https://doi.org/10.26219/heal.aueb.6279

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

REPRODUCING KERNEL HILBERT SPACES AND THEIR APPLICATIONS IN PROBABILITY AND STATISTICS

By

Demetrios Athanasios Dristellas

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
July 2015

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ

ΧΩΡΟΙ HILBERT ΜΕ ΑΝΑΠΑΡΑΓΩΓΙΚΟ ΠΥΡΗΝΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ ΣΤΗΝ ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΣΤΙΣ ΠΙΘΑΝΟΤΗΤΕΣ

Δημήτριος Αθανάσιος Δριστέλλας

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Ιούλιος 2015

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

DEDICATION

To my family and my beloved friends.

https://doi.org/10.26219/heal.aueb.6279

ACKNOWLEDGEMENTS

First and foremost, my utmost gratitude to mr. Athanasios Giannakopoulos, Professor at Athens University of Economics and Business, department of Statistics, whose sincerity and encouragement I will never forget.

My co-postgraduate students for this wonderful experience. And last but not least important, my family and friends for their moral support, by giving me strength to plod on despite all the difficulties I faced during the program.

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

VITA

After graduating from high school in 2007 I studied mathematics at the University of Crete, department of Mathematics where I was graduated in 2013. Almost immediately after I was accepted in the Athens University of Economics and Business at the full time postgraduate program of Statistics.

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

ABSTRACT

Demetrios Dristellas

TITLE Reproducing Kernel Hilbert spaces and their Applications In Probability and Statistics

July 2015

In a large variety of problems in Statistics and Stochastic processes, the random variables which are used do not present finite dimensionality, in contrary their dimension is infinite. On the other hand, the observations of those random variables are actually finite dimensional approximations of corresponding infinite dimensional subjects. Reproducing Kernel Hilbert Spaces (RKHS) are a useful theoretical and practical tool which provides us a series useful representations even for non-linear data. This work will be an introduction to the theory of RKHS in the context of the smoothness of a data set.

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

ΠΕΡΙΛΗΨΗ

Δημήτριος Δριστέλλας

Χώροι Hilbert με Πυρήνα Αναπαραγωγής και Εφαρμογές τους στην Στατιστική και στις Στοχαστικές Διαδικασίες

Ιούλιος 2015

Σε πολλές εφαρμογές οι τυχαίες μεταβλητές τις οποίες χρησιμοποιούμε δεν είναι πεπερασμένης διάστασης αλλά απειροδιάστατες. Από την άλλη βέβαια οι παρατηρήσεις αυτών των τυχαίων μεταβλητών γίνονται με διακριτό τρόπο και ουσιαστικά αποτελούν (πεπερασμένης διάστασης) προσεγγίσεις των πραγματικών απειροδιάστατων αντικειμένων. Οι χώροι Hilbert με πυρήνα αναπαραγωγής αποτελούν ένα ενδιαφέρον θεωρητικό αλλα και πρακτικό εργαλείο για την αναπαράσταση απειροδιάστατων δεδομένων και μας επιτρέπουν να πάρουμε γραμμικές αναπαραστάσεις ακόμη και για δεδομένα τα οποία είναι μη γραμμικά. Η εργασία αυτή θα αποτελεί μια εισαγωγη στην θεωρία και τις εφαρμογές των RKHS με απώτερο σκοπό την σύνδεση του με το θεμα της ομαλοποίησης των δεδομένων.

https://doi.org/10.26219/heal.aueb.6279

Contents

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

Chapter 1

Introduction

1.1 Hilbert Spaces and their usefulness

In the field of Applied Mathematics and especially in the field of Statistics, scientists are dealing with problems which are closely related with notions such as distance, angles, vectors or some kind of transformations, like a projection of a data set into a much more ”convenient” space which is equipped with tools which make a problem easier to manipulate. All of those notions are associated with a well known and widespread field, which is known as Euclidean geometry. A natural way for the treatment of such problems, is through a mathematical model which enables us to generalize the concepts of the Euclidean geometry into much more abstract sets, such as spaces of functions, etc. Hilbert spaces are such sets, which are actually the generalization of the finite Euclidean space, to some infinite dimensional set.

