terms of the reproducing kernel of another Hilbert space, by means of a linear operator defined on one of them and taking values in the other.

Theorem 2.4.3. Let H_1 and H_2 be two Hilbert spaces of functions defined on the sets X and Y admitting the reproducing kernels $K_1(x, y)$ and $K_2(u, v)$, respectively. If $T : H_1 \rightarrow H_2$ is a bounded linear operator such that the range of T is dense in H_2 , then the reproducing kernel K_2 is obtained from K_1 by the following formula:

$$K_2(u, v) = (T^*)_y^{-1}(\overline{(T_x K(x, y))_u})_v$$

Proof. For a fixed $v \in Y$ and $f \in H_1$, $T\{f\}(v)$ is a continuous linear operator H_2 . Therefore, by the Riesz representation theorem, there exists an $\phi_n \in H_1$ such that

$$T\{f\}(v) = \langle f(x), \phi_n \rangle_{H_1}$$

Moreover, we observe that

$$T_x\{K_1(x, y)\}_v = \langle K_1(x, y), \phi_n(y) \rangle_{H_1} = \overline{\langle \phi_n(y), K_1(x, y) \rangle_{H_1}} = \overline{\phi_n(y)}$$

And, since H_2 has a reproducing kernel $K_2(u, v)$, we derive that

$$\begin{aligned} T\{f\}(v) &= \langle T\{f\}, K_2(\cdot, v) \rangle_{H_2} \\ &= \langle T\{f\}(u), K_2(u, v) \rangle_{H_2} \\ &= \langle f(x), T_v^*(K_2(u, v))_y \rangle_{H_1} \end{aligned}$$

From the two equations above, it is clear that

$$T_v^*(K_2(u, v))_y = \overline{T_x K(x, y)}_u$$

Now, since the range of T is dense in H_2 , then $(T^*)^{-1} : H_1 \rightarrow H_2$ clearly exists, and by a formally multiplying the equation above, and substituting the indices appropriately, we obtain the required results

$$K_2(u, v) = (T^*)_y^{-1}(\overline{T_x K(x, y)}_u)_v$$

□

2.5 Operations on reproducing kernels

In many applications of the theory of reproducing kernel Hilbert spaces, it becomes necessary to consider various algebraic or set theoretic operations on reproducing kernels, and to determine the Hilbert space to which they correspond. These problems form the object of this chapter.

Consider a Hilbert space H of functions $f(x)$ defined on X equipped with a reproducing kernel $K(x, y)$ defined on $X \times X$, which necessarily must be a positive definite function. If $K(x, y)$ is restricted to the set $X_1 \times X_1$, $X_1 \subset X$, K will clearly remain positive definite, and thus, by Theorem 2.2.4., corresponds to a class H_1 of functions defined on X_1 with norm $\|\cdot\|_{H_1}$. The following theorem provides the explicit determination of H_1 .

Theorem 2.5.1. Let $K(x, y)$ is the reproducing kernel of the Hilbert space H of functions $f(x)$ defined on X with norm $\|\cdot\|_H$, and X_1 be a subset of X . Then the restriction $K_1(x, y)$ of $K(x, y)$ to $X_1 \times X_1$ is the reproducing kernel of the class H_1 of all restrictions of functions $f(x) \in H$ to the subset X_1 . For any $f_1(x) \in H_1$, the norm $\|f_1\|_{H_1}$ is defined by

$$\|f_1\|_{H_1} = \min_{\phi(x) \in H} \{ \|f(x)\|_H : f_1(x) = \phi(x)|_{X_1} \}$$

Proof. Define, for any function $f(x) \in H$, an operator $T\{f(x)\} = f(x)|_{X_1}$ and let $N = \{f(x) \in H \mid f(x)|_{X_1} = 0\}$ be its associated kernel. Clearly, the operator T is both linear and surjective.

To show that N is closed, let us note that, if $\{f_k\}_{k=1}^\infty$ is a sequence of functions in N converging to $f \in H$, it necessarily follows that f satisfies $f|_{X_1} = 0$, for $f_k(x)|_{X_1} = 0$ for all k . Thus, N is closed.

Let N^\perp denote the orthogonal complement of N in H , and W denote the restriction of T to N^\perp , i.e. $W\{f(x)\} = T\{f(x)\}|_{N^\perp}$. Since $N \cap N^\perp = \{0\}$,

the kernel of W contains only the zero element, and thus W is an injective linear operator. Thus, one can define an inner product for any two $f, g \in H_1$ as

$$\langle f, g \rangle_{H_1} = \langle W^{-1}\{f\}, W^{-1}\{g\} \rangle_H$$

H_1 , considered together with the inner product above, becomes a Hilbert space with norm $\|f\|_{H_1} = \|W^{-1}f\|_H$. To show that $K_1(x, y)$ is the reproducing kernel of H_1 , we note that, since $K(x, \cdot)$ belongs to H , then $K_1(x, \cdot) = K(x, \cdot)|_{X_1 \times X_1}$ necessarily belongs to H_1 . And, without loss of generality, we may put $W\{K(x, y)\} = K_1(x, y)$, so that, for any $f(x) \in H_1$, we have

$$\begin{aligned} \langle f(x), K_1(x, y) \rangle_{H_1} &= \langle W^{-1}\{f(x)\}, W^{-1}\{K(x, y)\} \rangle_H \\ &= \langle W^{-1}\{f(x)\}, K(x, y) \rangle_H \\ &= W^{-1}\{f(y)\} \\ &= f(y). \end{aligned}$$

Thus, $K_1(x, y)$ is the reproducing kernel of H_1 .

To prove the second assertion contained in the theorem, let $\phi(x)$ be an arbitrary element of H with $\phi(x)|_{X_1} = f_1(x)$. Now, since

$$W\{\phi(x) - W^{-1}\{f_1(x)\}\} = \phi(x)|_{X_1} - f_1(x) = 0$$

The function $\phi(x) - W^{-1}\{f_1(x)\}$ belongs to N , and, by definition, $W^{-1}\{f_1(x)\}$ belongs to N^\perp . Thus, we immediately obtain

$$\|\phi\|_H^2 = \left\| \phi - W^{-1}\{f_1\} \right\|_H^2 + \left\| W^{-1}\{f_1\} \right\|_H^2$$

Hence, we deduce the required fact

$$\|f_1\|_{H_1} = \|W^{-1}\{f_1\}\|_H \leq \|\phi\|_H$$

With equality being characterized by the condition $\phi(x) = W^{-1}\{f_1(x)\}$ \square

An obvious question that arises in theory is as follows: Given two reproducible kernels $K_1(x, y)$ and $K_2(x, y)$ associated to two, distinct Hilbert spaces H_1 and H_2 , what can be said about the sum $K(x, y) = K_1(x, y) + K_2(x, y)$ of the reproducible kernels? Is it the reproducing kernel of a given Hilbert space, derived from H_1 and H_2 ? An answer to this question is given in the following theorem.

Theorem 2.5.2. Let H_1, H_2, \dots, H_n be reproducing kernel Hilbert space on X with kernels K_1, K_2, \dots, K_n and respective norms $\|\cdot\|_{H_1}, \|\cdot\|_{H_2}, \dots, \|\cdot\|_{H_n}$. Then, $K(x, y) = \sum_{i=1}^n K_i(x, y)$ is the reproducing kernel of the Hilbert space $H = \bigoplus_{i=1}^n H_i = \{f \mid f = \sum_{i=1}^n f_i(x), f_i(x) \in H_i\}$ with the norm $\|\cdot\|_H$ defined by

$$\|f\|_H^2 = \min_{f = \sum_{i=1}^n f_i} \left(\sum_{i=1}^n \|f_i\|_{H_i}^2 \right)$$

Proof. First, we consider the direct sum of all the Hilbert spaces, $H^* = \bigoplus_{i=1}^n H_i = \{(f_1, f_2, \dots, f_n) : f_i \in H_i\}$, with the inner product defined as

$$\langle (f_1, f_2, \dots, f_n), (g_1, g_2, \dots, g_n) \rangle_{H^*} = \sum_{i=1}^n \langle f_i, g_i \rangle_{H_i}$$

From this formula, we immediately deduce the following equation

$$\|(f_1, f_2, \dots, f_n)\|_{H^*}^2 = \sum_{i=1}^n \|f_i\|_{H_i}^2$$

Define, for any ordered set of functions $(f_1, f_2, \dots, f_n) \in H^*$, the operator $T\{(f_1, f_2, \dots, f_n)\} = f_1 + f_2 + \dots + f_n$, whose kernel N is a subspace of H . To show H^* is closed, let $\{f_k^1, f_k^2, \dots, -f_k^n\}_{k=1}^\infty$ be a sequence of elements of N that converge to (f_1, f_2, \dots, f_n) . Since $\lim_{k \rightarrow \infty} \|(f_k^1, f_k^2, \dots, -f_k^n) - (f_1, f_2, \dots, f_n)\|_H = 0$, it necessarily follows that

$$\begin{aligned} \lim_{k \rightarrow \infty} \|f_k^i - f_i\|_{H_i} &= 0 \\ \lim_{k \rightarrow \infty} \|-f_k^n - f_n\|_{H_n} &= 0 \end{aligned}$$

And hence, in the limit, $f_1 + f_2 + \dots + f_{n-1} = -f_n$, implying N is a closed subspace.

Thus, by letting N^\perp be the orthogonal complement of N in H^* , we can define the operator $W = T\{(f_1, f_2, \dots, f_n)\}_{N^\perp}$. Using the same reasoning as in the previous theorem, W is necessarily an injective operator, and hence we can define an inner product on H as

$$\langle f, g \rangle_H = \langle W^{-1}\{f\}, W^{-1}\{g\} \rangle_{H^*}$$

With this inner product, H becomes a Hilbert space of functions defined on X . It is plain that, for any $y \in X$, $K(x, y)$ belongs to H , considered as

a function of X . We now need to verify the reproducing property and the expression for the norm in H .

Let f be an arbitrary function in H , $W^{-1}\{f\} = (f_1, f_2, \dots, f_n)$ and $W^{-1}\{K(x, y)\} = (K_1^*(x, y), K_2^*(x, y), \dots, K_n^*(x, y))$ for all $y \in X$

Yet, it is clear that $\sum_{i=1}^n K_i^*(x, y) - K(x, y) = K(x, y) - K(x, y) = 0$, which, by definition, implies that $(K_1^* - K_1, K_2^* - K_2, \dots, K_n^* - K_n)$ belongs to N , and for $W^{-1}\{f\} = (f_1, f_2, \dots, f_n) \in N^\perp$, which gives:

$$\sum_{i=1}^n \langle f_i, K_i^*(x, y) \rangle_{H_i} = \sum_{i=1}^n \langle f_i, K_i(x, y) \rangle_{H_i}$$

From which we obtain

$$\begin{aligned} \langle f, K(x, y) \rangle_H &= \langle W^{-1}\{f\}, W^{-1}\{K(x, y)\} \rangle_{H'} \\ &= \langle (f_1, f_2, \dots, f_n), (K_1^*(\cdot, y), K_2^*(\cdot, y), \dots, K_n^*(\cdot, y)) \rangle_{H'} \\ &= f_1(y) + f_2(y) + \dots + f_n(y) = f(y). \end{aligned}$$

Thus, K is a reproducing kernel on the Hilbert space H .