The mathematical objects that we now call Hilbert spaces did not begin with David Hilbert (1862-1943). Much of the theory was developed for solving physical problems such as integral equations of the form

$$x(s) - \int_0^1 K(s,t)x(t)dt = f(s)$$

for known kernel $K(s,t)$ and function $f(s)$. Contributions to the solutions of such problems trace from Daniel Bernoulli, through H. A. Schwartz, E.I. Fredholm, and others, to D. Hilbert, who expanded the solution in terms of linear combinations of the eigenfunctions, an orthogonal basis for the Hilbert space. The modern theory of Hilbert spaces owes much to J. Von Neumann.

Many geometric objects such as lines, planes or spheres have standard extensions to Hilbert spaces. The tools provided by them allows us to handle mathematical objects in the same manner as we do with points or vectors. For example, similarly with points or vectors that lie in an Euclidean space $\mathbb{R}^n$ we can measure the distance between two functions $f,\ g$ and a function g belongs to a Hilbert Space H through its corresponding inner product or finding their projections into a subspace of H. Hilbert

https://doi.org/10.26219/heal.aueb.6279

spaces do not provide new tools, but they show how simple and familiar tools can be employed to face wide classes of problems.

1.2 A brief introduction to RKHS

Among different classes of Hilbert spaces, Reproducing Kernel Hilbert Spaces (RKHS) are the most natural extension for the actual space that we live in, the Euclidean space $\mathbb{R}^3$. We assume that we study the elements of some abstract set A. This set could have many problems, for example non suitable topological properties or undefined ordinance. A good way for someone to surpass this, is finding another set S' with the proper structure for the treatment of the problem and consider the elements of S as elements of S'. For this purpose, we find ourselves in need for some kind of an ”imbedding” theorem or a ”representation” theorem, through which we can map the elements from S to their representers in S'. We will see that a Reproducing Kernel Hilbert Space provides us with some powerful tools and geometric concepts for many types of such problems. An important notion of a Reproducing Kernel Hilbert Space of functions on a set E is that it can be characterized only through a symmetric positive type function K on $E \times E$. This looming correspondence between positive type functions and the Reproducing Kernels plays critical role in the construction of these particular spaces. In addition, K possess the so called ”reproducing property” i.e. K can reproduce any value of a function on E through the corresponding inner product of H.

The notion of a reproducing kernel was first introduced in the 1907 work of Stanisław Zaremba concerning boundary value problems for harmonic and biharmonic functions. James Mercer simultaneously examined functions which satisfy the reproducing property in the theory of integral equations. The idea of the reproducing kernel remained untouched for nearly twenty years until it appeared in the dissertations of Gábor Szegő, Stefan Bergman, and Salomon Bochner. Polish mathematician Nachman Aronszajn (1907–1980)(Aronszajn 1950) developed the notion of Reproducing kernel Hilbert space. Some facts are also attributed to another Polish mathematisian, Stefan Bergman (1895–1977). In subsequent years, Poggio and Girosi (1989) introduced Tikhonov regularization in learning theory and worked with RKHS only implicitly, because they dealt mainly with hypothesis spaces on unbounded domains. Moreover ,RKHS were explicitly introduced in learning theory by Girosi (1997). RKHS were used much earlier in approximation theory (eg Wahba, 1990) and computer vision (eg Bertero, Torre, Poggio, 1988)(source https://en.wikipedia.org/wiki/Reproducing_kernel_Hilbert_space)

1.3 Main objectives of this thesis

This thesis provides an introduction of the constructive properties of the Reproducing Kernel Hilbert Spaces.

https://doi.org/10.26219/heal.aueb.6279

Especially in Chapter 2, we will present some useful facts of the reproducing kernels. We will try to bring out the most useful ways of their construction through some interesting tools such as specific types of functions or the eigenvalues of an operator generated by them.
In Chapter 3 we will demonstrate the most important consequences of this particular type of Hilbert spaces and we will highlight some applications regarding Gaussian stochastic properties.
In Chapter 4, we will present a brief introduction to ill-posed and well-posed linear operator equations whereby we can model many types of problems in Statistics. After providing the basic tools needed, we will develop the representer theorem which is the most important theorem of this thesis with many applications in data sciences.
Focusing on that and to the topological properties of the RKHS, in Chapter 5, we will present a particular type of handling functional data. In order do more precise, we will develop a particular type of minimizers, regarding functional linear regression which will allow us to extract accurate measures under much more general conditions than the usual way of regression.