To prove the second assertion contained in the theorem, let (f_1, f_2, \dots, f_n) be an element of H' , and set $f = \sum_{i=1}^n f_i$ be the corresponding element of H . We will also consider $\Phi = (\phi_1, \phi_2, \dots, \phi_n) = (f_1, f_2, \dots, f_n) - W^{-1}\{f\}$

Since $W\{\Phi\} = W\{(f_1, f_2, \dots, f_n)\} - f = f - f = 0$, it follows that Φ belongs to N .

Now, by the definition of the norm in H^* , we have

$$\|(f_1, f_2, \dots, f_n)\|_{H^*}^2 = \sum_{i=1}^n \|f_i\|_{H_i}^2$$

Moreover, since $\Phi \in N$ and $W^{-1}\{f\} \in N^\perp$, we deduce that

$$\begin{aligned} \|(f_1, f_2, \dots, f_n)\|_{H^*}^2 &= \|W^{-1}\{f\}\|_{H^*}^2 + \|\Phi\|_{H^*}^2 \\ &= \|W^{-1}\{f\}\|_{H^*}^2 + \sum_{i=1}^n \|\phi_i\|_{H_i}^2 \end{aligned}$$

Thus, we finally deduce that, for $f = f_1 + f_2 + \dots + f_n$,

$$\|f\|_H^2 = \|W^{-1}\{f\}\|_{H^*}^2 \leq \sum_{i=1}^n \|f_i\|_{H_i}^2$$

With equality being characterized by the condition $(f_1, f_2, \dots, f_n) = W^{-1}\{f\}$, implying the desired conclusion. \square

Moreover, we may also describe, in complete detail, the Hilbert space corresponding to the the reproducing kernel to the product $K(x_1, y_2, x_2, y_2) = K_1(x_1, y_1)K_2(x_2, y_2)$ of two reproducing kernels $K_1(x_1, y_1)$ and $K_2(x_2, y_2)$ associated to the Hilbert spaces H_1 and H_2 , by the following theorem:

Theorem 2.5.3. If H_1 and H_2 be two Hilbert spaces of functions defined on X having the reproducing kernels $K_1(x_1, y_1)$ and $K_2(x_2, y_2)$. Then, the direct product $H = H_1 \otimes H_2 = \{f(x_1, x_2) \mid f(x_1, x_2) = f_1(x_1)f_2(x_2), f_i(x_i) \in H_i\}$ is a reproducing kernel Hilbert space with kernel $K = K_1 \otimes K_2$ defined by

$$K(x_1, x_2, y_1, y_2) = K_1(x_1, y_1)K_2(x_2, y_2)$$

and

$$\langle f, K(\cdot, \cdot, y_1, y_2) \rangle_H = f(x_1, x_2)$$

Proof. For the complete demonstration, refer to Aronszajn [2]

The fundamental idea of the proof is analogous to that given for Theorem 3.2.4. By considering first those functions $f(x_1, x_2)$ that can be obtained by a certain combination of $f_1^{(1)}(x), f_1^{(2)}(x), \dots, f_1^{(n)}(x) \in H_1$ and $f_2^{(1)}(x), f_2^{(2)}(x), \dots, f_2^{(n)}(x) \in H_2$, i.e.

$$f(x_1, x_2) = \sum_{j=1}^n f_1^{(j)}(x_1)f_2^{(j)}(x_2)$$

One can proceed to define the inner product of any two function $f(x_1, x_2)$ and $g(x_1, x_2)$ as

$$\langle f, g \rangle_H = \sum_{k=1}^n \sum_{l=1}^n \langle f_1^{(k)}, g_1^{(l)} \rangle_{H_1} \langle f_2^{(k)}, g_2^{(l)} \rangle_{H_2}$$

Then, by completing the set of all elements of the form given above for $f(x_1, x_2)$ equipped with the inner product $\langle f, g \rangle_H$, one can construct the direct product space $H = H_1 \otimes H_2$ and demonstrate that it admits $K(x, y)$ as a reproducing kernel. \square

3 Reproducing Kernels of Analytic Functions

3.1 Preliminary remarks

As mentioned in the historical overview given in the previous section, the theory of reproducing kernel Hilbert spaces owes its origin to certain results obtained in the study of functions of a complex variable, primarily in those properties possessed by the solution of an elliptic partial differential equation.

Amongst those subjects in which the theory of reproducing kernels has been applied, one may justifiably assert that this particular branch of knowledge contains its most aesthetically pleasing and analytically demanding application.

Hence, to demonstrate the astounding generality contained in the theory previously given, we will precede its application in time series analysis by its most developed application in the theory of analytic functions of a complex variable, and their associated spaces.

3.2 Bergman kernel

The first well-known reproducing kernel arising in the theory of complex analysis arose in the work of Bergman, as detailed in his monographs [3] and [4], which provide the logical foundation for the subsequent developments that will be given. Our exposition of this theory, including the demonstrations, will closely follow those which are given by Davis [12] and Hille [19].

Let \mathbb{C} denote the set of complex numbers, written in the form $z = x + yi$, and Ω be a subset of \mathbb{C} , which we shall call a domain. Throughout the rest of this subsection, we will consider the space of square-integrable analytic functions $f(z)$ defined on Ω .

Definition 3.2.1. Let $\Omega \subset \mathbb{C}$ be an open set of complex numbers. The class of all functions which are both analytic and square integrable over Ω , i.e.

$$\iint_{\Omega} |f(z)|^2 dx dy < \infty$$

will be designated by the symbolic expression $A^2(\Omega)$. The integral above is to be interpreted as the Lebesgue integral in \mathbb{C} .

Our principal aim of this chapter is to demonstrate that $A^2(\Omega)$ forms a Hilbert space of analytic functions admitting a reproducing kernel, and thus whose analysis becomes subject to the general propositions given in the previous section.

Lemma 3.2.1. $A^2(\Omega)$ forms a linear space with respect to addition and scalar multiplication by complex numbers, i.e. if $f(z)$ and $g(z) \in L^2(\Omega)$, then $af(z) + bg(z) \in L^2(\Omega)$ for any $a, b \in \mathbb{C}$

Proof. It is well-known that the sum of any linear combination of analytic functions is analytic, and thus $af(z) + bg(z)$ is necessarily analytic.

From the equality

$$|af(z) + bg(z)|^2 + |af(z) - bg(z)|^2 = 2(|a|^2|f(z)|^2 + |b|^2|g(z)|^2)$$

From which we deduce

$$|af(z) + bg(z)|^2 \leq 2(|a|^2|f(z)|^2 + |b|^2|g(z)|^2)$$

Thus we obtain

$$\iint_{\Omega} |af(z) + bg(z)|^2 dx dy \leq 2|a|^2 \iint_{\Omega} |f(z)|^2 dx dy + 2|b|^2 \iint_{\Omega} |g(z)|^2 dx dy$$

The latter two expressions being finite, we derive

$$\iint_{\Omega} |af(z) + bg(z)|^2 dx dy < \infty$$

Hence, it follows that $af(z) + bg(z) \in L^2(\Omega)$ □

Lemma 3.2.2. If $f(z)$ and $g(z)$ are any two elements to $A^2(\Omega)$, then the following integral always exists:

$$\iint_{\Omega} f(z)\overline{g(z)} dx dy$$

Proof. The proof is an immediate consequence of the following identity

$$f(z)\overline{g(z)} = \frac{1}{2}|f(z) + g(z)|^2 + \frac{i}{2}|f(z) + ig(z)|^2 - \frac{1+i}{2}|f(z)|^2 - \frac{1+i}{2}|g(z)|^2$$

□

Lemma 3.2.3. For any two functions $f(z), g(z) \in L^2(\Omega)$, define

$$\langle f, g \rangle = \iint_{\Omega} f(z)\overline{g(z)} dx dy$$

The expression $\langle f(z), g(z) \rangle$ is an inner product on $A^2(\Omega)$. Thus, $A^2(\Omega)$ is an inner product space, and a normed linear space with

$$\|f\|^2 = \iint_{\Omega} |f(z)|^2 dx dy$$

Proof. The expression $\langle f, g \rangle$ is clearly an inner product, by applying the formal properties associated to the symbol of integration. \square

The following technical lemma, whose simple proof is taken from Hille [19], will be necessary to prove that $A^2(\Omega)$ equipped with $\langle f, g \rangle$ forms a complete inner product space.

Lemma 3.2.4. Let $f(z) \in L^2(\Omega)$. Define $\lambda(z_0)$ to be the distance from a point $z_0 \in \Omega$ to the boundary $\partial\Omega$ of Ω . Then,

$$|f(z_0)|^2 \leq \frac{\|f\|^2}{\pi\lambda^2(z_0)}$$

Proof. Let $B = B[z_0, r]$ be the closed ball in Ω centered at z_0 and of radius r , where r is taken to be small enough so that B lies completely in Ω .

By applying Cauchy's theorem to $f^2(z)$, we obtain

$$f^2(z_0) = \frac{1}{2\pi} \int_0^{2\pi} f^2(z_0 + re^{i\theta}) d\theta$$

Multiplying both sides of the equation above by rdr and integrating from 0 to $\lambda(z_0)$, we deduce

$$\begin{aligned} \frac{1}{2} \lambda^2(z_0) f^2(z_0) &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\lambda(z_0)} f^2(z_0 + re^{i\theta}) r dr d\theta \\ &= \frac{1}{2\pi} \iint_B f^2(z) dx dy \end{aligned}$$

From this, we obtain

$$\begin{aligned} \pi\lambda^2(z_0) |f(z_0)|^2 &\leq \iint_B |f(z)|^2 dx dy \\ &\leq \iint_{\Omega} |f(z)|^2 dx dy = \|f\|^2 \end{aligned}$$

And thus, by dividing both terms by $\pi\lambda^2(z_0)$, we get the required result

$$|f(z_0)|^2 \leq \frac{\|f\|^2}{\pi\lambda^2(z_0)}.$$

□

Theorem 3.2.1. The space $A^2(\Omega)$, equipped with the inner product $\langle f, g \rangle$, forms a Hilbert space.

Proof. Consider a Cauchy sequence $\{f_k(z)\}_{k=1}^\infty$ of functions in $A^2(\Omega)$. By definition, given any ϵ , however small, there corresponds an integer N such that, for all $m, n \geq N$,

$$\|f_m - f_n\| \leq \epsilon$$

Let $\Omega_0 \subset \Omega$ with δ being the Hausdorff distance of Ω_0 and $\partial\Omega$, the boundary of Ω . By the preceding lemma, we obtain, for all $z \in \Omega_0$, that

$$|f_m(z) - f_n(z)| \leq \frac{\|f_m - f_n\|}{\pi^{\frac{1}{2}}\lambda(z)}$$

Since $\lambda(z) \geq \delta$, we have that

$$|f_m(z) - f_n(z)| \leq \frac{\|f_m - f_n\|}{\pi^{\frac{1}{2}}\delta}$$

Thus, $\{f_k(z)\}_{k=1}^\infty$ is a Cauchy sequence in the space of analytic functions in Ω_0 . Furthermore, its limit $f(z)$ satisfies

$$\iint_{\Omega_0} |f(z)|^2 dx dy \leq \iint_{\Omega} |f(z)|^2 dx dy \leq \sup_n \iint_{\Omega} |f_n(z)|^2 dx dy = \sup_n \|f_n\|^2$$

Since this holds for any Ω_0 , we conclude that $f(z) \in \Omega$, is analytic and $\|f\|$ is finite. Finally, from the first equation given above, we deduce that

$$\lim_{m \rightarrow \infty} \|f_m - f_n\| = \|f - f_n\| \leq \epsilon.$$

Thus, every Cauchy sequence of functions in $A^2(\Omega)$ has a limit in the same space, thus $A^2(\Omega)$ forms a Hilbert space with the associated inner product.

□

Thus, by a well-known theorem in the theory of abstract Hilbert spaces, the space $A^2(\Omega)$ admits an orthonormal basis of functions $\{\phi_n(x)\}_{n=1}^\infty$. A simple application of Theorem 3.2.10 would provide a general form for the reproducing kernel in terms of the basis elements, assuming of course it exists. To formally demonstrate its existence, we will prove that Theorem 3.2.6 holds for this space, which will require the following technical lemma.

Lemma 3.2.5. Let $B_r = B[z_0, r]$ designate the closed ball in Ω of center z_0 and of radius r . If $f(z) = \sum_{n=0}^\infty a_n(z - z_0)^n$ be any analytic function in $B[z_0, r]$, then

$$\iint_{B_r} |f(z)|^2 dx dy = \pi \sum_{n=0}^\infty |a_n|^2 \frac{r^{2n+2}}{n+1}$$

Proof. Since $f(z) = \sum_{n=0}^\infty a_n(z - z_0)^n$ is analytic, the series necessarily converges both uniformly and absolutely on B_r . This fact formally justifies the following operations:

$$\begin{aligned} \iint_{B_r} |f(z)|^2 dx dy &= \iint_{B_r} \left\{ \sum_{n=0}^\infty a_n(z - z_0)^n \right\} \left\{ \sum_{m=0}^\infty \bar{a}_m(\bar{z} - \bar{z}_0)^m \right\} dx dy \\ &= \sum_{n=0}^\infty \sum_{m=0}^\infty a_n \bar{a}_m \iint_{B_r} (z - z_0)^n (\bar{z} - \bar{z}_0)^m dx dy \end{aligned}$$

By using the substitution $(z - z_0)^n = \rho e^{i\theta}$, the integral above reduces to the following:

$$\iint_{B_r} (z - z_0)^n (\bar{z} - \bar{z}_0)^m dx dy = \int_0^{2\pi} d\theta e^{(m-n)i\theta} \int_0^r \rho^{m+n+1} d\rho$$

When $m \neq n$, so that $m - n = k$, the inner integral vanishes, for

$$\int_0^{2\pi} d\theta e^{(m-n)i\theta} = \int_0^{2\pi} d\theta e^{i(k\theta)} = k \int_0^{2\pi} d\psi e^{i\psi} = 0$$

If $m = n$, the inner integral is clearly equal to π , while the latter integral is equal to $\frac{r^{2(n+1)}}{n+1}$. Thus, we deduce that

$$\iint_{B_r} |f(z)|^2 dx dy = \pi \sum_{n=0}^\infty |a_n|^2 \frac{r^{2n+2}}{n+1}$$

□

Theorem 3.2.2. Let a fixed positive integer $n \geq 0$ be given. If t designates any point in the interior of a set $\Omega \subset \mathbb{C}$, then the functional

$$L_t\{f\} = f^{(n)}(t)$$

is bounded for any function f in $A^2(\Omega)$.

Proof. Since t is a point in the interior of Ω , there corresponds a closed ball $B_r = B[t, r]$ that is completely contained in Ω

Since $f(z) = \sum_{k=0}^{\infty} \frac{f^{(k)}(t)}{k!} (z-t)^k$ in B_r , an application of the preceding lemma produces

$$\pi \sum_{k=0}^{\infty} \frac{|f^{(k)}(t)|^2}{(k!)^2 (k+1)} r^{2k+2} = \iint_{B_r} |f(z)|^2 dx dy \leq \iint_{\Omega} |f(z)|^2 dx dy$$

For a particular value of $k = n$, we obtain

$$\pi \frac{|f^{(n)}(t)|^2}{(n!)^2 (n+1)} r^{2n+2} \leq \iint_{\Omega} |f(z)|^2 dx dy = \|f\|^2$$

Thus, we finally obtain the required result, viz.

$$|L_t\{f\}| = |f^{(n)}(t)| \leq \frac{n! \sqrt{n+1}}{\sqrt{\pi} r^{n+1}} \|f\| .$$

Hence, $L_t\{f\} = f^{(n)}(t)$ is a bounded linear functional for all n . \square

Therefore, by replacing $n = 1$ in the theorem above, we have proven that the condition of Theorem 3.2.6 holds in $A^2(\Omega)$, hence this space admits a reproducing kernel $B(w, z)$. If $\{\phi_n(z) : n \in \mathbb{N}\}$ designates an orthonormal basis of $A^2(\Omega)$, then, by Theorem 3.2.10, we obtain the following representation

$$B(w, z) = \sum_{n=1}^{\infty} \phi_n(z) \overline{\phi_n(w)}$$

This general expression is called the *Bergman kernel* for the domain Ω of the space $A^2(\Omega)$. In certain particular cases, the kernel above reduces to a concrete algebraic expression, such as when $\Omega = S^1 = \{z \in \mathbb{C} : |z| = 1\}$. The following theorem illustrates this fact, whose proof is inspired by that given in Duren [13].

Theorem 3.2.3. The Bergman space $L^2(S^1)$ of all analytic functions $f(z)$ for which

$$\iint_{S^1} |f(z)|^2 dx dy < \infty$$

is a reproducing kernel Hilbert space with respect to the inner product

$$\langle f, g \rangle = \iint_{S^1} f(z) \overline{g(z)} dx dy$$

admitting the reproducing kernel $B(w, z)$ defined by

$$B(w, z) = \frac{1}{\pi} \frac{1}{(1 - z\bar{w})^2}$$

Proof. That $L^2(S^1)$ equipped with the inner product $\langle f, g \rangle$ forms a Hilbert space is clearly a special case of Theorem 4.2.1, and that it admits a reproducing kernel $B(w, z) = \sum_{n=1}^{\infty} \phi_n(z) \overline{\phi_n(w)}$ follows by an analogous argument to the one given in the paragraph above.

To obtain an explicit construction for $B(w, z)$, it will be shown that the set $\{\phi_s(z) : s \in \mathbb{N}\}$, whose general term $\phi_n(z)$ is given by

$$\phi_n(z) = \sqrt{\frac{n+1}{\pi}} z^n$$

forms an orthonormal basis for the space $L^2(S^1)$.

By making the substitution $r = \rho e^{i\theta}$, and applying the same reasoning as in Lemma 4.2.6., we obtain

$$\begin{aligned} \langle \phi_n(z), \phi_m(z) \rangle &= \sqrt{\frac{(n+1)(m+1)}{\pi^2}} \iint_{S^1} z^n \overline{z^m} dx dy \\ &= \frac{\sqrt{(n+1)(m+1)}}{\pi} \int_0^{2\pi} e^{i(n-m)\theta} d\theta \int_0^1 \rho^{n+m+1} d\rho \\ &= \frac{\sqrt{(n+1)(m+1)}}{\pi} \frac{2\pi}{n+m+2} \delta_{nm} \\ &= \frac{2\sqrt{(n+1)(m+1)}}{n+m+2} \delta_{nm} \end{aligned}$$

Hence, we obtain that $\|\phi_n\| = 1$ for all n and $\langle \phi_n(z), \phi_m(z) \rangle = 0$ whenever $m \neq n$. Therefore, the set $\{\phi_n(z) : n \in \mathbb{N}\}$ is an orthonormal set.

Now, since any function $f(z)$ in $L^2(S^1)$ is analytic, it will be expressed in the form $f(z) = \sum_{k=0}^{\infty} a_k z^k$. To prove that the set $\{\phi_s(z) : s \in \mathbb{N}\}$ is complete in $L^2(S^1)$, it is sufficient to demonstrate Parseval's relation holds, viz.

$$\sum_{n=0}^{\infty} |\langle f, \phi_n \rangle|^2 = \|f\|^2$$

By using a result from Lemma 4.2.4, we immediately deduce that

$$\begin{aligned} |\langle f, \phi_n \rangle| &= \iint_{S^1} \left\{ \sqrt{\frac{n+1}{\pi}} z^n \right\} \left\{ \sum_{k=0}^{\infty} a_k z^k \right\} dx dy \\ &= \sqrt{\frac{n+1}{\pi}} \sum_{k=0}^{\infty} a_k \iint_{S^1} z^k \bar{z}^n dx dy \\ &= \sqrt{\frac{n+1}{\pi}} \sum_{k=0}^{\infty} a_k \frac{\pi}{n+1} \delta_{nk} = \sqrt{\frac{\pi}{n+1}} \sum_{k=0}^{\infty} a_k \delta_{nk} \end{aligned}$$

By simple algebraic manipulations, we obtain the following result

$$\sum_{n=0}^{\infty} |\langle f, \phi_n \rangle|^2 = \pi \sum_{n=0}^{\infty} \frac{|a_n|^2}{n+1}$$

By letting $s_n(z) = \sum_{k=0}^n a_k z^k$ to be the partial sums of $f(z)$ and putting $S^q = \{z : |z| < q < 1\}$, we have

$$\begin{aligned} \iint_{S^q} |s_n(z)|^2 dx dy &= \sum_{k=0}^n \sum_{l=0}^n a_k \bar{a}_l \int_0^{2\pi} e^{(k-l)i\theta} d\theta \int_0^q \rho^{k+l+1} d\rho \\ &= \sum_{k=0}^n \sum_{l=0}^n a_k \bar{a}_l \pi \frac{q^{2n+2}}{n+1} \delta_{kl} \\ &= \pi \sum_{k=0}^n \frac{|a_k|^2}{k+1} q^{2n+2} \end{aligned}$$