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

Chapter 2

Basic definitions and construction of RKHS

The aim of this chapter is to introduce the notion of a reproducing kernel but also to prove some useful theorems and facts that will lead to the construction of another space of functions named Reproducing Kernel Hilbert Space. We will follow closely Berlinet and Thomas-Agnan, 2011

2.1 Notations and examples of Hilbert spaces

Let E be an non empty set. Let H be a vector space of functions defined on E with values in $\mathbb{C}$. H is endowed with the structure of Hilbert space defined by an inner product $< .,. >_H$.

$$H \times H \to \mathbb{C}$$

Let $\|.\|_H$ denote the associated norm:

$$\forall \phi \in H, \|f\|_H =< \phi, \phi >_H^{\frac{1}{2}}$$

.

For any t$\in$ E, we will denote by e_t the evaluation functional at the point t, i.e. the mapping

$$H \to \mathbb{C}$$

$$g \to e_t(g) = g(t)$$

The set of complex functions defined on E will be denoted by $\mathbb{C}^E$

The examples given below will be very helpful in the understanding of the structure of such spaces H with the above properties.

https://doi.org/10.26219/heal.aueb.6279

Example 2.1.1. *Let H be a finite dimensional complex vector space of functions with basis $(f_1, ..., f_n)$.Any vector of H can be written in a unique way as a linear combination of $f_1, ..., f_n$.Therefore an inner product $< ., . >_H$ on H is entirely defined on H by the numbers*

$$g_{ij} =< f_i, f_j >, 1 \leq i, j \geq n$$

if

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

then

$$< v, w >=< \sum_{i=1}^{n} v_i f_i, \sum_{j=1}^{n} w_j f_j >= \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \bar{w}_j g_{ij}$$

The matrix $G = (g_{ij})$ is called the Gram matrix of the basis. G is Hermitian (observe that $g_{ij} = \bar{g}_{ji}$) and positive definite ($v^\star G v = \| \sum_{i=1}^{n} v_i f_i \|_H > 0$)

Furthermore we denote that any finite dimensional space endowed with any inner product is complete (i.e. every Cauchy sequence is convergent).

Example 2.1.2. *Let $E = (a, b)$, $-\infty \leq a < b \leq \infty$ and $\mathscr{L}^2(a, b)$ be the set of all complex measurable functions over (a, b) such that*

$$\int_a^b |f(x)|^2 d\lambda(x) < \infty$$

where $\lambda(x)$ is the Lebesque measure on $\mathbb{R}$ Identifying two functions f, g on $\mathcal{L}^2(a, b)$ which are equal except on a set of Lebesque measure equal to zero, we get a vector space $\mathcal{L}^2(a, b)$ which is Hilbert space with inner product

$$< f, g >_{\mathscr{L}^2(a,b)}= \int_a^b f(x)\overline{g(x)} d\lambda(x)$$

Functions belonging on $\mathcal{L}^2(a, b)$ endowed with the inner product defined above will be further examined in Chapter 4.