Since $s_n(z)$ converges uniformly to $f(z)$ on S_q^1 , the following necessarily holds:

$$\|f\|^2 = \lim_{q \rightarrow 1} \iint_{S^q} |f(z)|^2 dx dy = \lim_{q \rightarrow 1} \pi \sum_{k=0}^{\infty} \frac{|a_k|^2}{k+1} q^{2n+2} = \pi \sum_{k=0}^{\infty} \frac{|a_k|^2}{k+1}$$

From which we conclude that $\sum_{n=0}^{\infty} |\langle f, \phi_n \rangle|^2 = \|f\|^2$, hence the family of functions $\{\phi_s(z) : s \in \mathbb{N}\}$ is an orthonormal basis of $L^2(S^1)$

Thus, by applying Theorem 3.2.10, we derive that the reproducing kernel $B(w, z)$ of the space $L^2(S^1)$ is given by the formula

$$\begin{aligned} B(w, z) &= \sum_{n=1}^{\infty} \phi_n(z) \overline{\phi_n(w)} \\ &= \sum_{n=1}^{\infty} \sqrt{\frac{n+1}{\pi}} w^n \sqrt{\frac{n+1}{\pi}} \bar{z}^n = \frac{1}{\pi} \sum_{n=1}^{\infty} (n+1) (w\bar{z})^n \\ &= \frac{1}{\pi} \frac{d}{du} \left(\sum_{n=1}^{\infty} u^n \right) \Big|_{u=w\bar{z}} = \frac{1}{\pi} \frac{1}{(1-u)^2} \Big|_{u=w\bar{z}} \\ B(w, z) &= \frac{1}{\pi} \frac{1}{(1-w\bar{z})^2} \end{aligned}$$

□

3.3 Bargmann-Fock kernel

Another well-known example of a Hilbert space of analytic functions admitting a reproducing kernel is the Bargmann-Fock space, originally considered in the field of quantum mechanics. Our presentation of this given subject will closely follow those given by Hall in [16] and [15].

Definition 3.3.1. The class of all functions $f(z)$ which are analytic in an open set $\Omega \subset \mathbb{C}$ and for which

$$\|f\|_{F^2(\Omega, w_h)}^2 = \iint_{\Omega} |f(z)|^2 (\pi h)^{-1} \exp(-|z|^2/h) dx dy < \infty$$

is said to be the Bargmann-Fock space $F^2(\Omega, w_h)$ with the associated weight function $w_h(z) = (\pi h)^{-1} \exp(-|z|^2/h)$, with $h > 0$. The integral above is to be understood as the Lebesgue integral in \mathbb{C} .

The following theorem will prove the Bargmann-Fock space defined above forms a Hilbert space and that it admits a reproducing kernel, by verifying the condition of Theorem 3.2.6.

Theorem 3.3.1. Let f denote an arbitrary function in the space $F^2(\Omega, w_h)$ defined above. The following properties are possessed by $F^2(\Omega, w_h)$.

1. If z designates any point in \mathbb{C} , then the functional

$$L_z\{f\} = f(z)$$

is bounded for any function f in $F^2(\mathbb{C}, w_h)$.

2. $F^2(\Omega, w_h)$ is a closed subspace of $L^2(\Omega, w_h)$, and therefore is a Hilbert space with respect to the corresponding inner product

$$\langle f, g \rangle_{F^2(\Omega, w_h)} = \iint_{\Omega} f(z) \overline{g(z)} w_h(z) dx dy$$

Proof. Let $B_r = B[t, r] = \{z \in \mathbb{C} : |z - t| < r\}$ denote a closed ball in Ω such that r is taken to be sufficiently small so that the closure of B_r is entirely contained in Ω , and let t denote an arbitrary point in B_r .

Since $f(t)$ is analytic on B_r , it will have the following Taylor expansion

$$f(t) = f(z) + \sum_{n=1}^{\infty} a_n (t - z)^n$$

which converges uniformly for all values of t belonging to $B_r \subset \Omega$.

Thus, by letting $(t - z)^n = \rho^n e^{in\theta}$, we obtain the following result

$$\begin{aligned} (\pi r^2)^{-1} \iint_{B_r} f(t) dx dy &= (\pi r^2)^{-1} \iint_{B_r} [f(z) + \sum_{n=1}^{\infty} a_n (t - z)^n] dx dy \\ &= (\pi r^2)^{-1} \left[\iint_{B_r} f(z) dx dy + \sum_{n=1}^{\infty} \iint_{B_r} (t - z)^n dx dy \right] \\ &= (\pi r^2)^{-1} (\pi r^2) f(z) + \sum_{n=1}^{\infty} a_n \left[\int_0^{2\pi} e^{in\theta} d\theta + \int_0^r \rho^{n+1} d\rho \right] \\ &= f(z) \end{aligned}$$

Let $\mathbb{1}_{B_r}(t)$ be the indicator function for the set $B_r \subset \mathbb{C}$. Since B_r was chosen arbitrarily, the following equations necessarily holds

$$\begin{aligned} f(z) &= (\pi r^2)^{-1} \iint_{\Omega} \mathbb{1}_{B_r}(t) f(t) dx dy \\ &= (\pi r^2)^{-1} \iint_{\Omega} \left\{ \mathbb{1}_{B_r}(t) \frac{1}{w_h(t)} \right\} \{f(t) w_h(t)\} dx dy \\ &= (\pi r^2)^{-1} \langle \mathbb{1}_{B_r} w_h^{-1}, f \rangle_{F^2(\Omega, w_h)} \end{aligned}$$

By applying the Cauchy-Schwarz inequality, we finally derive that

$$|f(z)|^2 \leq (\pi r^2)^{-1} \left\| \mathbb{1}_{B_r} w_h^{-1} \right\|_{F^2(\Omega, w_h)}^2 \|f\|_{L^2(\Omega, w_h)}^2 .$$

Since $w_h(t)$ is bounded and positive on \mathbb{C} , its reciprocal $w_h^{-1}(t)$ is necessarily bounded on $B_r(t)$, hence the norm $\left\| \mathbb{1}_{B_r} w_h^{-1}(t) \right\|_{F^2(\Omega, w_h)}$ is necessarily finite. By setting $M_z^2 = (\pi r^2)^{-1} \left\| \mathbb{1}_{B_r} w_h^{-1} \right\|_{F^2(\Omega, w_h)}^2$, we obtain the desired result for (1), viz.

$$|L_z\{f\}| = |f(z)| \leq M_z \|f\|_{F^2(\Omega, w_h)} .$$

To demonstrate the second assertion, let us suppose we have a sequence $\{f_n(z)\}_{n=0}^\infty$ of functions in $F^2(\mathbb{C}, w_h)$ converging to a function $f(z)$ in $L^2(\Omega, w_h)$. Then, it necessarily follows that f_n is a Cauchy sequence, with the added property from the reasoning above being that

$$\lim_{n, m \rightarrow \infty} \sup_{z \in \Omega} |f_n - f_m| \leq M_z \|f_n - f_m\|_{F^2(\Omega, w_h)} = 0$$

From this equation, it follows that $\{f_n(z)\}_{n=1}^\infty$ converges locally uniformly to a function that must necessarily be $f(z)$, for $F^2(\Omega, w_h)$ convergence is equivalent to pointwise convergence. But, it is standard result in complex analysis that any uniform limit of analytic functions is analytic, thus the limit f must belong to $F^2(\Omega, w_h)$, and thus $F^2(\Omega, w_h)$ is a closed subspace of $L^2(\Omega, w_h)$, and therefore is a Hilbert space. \square

To provided a concrete kernel $F^2(\mathbb{C}, w_h)$, the theorem given below will provide a concrete orthonormal basis for $F^2(\mathbb{C}, w_h)$, and apply Theorem 3.2.10.

Theorem 3.3.2. Let

$$\phi_n(z) = \frac{z^n}{\sqrt{n! t^n}}$$

Then, the Bargmann-Fock space $F^2(\mathbb{C}, w_h)$ is a Hilbert space having $\{\phi_n(z)\}_{n=0}^\infty$ as an orthonormal basis and admitting the reproducing kernel $F(x, y)$ given by

$$F(z, w) = \exp\left(\frac{z\bar{w}}{h}\right)$$

Proof. The method of demonstrating that the set of functions $\{\phi_n(z)\}_{n=0}^{\infty}$ is an orthonormal basis for $F^2(\mathbb{C}, w_k)$ is analogous to the one given in Theorem 4.2.3.

After applying the substitution $z = \rho e^{i\theta}$, we obtain that

$$\begin{aligned}
\langle z^n, z^m \rangle_{F^2(\Omega, w_h)} &= (\pi h)^{-1} \iint_{\Omega} z^n z^m \exp(-|z|^2 / h) dx dy \\
&= (\pi h)^{-1} \int_0^{2\pi} e^{i(n-m)\theta} d\theta \int_0^{\infty} \rho^{n+m} \exp\left(-\frac{\rho^2}{h}\right) \rho d\rho \\
&= (\pi h)^{-1} 2\pi \delta_{n,m} \int_0^{\infty} \rho^{2n} \exp\left(-\frac{\rho^2}{h}\right) \rho d\rho \\
&= h^{-1} \delta_{n,m} \int_0^{\infty} \left(\frac{\rho^2}{h}\right)^n \exp\left(-\frac{\rho^2}{h}\right) \rho d\left(\frac{\rho^2}{h}\right) h^{n+1} \\
&= \delta_{n,m} h^n \int_0^{\infty} x^n \exp(-x) dx \\
&= h^{n+1} \Gamma(n+1) \delta_{n,m} \\
\langle z^n, z^m \rangle_{F^2(\Omega, w_h)} &= n! h^n \delta_{n,m}.
\end{aligned}$$

From this, it immediately follows that $\langle \phi_n(z), \phi_m(z) \rangle_F = \delta_{n,m}$, and thus the collection of functions $\{\phi_n(z)\}_{n=0}^{\infty}$ forms an orthonormal set. To show that it forms an orthonormal basis, it is sufficient to prove that $\langle f(z), z^n \rangle_F = 0$ implies $f(z) = 0$.

Since $f(z)$ is supposed to be analytic in the whole complex domain, it will necessarily possess a uniformly convergent Taylor series expansion, expressed by $f(z) = \sum_{k=0}^{\infty} a_k z^k$.

Thus, we obtain

$$\begin{aligned}
\langle f(z), z^n \rangle_{F^2(\Omega, w_h)} &= (\pi h)^{-1} \iint_{\mathbb{C}} f(z) \bar{z}^n \exp(-|z|^2 / h) dx dy \\
&= (\pi h)^{-1} \iint_{\mathbb{C}} \left(\sum_{k=0}^{\infty} a_k z^k \right) \bar{z}^n \exp(-|z|^2 / h) dx dy \\
&= (\pi h)^{-1} \sum_{k=0}^{\infty} a_k \iint_{\mathbb{C}} z^k \bar{z}^n \exp(-|z|^2 / h) dx dy \\
&= \sum_{k=0}^{\infty} a_k \|z^n\|^2 \delta_{k,n} \\
\langle f(z), z^n \rangle_{F^2(\Omega, w_h)} &= a_n \|z^n\|^2 .
\end{aligned}$$

Hence, if $\langle f(z), z^n \rangle_F = 0$, it is necessary that $a_n = 0$ for all $n \in \mathbb{N}$, which is equivalent to $f(z) = 0$. Thus, the set of all functions $\{\phi_n(z)\}_{n=0}^{\infty}$ forms an orthonormal basis of $F^2(\mathbb{C}, w_k)$.

That the space $F^2(\mathbb{C}, w_k)$ admits a reproducing kernel is a consequence of a result demonstrated in the theorem above, namely that every evaluation functional $L_z\{f\}$ is bounded, which by Theorem 3.2.6. is a sufficient condition for its existence. Since $\{\phi_n(z)\}_{n=0}^{\infty}$ is an orthonormal basis for $F^2(\Omega, w_k)$, we obtain, by applying Theorem 3.2.10, that its reproducing kernel $F(z, w)$ is given by the following formula

$$\begin{aligned}
F(z, w) &= \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n! h^n}} \frac{\bar{w}^n}{\sqrt{n! h^n}} \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{z\bar{w}}{h} \right)^n \\
F(z, w) &= \exp\left(\frac{z\bar{w}}{h} \right) .
\end{aligned}$$

□

4 Application to Time Series Analysis

4.1 Introduction

In a lengthy technical report published by Parzen in the winter of 1959 [29], it was demonstrated by the author that the theory of time series analysis, of regression analysis, of the analysis of variance, amongst others, could all be united, logically speaking, into a single framework, namely that of reproducing kernel Hilbert spaces.

This method of reasoning naturally leads to solutions of the highest degree of generality, which for example, in the case of time series analysis, could be applied to stationary or non-stationary multiple time series observed at discrete or continuous points in time. The object of this chapter will thus consist in a formalized description of this incredible discovery, within the context of time series analysis.

Apart from the last subsection, the treatment given will follow, in both order and method of demonstration, the various papers contained in Parzen [30], in particular the technical report cited above. The last subsection will present a surprising application of the theories hitherto developed to derive general sampling theorems for time series.

4.2 Hilbert Space representations of time series

To be able to adequately present the following considerations, it will be necessary to define certain basic notions that will be commonly employed throughout the remainder of this chapter.

Definition 4.2.1. Let T be an indexing set, H be a real Hilbert space with norm $\|\cdot\|_H$, and $V = \{v(t) : t \in T\}$ be a collection of elements of H . The linear subspace spanned by the elements of V , which we denote by $\text{Sp}(V)$, is defined as

$$\text{Sp}(V) = \left\{ \sum_{i=1}^n a_i v(t_i) \mid n \in \mathbb{N}, a_i \in \mathbb{R}, v(t_i) \in V \right\}.$$

The closed linear subspace spanned by the elements of V , which we denote by $\overline{\text{Sp}}(V)$, is defined as

$$\overline{\text{Sp}}(V) = \{v \in H \mid \forall \varepsilon > 0 \exists v^* \in \text{Sp}(V) : \|v - v^*\|_H < \varepsilon\}.$$

If $\overline{\text{Sp}}(V) = H$, we say that $V = \{v(t) : t \in T\}$ *spans* the Hilbert space H . Loosely speaking, the closed linear subspace $\overline{\text{Sp}}(V)$ of V can be understood

as the set of all elements that can be expressed in the form $v = \sum_{i=1}^n a_i v(t_i)$ or $v' = \lim_{n \rightarrow \infty} \sum_{i=1}^n a'_i v(t'_i)$, where the infinite series converges in the norm.

Definition 4.2.2. Given an abstract probability space (Ω, \mathcal{F}, P) , we denote by $L^2(\Omega, \mathcal{F}, P)$ the Hilbert space of all \mathcal{F} measurable random variables X defined on Ω that are squared integrable, i.e. for which

$$\langle X, X \rangle_P = \int_{\Omega} X^2 dP = E[X^2] < \infty.$$

In order to define what we mean by the representation of a time series, we must first define the term *time series*. To do so, we will utilize the well-known notion of a stochastic process, defined on an arbitrary probability space, and hence obtain the following definition:

Definition 4.2.3. A *time series* is a collection $\{X(t) : t \in T\}$ of random variables in $L^2(\Omega, \mathcal{F}, P)$ indexed by the elements of a set T . The *mean value function* $m(t)$ and *covariance function* $K(s, t)$ are functions defined on T and $T \times T$, respectively, by the formulas

$$m(t) = E[X(t)], \quad K(s, t) = E[X(s)X(t)].$$

Furthermore, to any time series $\{X(t) : t \in T\}$, there corresponds a unique Hilbert space $L_2(X(t), t \in T)$, whose consideration will be useful for solving the fundamental problems of time series. This very important space is defined as such:

Definition 4.2.4. Let $\{X(t) : t \in T\}$ be an arbitrary time series. We denote by $L^2(X(t), t \in T)$ the closed linear subspace of $L^2(\Omega, \mathcal{F}, P)$, which may be heuristically defined as the set of all random variables $U \in L^2(\Omega, \mathcal{F}, P)$ that can be expressed in one of the forms given below

$$U = \sum_{i=1}^n a_i X(t_i), \quad U = \lim_{n \rightarrow \infty} \sum_{i=1}^n b_i X(t_i),$$

where the infinite series converges with respect to the inner product of $L^2(\Omega, \mathcal{F}, P)$, namely $\langle X, Y \rangle_P = E[XY]$.

The different Hilbert space representations of a time series all depend upon the following important theorem.

Theorem 4.2.1. Let H_1 and H_2 be two Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_{H_1}$ and $\langle \cdot, \cdot \rangle_{H_2}$, respectively. Let T be an index set, and $V = \{v(t) : t \in T\}$ and $W = \{w(t) : t \in T\}$ be two sets of elements that span H_1 and H_2 , respectively. If, for every $s, t \in T$,

$$\langle v(s), v(t) \rangle_{H_1} = \langle w(s), w(t) \rangle_{H_2},$$

then there exists a isometry (a bijective inner product preserving linear mapping) $\psi : H_1 \rightarrow H_2$ having the property that

$$\psi(v(t)) = w(t).$$

Proof. The idea of the proof consists in constructing the function ψ from the given sets spanning the respective Hilbert spaces, extending its definition to any linear combination of the elements of these sets, and then proving it has the satisfied properties.

Hence, define the function ψ as follows. For any element $v(t)$ in U , let $\psi(v(t)) = w(t)$. For any element $v = \sum_{i=1}^n a_i v(t_i)$, define

$$\psi(v) = \sum_{i=1}^n a_i w(t_i)$$

To prove the mapping ψ is well-defined, it is clear that it suffices to prove that

$$\sum_{i=1}^n a_i v(t_i) = 0 \quad \text{if and only if} \quad \sum_{i=1}^n a_i w(t_i) = 0$$

The assertion above immediately follows from the following calculation

$$\begin{aligned} 0 &= \left\| \sum_{i=1}^n a_i v(t_i) \right\|_{H_1}^2 = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \langle v(t_i), v(t_j) \rangle_{H_1} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \langle w(t_i), w(t_j) \rangle_{H_2} = \left\| \sum_{i=1}^n a_i w(t_i) \right\|_{H_2}^2 \\ &= \left\| \sum_{i=1}^n a_i w(t_i) \right\|_{H_2}^2. \end{aligned}$$

It is clear that ψ is a bijective and linear mapping from the span of V

to the span of W . To show it is inner-product preserving, we note that

$$\begin{aligned}
\langle \psi \left\{ \sum_{i=1}^n a_i v(t_i) \right\}, \psi \left\{ \sum_{j=1}^n a_j v(t_j) \right\} \rangle_{H_2} &= \left\langle \sum_{i=1}^n a_i w(t_i), \sum_{j=1}^n a_j w(t_j) \right\rangle_{H_2} \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \langle w(t_i), w(t_j) \rangle_{H_2} \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \langle v(t_i), v(t_j) \rangle_{H_1} \\
\langle \psi \left\{ \sum_{i=1}^n a_i v(t_i) \right\}, \psi \left\{ \sum_{j=1}^n a_j v(t_j) \right\} \rangle_{H_2} &= \left\langle \sum_{i=1}^n a_i v(t_i), \sum_{j=1}^n a_j v(t_j) \right\rangle_{H_1}.
\end{aligned}$$

Thus, since ψ is inner product preserving, any sequence $\{v_n\}_{n=1}^\infty$ in the span of $\{v(t) : t \in T\}$ is a Cauchy sequence in H_1 if and only if $\{\psi(v_n)\}_{n=1}^\infty$ is a Cauchy sequence in H_2 . From this, we immediately deduce that $\lim_{n \rightarrow \infty} v_n = 0$ if and only if $\lim_{n \rightarrow \infty} \psi(v_n) = 0$.

For $v = \lim_{n \rightarrow \infty} v_n$, define $\psi(v) = \lim_{n \rightarrow \infty} \psi(v_n)$. Since the families $V = \{v(t) : t \in T\}$ and $W = \{w(t) : t \in T\}$ is assumed to span H_1 , ψ is clearly now defined for any element of H_1 . By repeating an argument analogous to the one given above, it easily follows that ψ is a isometry from H_1 into H_2 . □

Having properly defined and demonstrated all the required preliminary notions, we are now set to define the Hilbert space representation of a time series.

Definition 4.2.5. A collection $F = \{f(t) : t \in T\}$ of elements in a Hilbert space H is said to be a representation of a time series $\{X(t) : t \in T\}$ if, for every $s, t \in T$,

$$\langle f(s), f(t) \rangle_H = K(s, t) = E[X(s)X(t)]$$

Therefore, there exists a isometry $\psi : \overline{\text{Sp}}(F) \rightarrow L^2(X(t), t \in T)$ satisfying

$$\psi(f(t)) = X(t)$$

Moreover, every random variable U in $L^2(X(t), t \in T)$ can be expressed as

$$U = \psi(g)$$

for a unique element g in $\overline{\text{Sp}}(F)$, by the bijective property of ψ .

It is well-known that the covariance function $K(s, t)$ of any given time series $\{X(t) : t \in T\}$ is a positive definite function, as defined in Definition 2.2.1. Therefore, from Theorem 2.2.5., to any covariance function K , there corresponds a unique Hilbert space H_K admitting K as the reproducing kernel. By applying Lemma 4.2.1., we obtain the following lemma, which provides an explicit description of H_K .