Example 2.1.3. *Let $E = \mathbb{R}$ and*

$$H = W_2^1(\mathbb{R}) = \{\phi : \ \phi \text{ is absolutely continuous}, \ \phi \text{and } phi' \in \mathscr{L}^2 \ \}$$

where ϕ' is the derivative of ϕ almost everywhere. H is a Hilbert space with the inner product

$$< \phi, \psi >_H= \int_{\mathbb{R}} (\phi\bar{\psi} + \phi'\bar{\psi}') d\lambda \tag{2.1}$$

https://doi.org/10.26219/heal.aueb.6279

Example 2.1.4. *Let* (Ω, A, P) *be a probability space, Let* $\mathcal{F}$ *be a Hilbert space with inner product* $< .,. >_{\mathcal{F}}$ *and norm* $\|.\|_{\mathcal{F}}$ *and let* $\mathscr{L}^2(\Omega, A, P)$ *be the set of random variables* X *with values in* F *such that*

$$E_P(\|X\|^2_{\mathcal{F}}) = \int \|X\|^2_{\mathcal{F}} dP < \infty. \tag{2.2}$$

Identifying two random variables X *and* Y *such that* $P(X \neq Y) = 0$ *we get the space* $\mathscr{L}^2(\Omega, A, P)$ *which is a Hilbert space when endowed with the inner product*

$$< X, Y >= E_P(< X, Y >_{\mathcal{F}})$$

2.2 Definitions and facts regarding Reproducing Kernels

Let us now introduce the the definition of reproducing kernel

Definition 2.2.1. *A function*

$$K : E \times E \to \mathbb{C}$$
$$(s, t) \to K(s, t)$$

is called a reproducing kernel of the Hilbert space H *if and only if*

a) $\forall t \in E$, $K(., t) \in H$

b) $\forall t \in E$, $\forall \phi \in H < \phi, K(., t) >= \phi(t)$

The last condition is commonly called the "reproducing property" of K. Every value t of a function $\phi \in H$ can be "reproduced" by the inner product of ϕ with K(.,t). By conditions a),b) follows that $K(s, t)$ can be written as $K(s, t) =< K(., s), K(.t) >$, $t \in E$.

A Hilbert space of complex valued functions which possesses a reproducing kernel is called a REPRODUCING KERNEL HILBERT SPACE (RKHS).

We will present some examples of Hilbert spaces H and see how a function K which satisfies conditions a) and b) influences with its elements.

Example 2.2.2. *Let* $(e_1, ..., e_n)$ *be an orthonormal basis in* H *and define*

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

 https://doi.org/10.26219/heal.aueb.6279

Then for any y in E,

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

belongs to H and for any function

$$\phi(.) = \sum_{i=1}^{n} \lambda_i e_i(.)$$

in H, we have

$$\begin{aligned}\forall y \in E, \ < \phi, K(.,y) >_H &= < \sum_{i=1}^{n} \lambda_i e_i(.), \sum_{i=1}^{n} e_i(.)\bar{e}_i(y) >_H \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \bar{\bar{e}}_j(y) < e_i, e_j >_H \\ &= \sum_{i=1}^{n} \lambda_i e_i(y) = \phi(y)\end{aligned}$$

Any finite dimensional Hilbert space of functions possesses a reproducing kernel

Example 2.2.3. *Let $K(i,j) = \delta_{ij}$ the Kronecker delta function,*

where $\delta_{ij} = \begin{cases} 1 \, if \ i = j \\ 0 \, else \end{cases}$ *Then*

$\forall \ j \in \mathbb{N}, \ K(.,j) \ = \ (0,0,..,1,,..) \in H$ *(1 at the j-th place)*

$\forall \ j \in \mathbb{N} \ \forall \ (x_i)_{i \in \mathbb{N}} \in H \ < x, K(.,j) >_H = \sum_{i=1}^{n} x_i \bar{\delta_{ij}} = x_j$

K is the reproducing kernel of H

Example 2.2.4. *A set of functions of particular interest especially in the field of statistics, is the space $\mathscr{L}^[a,b]$. We are in interest of finding a function $K(.,t)$ in (a,b) such that*

$$\forall \phi \in (a,b) \qquad < \phi, K(.,t) >_{\mathscr{L}^2} \int_{[a,b]} \phi \overline{K(.,t)} d\lambda = \phi(t) \tag{2.3}$$