Lemma 4.2.1. The covariance function $K(s, t)$ of a time series $\{X(t) : t \in T\}$ generates a unique Hilbert space H_K consisting of functions $g(t)$ defined on T of the form

$$g(t) = E[UX(t)],$$

for some unique random variable $U \in L^2(X(t), t \in T)$ and equipped with the norm

$$\|g\|_{H_K} = E[U^2].$$

Moreover, H_K admits the covariance function $K(s, t)$ as its reproducing kernel.

Proof. The existence and uniqueness of the Hilbert space H_K corresponding to the covariance function K follows from Theorem 2.2.5.

Hence, it is sufficient to prove that $K(s, t)$ is a reproducing kernel to the Hilbert space H_K defined above.

Clearly, since $K(s, t) = E[X(s)X(t)]$, where $X(s) \in L^2(X(t), t \in T)$, it follows that, for all $s \in T$, $K(s, t)$ belongs in H_K as a function of t .

Also, since $\langle g(t), K(s, t) \rangle_{H_K} = E[UX(s)] = g(s)$, we deduce that $K(s, t)$ is the reproducing kernel of H_K . \square

The following theorem provides the reproducing kernel Hilbert space representation of any time series of finite second moment.

Theorem 4.2.2. Let $\{X(t) : t \in T\}$ be a time series with covariance function $K(s, t)$. Then, the set $\mathcal{K} = \{K(\cdot, t) : t \in T\}$ of functions in H_K is a representation for $\{X(t) : t \in T\}$.

Moreover, there exists a isometry $\psi : H_K \rightarrow L^2(X(t), t \in T)$ such that, for every $t \in T$ and $g, h \in H_K$, we have the following conditions:

$$\begin{aligned} \psi(K(\cdot, t)) &= X(t) \\ \langle g, h \rangle_{H_K} &= E[\psi(g)\psi(h)] \end{aligned}$$

Proof. The first assertion contained in the theorem above follows immediately from the obvious relation

$$K(s, t) = \langle K(\cdot, s), K(\cdot, t) \rangle_{H_K} = E[X(s)X(t)]$$

To prove the second assertion, notice that by Definition 4.2.5., there exists a isometry ψ from $\overline{\text{Sp}}(\mathcal{K})$ to $L^2(X(t), t \in T)$ having the required properties.

Therefore, we need to show that \mathcal{K} spans H_K , so that $\overline{\text{Sp}}(\mathcal{K}) = H_K$.

It suffices to prove that the only function $f \in H_K$ that is orthogonal to each function $K(\cdot, t)$ is $f = 0$. But, from the reproducing property of $K(s, t)$, it is plain that $\langle f(s), K(\cdot, t) \rangle = 0$ for every $t \in T$ implies that $f = 0$, since $\langle f(s), K(\cdot, t) \rangle = f(t)$ \square

4.3 Prediction and Smoothing using RKHS methods

The two central problems of time series analysis, and its subordinate fields of communication and control, of signal processing, of econometric analysis, and so forth, are without a doubt those of *prediction* and *smoothing*. Thus, it could be argued with good reason that the prior developments given are of little value unless they provide new solutions to these fundamental problems.

As will be shown in the following paragraphs, the method of reproducing kernel Hilbert spaces provides solutions that are unparalleled in their degree of generality, yet that are, practically speaking, not easily computable. Though this clearly is a deficiency in the theory, it is the author's position that this fact does not detract from either its theoretical merit or analytical power.

The problems of *prediction* and *smoothing* can be intuitively thought of as estimating and forming certain hypotheses about a given quantity having given a number of observations of another quantity depending on chance. More often than not, the given quantity is supposed to be observed at successive points in time, which thus leads us naturally into the consideration of time series.

Let T denote be an arbitrary index set, whose elements represent the points on an arbitrary time axis, and $X(t)$ denote the observation made at time t . Given the time series $\{X(t) : t \in T\}$, the problems of prediction and smoothing may be formally stated as such:

1. *Prediction*: Observe the time series $X(t)$ over the interval $[0, s]$, and then predict $X(\tau)$ for any positive value of $\tau > s$.
2. *Smoothing*: Observe the sum $X(t) = S(t) + N(t)$ of two time series $S(t)$ and $N(t)$ over the interval $[0, s]$; then estimate $S(\tau)$ for any value $\tau \in [0, s]$. The time series $S(t)$ is often interpreted as the signal, while $N(t)$ is interpreted as the noise, and hence the problem consists in filtering out the noise from the observed process to obtain the true signal.

It is clear that each of the problems above consists in estimating or predicting a random variable W from a given time series $\{X(t) : t \in T\}$. In the solutions which we shall develop, W will always be estimated or predicted by means of a linear functional of $\{X(t) : t \in T\}$. We will designate by

$$\mathcal{E}[W \mid X(t), t \in T]$$

as being that linear functional which minimizes the mean-squared error. As is well-known, in the case of non-linear functionals, this problem reduces to the computation of the conditional expectation $E[W \mid X(t) : t \in T]$.

The following theorem provides a general solution to the problem of prediction, by employing the reproducing kernel Hilbert space representation.

Theorem 4.3.1. Let $\{X(t) : t \in T\}$ be a time series with covariance function $K(s, t)$, H_K be the corresponding reproducing kernel Hilbert space, and ψ denote the isometry from H_K to $L^2(X(t), t \in T)$. If W is a random variable with finite second moment and $\nu_W(t) = E[WX(t)]$, then

$$\mathcal{E}[W \mid X(t), t \in T] = \psi(\nu_W(t))$$

with the associated mean squared error given by the formula

$$E \left[(W - \mathcal{E}[W \mid X(t), t \in T])^2 \right] = E(W^2) - \|\nu_W(t)\|_{H_K}^2 .$$

Proof. From the second condition in Theorem 4.2.2., and by the definition of $\nu_W(t)$, it follows that, for any $h(t) \in H_K$,

$$E[W\psi(h(t))] = \langle \nu_w(t), h(t) \rangle_{H_K} .$$

Since the isometry ψ is necessarily bijective, it follows that any random variable in $L^2(X(t) : t \in T)$ is equal to $\psi(h(t))$, for some $h(t) \in H_K$.

Thus, the mean-square error between any linear predictor $\psi(h(t))$ and W is clearly given by:

$$\begin{aligned} E[|\psi(h) - W|^2] &= E[\psi^2(h)] + E[W^2] - 2E[W\psi(h)] \\ &= E[W^2] + \langle h, h \rangle_{H_K} - 2\langle \nu_W, h \rangle_{H_K} \\ &= E[W^2] - \|\nu_W\|_{H_K}^2 + \langle h - \nu_W, h - \nu_W \rangle_{H_K} . \end{aligned}$$

We clearly see that $h(t) = \nu_w(t)$ minimizes the mean-square error, and thus is the best linear predictor. □

To address the problem of *smoothing*, we will reformulate the statement of the problem given above, so as to reduce it to one of continuous linear regression.

Consider a time series $\{X(t) : t \in T\}$ with a proper covariance function $R(s, t) = \text{Cov}(X(s), X(t))$. We will regard the time series $X(t)$ as being defined by

$$X(t) = m(t) + Y(t)$$

with $m(t)$ being considered as the *mean* function and $Y(t)$ as the *disturbance* function, with $m(t) = E[X(t)]$.

It will be assumed that the mean value function $m(t)$ belongs to a class of functions M , which will we suppose to be a subset of H_R , the reproducing kernel Hilbert space generated by $R(s, t)$ ³.

The following theorem provides the solution for the minimum variance unbiased linear estimate of the mean function $m(t)$.

Theorem 4.3.2. Let the conditions be given as described in the preceding discussions. The following series of assertions provides the complete solution to the linear *smoothing* problem, as stated in this thesis.

1. There exists a bijective mapping $\phi : H_R \rightarrow L^2(X(t) : t \in T)$ with the following properties, for any $t \in T$ and $h, g \in H_R$:

$$\begin{aligned} \phi(R(\cdot, t)) &= X(t) \\ E_m[\phi(h)] &= \langle h, m \rangle_{H_R}, \quad \forall m \in M \\ \text{Cov}[\phi(h), \phi(g)] &= \langle h, g \rangle_{H_R} \end{aligned}$$

where the subscript m indicates that the expectation is taken supposing that $m(t)$ is the true mean function.

2. A random variable $\phi(h)$ belonging to $L^2(X(t), t \in T)$ is said to be an unbiased linear estimator of $m(t)$ at time t provided that

$$E_m[\phi(h)] = \langle h, m \rangle_{H_R} = m(t) \quad \forall m \in M$$

The uniformly minimum variance unbiased linear estimate $\mu(t)$ of $m(t)$ is provided by the following formula

$$\mu(t) = \phi(P_{\mathcal{M}}\{R(\cdot, t)\}),$$

where \mathcal{M} is the smallest Hilbert subspace of H_R containing M , and $P_{\mathcal{M}}\{R(\cdot, t)\}$ is the projection of $R(\cdot, t)$ onto \mathcal{M} .

³Though this Hilbert space has never been explicitly constructed, it may be easily proven that it has the same properties as those indicated in Lemma 4.2.1 and Theorem 4.2.2.

Proof. Since the proper covariance function $R(s, t)$ is well-known to be a positive definite function, there necessarily corresponds a Hilbert space H_R admitting $R(s, t)$ as its reproducing kernel.

To define the required mapping ϕ , we will first define it on the linear subspace spanned by $\mathcal{R} = \{R(\cdot, t) : t \in T\}$, and then extend it to the closure of the span, which by an argument in Theorem 4.2.2., is equal to H_R .

Hence, since any function h in $\text{Sp}(\mathcal{R})$ can be expressed as

$$h = \sum_{i=1}^n \alpha_i R(\cdot, t_i).$$

Let

$$\phi(h) = \sum_{i=1}^n \alpha_i X(t_i).$$

For any mean function $m(t) \in M$, we clearly have

$$\begin{aligned} E_m \left[\left\{ \sum_{i=1}^n \alpha_i X(t_i) \right\}^2 \right] &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j [R(t_i, t_j) + m(t_i)m(t_j)] \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j R(t_i, t_j) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j m(t_i)m(t_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \langle R(\cdot, t_i), R(\cdot, t_j) \rangle_{H_R} \\ &\quad + \left| \langle m(\cdot), \sum_{i=1}^n \alpha_i R(\cdot, t_i) \rangle_{H_K} \right|^2 \\ &= \left\| \sum_{i=1}^n \alpha_i R(\cdot, t_i) \right\|_{H_R}^2 + \left| \langle m(\cdot), \sum_{i=1}^n \alpha_i R(\cdot, t_i) \rangle_{H_K} \right|^2. \end{aligned}$$

Hence, it is clear that $\phi(h) = 0$ if and only if $h = 0$, thus h is necessarily injective. Furthermore, ϕ is clearly surjective on $\text{Sp}(X(t), t \in T)$, and satisfies the first property $\phi(R(\cdot, t)) = X(t)$.

We have that

$$\begin{aligned}
E_m[\phi(h)] &= E_m \left[\sum_{i=1}^n \alpha_i X(t_i) \right] = \sum_{i=1}^n \alpha_i m(t_i) \\
&= \sum_{i=1}^n \alpha_i \langle m(\cdot), R(\cdot, t_i) \rangle_{H_R} \\
&= \langle m(\cdot), \sum_{i=1}^n \alpha_i R(\cdot, t_i) \rangle_{H_R} \\
&= \langle m, h \rangle_H.
\end{aligned}$$

And, for $g = \sum_{j=1}^n \beta_j R(\cdot, t_j)$, we also have that

$$\begin{aligned}
\text{Cov}[\phi(h), \phi(g)] &= \text{Cov} \left[\sum_{i=1}^n \alpha_i X(t_i), \sum_{j=1}^n \beta_j X(t_j) \right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_j \text{Cov}[X(t_i), X(t_j)] \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_j R(t_i, t_j) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_j \langle R(\cdot, t_i), R(\cdot, t_j) \rangle_{H_R} \\
&= \langle \sum_{i=1}^n \alpha_i R(\cdot, t_i), \sum_{j=1}^n \beta_j R(\cdot, t_j) \rangle_{H_R} = \langle h, g \rangle.
\end{aligned}$$

Now, by applying the same argument as in Theorem 4.2.2., the function ϕ may be extended to $\overline{\text{Sp}}(\mathcal{R}) = H_R$, so as to be a bijective mapping from H_R to $L^2(X(t), t \in T)$ with the required properties.

Now, any unbiased linear estimator $\phi(h)$ of $m(t)$ must satisfy, by definition,

$$E_m[\phi(h)] = \langle h, m \rangle_{H_R} = m(t). \quad (1)$$

Yet, by the third property of ϕ , the variance of the estimator ϕ is given by

$$\text{Var}[\phi(h)] = \langle h, h \rangle_{H_R}. \quad (2)$$

Hence, the minimum variance unbiased linear estimator $\mu(t)$ of $m(t)$ consists in finding that function h_μ in H_R which minimizes (1), while satisfying (2).

The solution to this problem is provided by the well-known projection theorem, from which we obtain the required result, viz.

$$\mu(t) = \phi(P_{\mathcal{M}}\{R(\cdot, t)\}).$$

□

4.4 Stationary time series using Hilbert space methods

The previous chapter has conclusively demonstrated that the reproducing kernel Hilbert space approach to time series analysis can provide general solutions to some its most important problems, marked by their surprising simplicity.

All prior developments in this approach have been, for good reason, completely independent of the *stationary* hypothesis, so as to demonstrate that this assumption is not necessary to properly edify a theory of time series.

Yet, in many respects, this framework provides the clearest known view of this hypothesis and its logical consequences, namely by its thorough description of the representation of a time series by a stochastic integral. To be able to demonstrate this point clearly, we will first need to define certain measure theoretic notions.

Definition 4.4.1. Let $\mathcal{M} = (Q, \mathcal{A}, \mu)$ be an arbitrary measure space. We denote by $L_2(Q, \mathcal{A}, \mu)$ the Hilbert space of all \mathcal{A} measurable functions f defined on Q that are square integrable, in the sense that

$$\langle f, f \rangle_{\mathcal{M}} = \int_Q f \bar{f} d\mu < \infty.$$

Definition 4.4.2. Let $\mathcal{M} = (Q, \mathcal{A}, \mu)$ be an arbitrary measure space, and for every $A \in \mathcal{A}$, let $Z(A)$ be a random variable. The collection of all random variables $\{Z(A) : A \in \mathcal{A}\}$ is said to an *orthogonal random measure* with covariance function μ if for any $A_1, A_2 \in \mathcal{A}$,

$$E[Z(A_1)Z(A_2)] = \mu(A_1 \cap A_2).$$

The Hilbert space $L^2(Z(A), A \in \mathcal{A})$ spanned by a orthogonal random measure is defined as the closed linear subspace $\overline{\text{Sp}}(Z(A), A \in \mathcal{A})$ contained in $L^2(Q, \mathcal{A}, \mu)$.

As is well-known in measure theory, the Hilbert space $L^2(Q, \mathcal{A}, \mu)$, as described above, may equivalently be considered to be spanned by the set of indicator functions $\{I_A : A \in \mathcal{A}\}$, where for any point $q \in Q$

$$I_A(q) = \begin{cases} 1 & \text{if } q \in A \\ 0 & \text{if } q \notin A \end{cases}$$

From this remark, we immediately deduce the following lemma:

Lemma 4.4.1. Let $\mathcal{M} = (Q, \mathcal{A}, \mu)$ be an arbitrary measure space equipped with an orthogonal random measure $\{Z(A) : A \in \mathcal{A}\}$ with a covariance function μ . Then, there exists a isometry $\phi : L^2(Q, \mathcal{A}, \mu) \rightarrow L^2(Z(A), A \in \mathcal{A})$ such that for any $A \in \mathcal{A}$,

$$\phi(I_A) = Z(A)$$

Proof. Let us note that, for any $A_1, A_2 \in \mathcal{A}$,

$$\langle I_{A_1}, I_{A_2} \rangle_{\mathcal{M}} = \int_Q I_{A_1} I_{A_2} d\mu = \mu(A_1 \cap A_2) = E[Z(A_1)Z(A_2)]$$

Therefore, Theorem 4.2.2 is applicable, and thus there exists a isometry ϕ with the required property. \square

This lemma allows us to rigourously define the stochastic integral of a measurable function with respect to a random orthogonal measure, as follows

Definition 4.4.3. Let $\mathcal{M} = (Q, \mathcal{A}, \mu)$ be an arbitrary measure space equipped with an orthogonal random measure $\{Z(A) : A \in \mathcal{A}\}$ with a covariance function μ . For any function $f \in L^2(Q, \mathcal{A}, \mu)$, the stochastic integral of f with respect to $\{Z(A) : A \in \mathcal{A}\}$, is defined as

$$\int_Q f dZ = \phi(f),$$

where ϕ is the isometry between $L^2(Q, \mathcal{A}, \mu)$ and $L^2(Z(A), A \in \mathcal{A})$ defined in the preceding lemma.

This stochastic integral obeys the following formal laws, for any functions $f, g \in L^2(Q, \mathcal{A}, \mu)$,

$$E\left[\left(\int_Q f dZ\right)\left(\int_Q \bar{g} dZ\right)\right] = \int_Q f \bar{g} d\mu$$

$$E\left[\left\{\int_Q f dZ - \int_Q g dZ\right\}\left\{\int_Q \bar{f} dZ - \int_Q \bar{g} dZ\right\}\right] = \int_Q |f - g|^2 d\mu$$

The following theorem, due to Karhunen [22], gives a necessary condition for a given time series $\{X(t), t \in T\}$ with finite second moments with covariance function $K(s, t)$ to be expressible as a stochastic integral.

Theorem 4.4.1. Let $\{X(t), t \in T\}$ be a time series with finite second moments and covariance function $K(s, t)$, defined on the probability space $\{\Omega, \mathcal{F}, P\}$. Let $\mathcal{M} = (Q, \mathcal{A}, \mu)$ be a measure space such that the dimension of $L^2(\Omega, \mathcal{A}, P)$ is greater or equal to the dimension of $L^2(Q, \mathcal{A}, \mu)$.

If there exists a set $\{f(t, \cdot) : t \in T\}$ of elements in $L^2(Q, \mathcal{A}, \mu)$ such that⁴

$$K(s, t) = \int_Q f(s, \lambda) f(t, \lambda) d\mu(\lambda), \quad (3)$$

then there exists an orthogonal random measure $\{Z(A), A \in \mathcal{A}\}$ with covariance kernel μ such that, for any $t \in T$,

$$X(t) = \int_Q f(t, \lambda) dZ(\lambda).$$

Proof. For purposes of notational brevity, we will designate $f(t, \cdot)$ by $f(t)$, as is commonly done for time series or stochastic processes.

From (3), we may apply Theorem 4.2.2. to deduce the existence of a isometry $\phi : L^2(f(t), t \in T) \rightarrow L^2(X(t), t \in T)$, satisfying $\phi(f(t)) = X(t)$

Now, if $L^2(f(t), t \in T) \neq L^2(Q, \mathcal{A}, \mu)$, it is clear we can complete, by constructing the appropriate Cauchy sequences, the isometry ϕ so that it maps any $f \in L^2(Q, \mathcal{A}, \mu)$ to a subspace of $L^2(\Omega, \mathcal{A}, P)$ containing $L^2(X(t), t \in T)$, while retaining the property $\phi(f(t)) = X(t)$.

Moreover, let $Z(A) = \phi(I_A)$ for any $A \in \mathcal{A}$, from which it follows that $E[Z(I_{A_1})Z(I_{A_2})] = \mu(A_1 \cap A_2)$ for any $A_1, A_2 \in \mathcal{A}$. Thus, this implies that $\{Z(A) : A \in \mathcal{A}\}$ is an orthogonal random measure with covariance kernel μ .

Hence, the result is immediate, for $X(t) = \phi(f(t))$ with the corresponding orthogonal random measure $\{Z(A) : A \in \mathcal{A}\}$. \square

Now, if the covariance kernel $K(s, t)$ of a time series $\{X(t) : t = 0, \pm 1, \dots\}$ is assumed to depend upon the difference of its arguments, i.e.

$$K(s, t) = K(s - t),$$

then we know, by the discrete version of the Bochner-Khinchin theorem⁵, that there will necessarily exist some non-decreasing bounded function $F(\lambda)$ defined over $[-\lambda, \lambda]$ such that the following representation holds

$$K(s - t) = \int_{-\pi}^{\pi} e^{i(s-t)\lambda} dF(\lambda).$$

⁴This condition is equivalent to assuming $\{f(t, \cdot) : t \in T\}$ forms a representation of $\{X(t) : t \in T\}$ on $L^2(Q, \mathcal{A}, \mu)$

⁵We refer to the fourth section of the textbook of Brockwell and Davis [8] for a proper statement and proof of this elementary theorem.

Combining this with the result obtained in the previous theorem, we obtain the spectral representation for stationary time series, obtained in this form by Parzen [29].

Theorem 4.4.2. A discrete time series $\{X(t) : t = 0, \pm 1, \dots\}$ with covariance function $K(s, t)$ is stationary if and only if for some Lebesgue measure μ defined on the interval $Q = \{\lambda : -\pi \leq \lambda \leq \pi\}$ the complex exponentials $\{e^{i\lambda t} : t = 0, \pm 1, \dots\}$ constitute a representation in $L^2(Q, \mathcal{A}, \mu)$, where \mathcal{A} is the Borel σ -algebra of subsets of Q .

Then, there exists a orthogonal random measure $\{Z(A) : A \in \mathcal{A}\}$ such that

$$X(t) = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda).$$

4.5 Sampling theorems and reproducing kernels

Though not so well-known in the statistical litterature, a problem of a fundamental concern arising in the analysis of empirical continuous time series is that of proper sampling, being informally defined as that which *captures* all the information.

If we consider any continuous time series $\{X(t) : t \in \mathbb{R}\}$ as being given, and we suppose the samples to be taken at a fixed sample rate h , it is clear the problem will admit of a solution if there exists a functional Φ such that

$$X(t) = \Phi(n, h, t, X_0, X_{\pm 1/h}, X_{\pm 2/h}, \dots).$$

Thus, the problem reduces to determining those conditions that may be imposed on the time series $\{X(t) : t \in T\}$ for such a functional Φ to exist.

In the context of a deterministic function $f(t)$ of a real variable, the solution to this problem is provided by the *sampling* theorem, a well-known result of Whittaker, and later popularized by Shannon⁶, which states that if f admits an inverse Fourier transform representation, for some positive constant h ,

$$f(t) = \int_{-\pi h}^{\pi h} e^{i\lambda t} dg(\lambda), \tag{4}$$

where g is a square-integrable function on $[\pi h, \pi h]$, then $\phi(t)$ will admit the representation

⁶For a brief historical discussion and well-written presentation of the classical results associated to this problem, consult the work of Higgins [17]

$$f(t) = \sum_{n=-\infty}^{n=\infty} f\left(\frac{n}{h}\right) \frac{\sin \pi(ht - n)}{\pi(ht - n)}.$$

Many authors, including Lloyd [23] and Parzen [28], were able to demonstrate, by using the spectral representation theorem given in this previous section, that this results easily extends to stationary time series. That is, if $\{X(t) : t \in T\}$ is a stationary time series with covariance function $K(s, t) = K(s - t)$ admitting the spectral representation

$$K(s - t) = \int_{\pi h}^{\pi h} e^{i(s-t)\lambda} dF(\lambda),$$

then $X(t)$ will have the following mean-square convergent expansion

$$X(t) = \sum_{n=-\infty}^{n=\infty} X\left(\frac{n}{h}\right) \frac{\sin \pi(ht - n)}{\pi(ht - n)}.$$

Various generalizations of the solution given above were developed in the literature, such as allowing for a time-varying sampling rate or in replacing the Fourier transform for another transform, such as the Sturm-Liouville or the Jacobi transform.⁷

After many of these developments were published, it was discovered by Yao [35] that the essence of these *sampling* theorems lied in the reproducing kernel associated to the corresponding space of functions considered. In particular, it was shown that the space of square integrable functions $f(t)$ admitting an inverse Fourier transform representation defined in (4) forms a Hilbert space for which the sinc function

$$w(\tau, n) = \frac{\sin \pi(\tau - n)}{\pi(\tau - n)}$$

is its reproducing kernel, with $\tau = ht$, and which, when taken in conjunction, forms a complete orthonormal set in the corresponding space.

Various authors were led to derive general *sampling* theorems in reproducing kernel Hilbert space, notably Higgins [18] and Nashed-Walter [27], which, in their simplest form, reduce to the theorem given below, which depends upon the following elementary lemma.

⁷Consult the paper of Jerri [21] for a complete survey of all known extensions.

Lemma 4.5.1. Let H be a Hilbert space of functions admitting the reproducing kernel $K(s, t)$, and $\{f_n\}$ be a sequence of functions converging pointwise to f . Then, $\{f_n\}$ converges in norm to f , and this convergence is uniform on every subset where $K(t, t)$ is bounded.

Proof. By the reproducing property of K , we have that

$$f(t) - f_n(t) = \langle f - f_n, K(\cdot, t) \rangle_H$$

From which we deduce, by Cauchy-Schwarz inequality and the third assertion of Theorem 2.2.1, that

$$\begin{aligned} |f(t) - f_n(t)| &= |\langle f - f_n, K(\cdot, t) \rangle_H| \\ &\leq \|f - f_n\|_H \|K(\cdot, t)\|_H \\ &= \|f - f_n\|_H K(t, t)^{\frac{1}{2}}, \end{aligned}$$

from which the theorem immediately follows. \square

Theorem 4.5.1. Let H be a Hilbert space of functions defined on a set X admitting the reproducing kernel $K(s, t)$. If there exists a sequence $\{t_n\}_{n=1}^{\infty}$ of points in X such that $\{K(\cdot, t_n)\}_{n=1}^{\infty}$ is an orthonormal basis for H , then any function $f \in H$ can be expressed as

$$f(t) = \sum_{n=1}^{\infty} f(t_n) \frac{K(t, t_n)}{K(t_n, t)},$$

with uniform convergence on those subsets of X for which $K(t, t)$ is bounded.

Proof. Since $\|K(\cdot, t_n)\| = \langle K(\cdot, t_n), K(\cdot, t_n) \rangle = \sqrt{K(t_n, t_n)}$, it follows that the set $\{K(\cdot, t_n)/\sqrt{K(t_n, t_n)}\}$ is an orthonormal basis for H .

Thus, we obtain, for any $f \in H$, the following representation

$$f = \sum_{n=1}^{\infty} \langle f, \frac{K(\cdot, t_n)}{\sqrt{K(t_n, t_n)}} \rangle_H \frac{K(\cdot, t_n)}{\sqrt{K(t_n, t_n)}} = \sum_{n=1}^{\infty} f(t_n) \frac{K(\cdot, t_n)}{K(t_n, t)},$$

where the condition of uniform convergence follows from the lemma above. \square

5 Conclusion

In this thesis, we aimed to demonstrate that the theory of reproducing kernel Hilbert spaces could serve as a genuine foundation upon which a general theory of time series analysis could be constructed. Though incomplete, and lacking in certain details, it could be well argued that the developments previously given could be extended so as to reinforce the foundation at which it aims to be.

As mentioned in the previous chapter, the theory exposed is certainly lacking in one fundamental area, namely that of computability. The various expressions and functions which it consider are by no means simple to compute, whether numerically or algebraically. Thus, to those interested in purely practical subject matter, this may indeed serve as a considerable deterrent for its consideration. Yet, in the author's opinion, that a mathematical theory is *computable* at any given point in time is not necessarily intrinsic to the theory itself, but rather highly dependent upon the current stage of our analytical abilities. Thus, a future area of research that would provide tremendous practical value to the theory of reproducing kernels would be that which develops its numerical approximations.

Another point that may be made, from a theoretical point of view, is that the theory expounded previously limits itself to the context of Hilbert spaces, while it is well-known that there are known spaces far greater in generality, such as topological vector spaces. And thus, an obvious generalization would be to determine if the theory of reproducing kernels can be extended to these spaces. The first step in this direction has been taken by Schwartz [33], and it remains to be seen if this generalized theory can be applied to time series analysis.

References

- [1] Nachman Aronszajn. La théorie des noyaux reproduisants et ses applications première partie. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 39, pages 133–153. Cambridge Univ Press, 1943.
- [2] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.
- [3] Stefan Bergman. *The kernel function and conformal mapping*. Number 5. American Mathematical Soc., 1970.
- [4] Stefan Bergman and Menahem Schiffer. *Kernel functions and elliptic differential equations in mathematical physics*. Courier Corporation, 2013.
- [5] Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
- [6] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.
- [7] Salomon Bochner. *Lectures on Fourier integrals*. Number 42. Princeton University Press, 1959.
- [8] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer, 2013.
- [9] Bruce L Chalmers. Images of bounded linear operators on a hilbert space with kernel function. *Archiv der Mathematik*, 26(1):313–316, 1975.
- [10] John B. Conway. *Functions of One Complex Variable I*. Springer, New York, 1978.
- [11] John B. Conway. *A course in functional analysis*, volume 96. Springer Science & Business Media, 2013.
- [12] Philip J Davis. *Interpolation and approximation*. Courier Corporation, 1975.
- [13] Peter L Duren and Alexander Schuster. *Bergman spaces*. Number 100. American Mathematical Soc., 2004.

- [14] Robert Fortet. *Vecteurs, fonctions et distributions aleatoires dans les espaces de Hilbert: Analyse harmonique & prevision*. Hermès, 1995.
- [15] Brian C. Hall. Holomorphic methods in analysis and mathematical physics. *Contemp. Math*, 260.
- [16] Brian C. Hall. *Quantum theory for mathematicians*. Springer, 2013.
- [17] John R Higgins. Five short stories about the cardinal series. *Bulletin of the American Mathematical Society*, 12(1):45–89, 1985.
- [18] John Rowland Higgins. An interpolation series associated with the bessel–hankel transform. *Journal of the London Mathematical Society*, 2(4):707–714, 1972.
- [19] Einar Hille. Introduction to general theory of reproducing kernels. *Journal of mathematics*, 2(3), 1972.
- [20] Vasile Ion Istratescu. *Inner product structures: theory and applications*, volume 25. Springer Science & Business Media, 2012.
- [21] Abdul J Jerri. The shannon sampling theorem – its various extensions and applications: A tutorial review. *Proceedings of the IEEE*, 65(11):1565–1596, 1977.
- [22] Kari Karhunen. *On linear methods in probability theory*. Rand Corporation, 1960.
- [23] SP Lloyd. A sampling theorem for stationary (wide sense) stochastic processes. *Transactions of the American Mathematical Society*, 92(1):1–12, 1959.
- [24] M Loève. *Probability theory*. Springer-Verlag, New York, 1977.
- [25] James Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character*, 209:415–446, 1909.
- [26] Eliakim Hastings Moore. *General Analysis*. Philadelphia, the American Philosophical Society, 1935.
- [27] M Zuhair Nashed and Gilbert G Walter. General sampling theorems for functions in reproducing kernel hilbert spaces. *Mathematics of Control, Signals and Systems*, 4(4):363–390, 1991.

- [28] Emanuel Parzen. A simple proof and some extensions of the sampling theorem. Technical report, DTIC Document, 1956.
- [29] Emanuel Parzen. *Statistical inference on time series by Hilbert space methods*. Stanford Univ., 1959.
- [30] Emanuel Parzen. *Time series analysis papers*. Holden-Day, 1967.
- [31] Vladimir Semenovich Pugachev. *Theory of Random Functions: And Its Application to Control Problems*. Pergamon Press, 1965.
- [32] Saburo Saitoh. *Theory of reproducing kernels and its applications*, volume 189. Longman, 1988.
- [33] Laurent Schwartz. Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *Journal d'analyse mathématique*, 13(1):115–256, 1964.
- [34] Grace Wahba et al. Support vector machines, reproducing kernel hilbert spaces and the randomized gacv. *Advances in Kernel Methods-Support Vector Learning*, 6:69–87, 1999.
- [35] Kung Yao. Applications of reproducing kernel hilbert spaces–bandlimited signal models. *Information and Control*, 11(4):429–444, 1967.