A function that one could natural think is the Dirac delta function since it possesses the reproducing property

$$f(x) = \int_{[a,b]} \delta(x-u) f(u) du$$

https://doi.org/10.26219/heal.aueb.6279

But unfortunately this function is not square integrable since

$$\int_{[a,b]} \delta^2(u)du \nless \infty$$

The main problem of this space of regards on what does it mean for a function f and a function g to be equal. In this kind of spaces, a function $f \in \mathscr{L}^[a,b]$ is not individual but a whole class of functions. This comes from the fact that this space enables us to handle its elements with a much weaker condition, the almost everywhere condition. This allows a function $f \in \mathscr{L}^[a,b]$ to be equal to any g with the same domain, except on some set A with $\lambda(A) = 0$. Therefore Definition 2.2.1 does not clearly apply in this case. Moreover this space or a subset of it, could contain functions with undesirable properties to work with, for example discontinuity. Concluding the space $\mathscr{L}^2$ is not an RKHS

A natural question that comes out directly from Definition 1 is that when a Hilbert space H possesses a function K with properties a) and b).The answer arises from Theorem 2.2.5

Theorem 2.2.5. *A Hilbert space of complex valued functions on E has a reproducing kernel if and only if all the evaluation functionals e_t , $t \in E$ are continuous in H*

Proof. If H has a reproducing kernel K then for any $t \in E$, we have

$$\forall\, t \in \; H \quad e_(\phi) = <\phi, K(.,t)>_H .$$

thus the evaluation functional e_t is linear and, by Cauchy-Schwartz inequality, continuous

$|e_t(\phi)| = | <\phi, K(.,t)>_H | \leq \|\phi\|\|K(.,t)\| = |\phi\|[K(t,t)]^{1/2}$. Moreover, for $\phi = K(.,t)$, the upper bound is obtained so that the norm of the continuous linear functional e_t is given by

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

Conversely by Riesz's representation theorem, if the linear mapping

$$H \to \mathbb{C}$$

$$\phi \to e_t(\phi) = \phi(t)$$

is continuous then there exist $N_t \in H$ such that

$$\forall\, \phi \in H \;\; <\phi, N_t>= \phi(t).$$

https://doi.org/10.26219/heal.aueb.6279

If this property holds for any $\in E$, then it is clear that$K(s,t) = N_t(s)$ is the reproducing kernel of H. □

Lemma 2.2.6 is a direct consequence of the continuity of the evaluation functionals e_t on a RKHS H

Lemma 2.2.6. *In an RKHS a sequence converging in the norm sense converges pointwise to the same limit*

Proof. If (ϕ_n) converges to ϕ in the norm sense we have for any $t \in E$,

$$|\phi_n(t) - \phi(t)| = |e_t(\phi(n)) - e_t(\phi)|$$

and $e_t(\phi_n)$ converges to $e_t(\phi)$ by continuity of e_t □

The above property is not generally hold true in an abstract space H. Consider for example, the space of polynomials over $[0,1]$ endowed with the $\mathscr{L}^p_{\mathcal{P}}[0,1]$ metric

$$d(P_1, P_2) = \left(\int_0^1 |P_1(x) - P_2(x)|^p d\lambda(x)\right)^{1/p}$$

where $p > 0$ and λ is the Lebesque measure on the set R. Let$(Q_n)_{n\in\mathbb{N}}$ be the sequence of polynomials in $\mathscr{L}^p_{\mathcal{P}}[0,1]$ with $Q_n(x) = x^n$. Then,

$$d(Q_n, 0) = \left(\int_0^1 x^{np} d\lambda\right)^{1/p} = (np+1)^{-1/p} \to 0$$

as n tends to infinity. But the sequence $(Q_n(1))_{n\geq 0} = 1$ and therefore lemma 2.2.6 does not hold.

The property that if two functions f and g are close in the norm sense, then the values $f(x)$ and are close to the values $g(x)$ is one of the most important facts in RKHS theory and one of the main reasons that these particular spaces are used constantly, with wide variety of applications especially in statistics and stochastic processes. This is not generally hold true for an abstract Hilbert space

2.3 Reproducing kernels and positive type functions

We now introduce a class of functions named by positive type (or positive definite) functions, with many important properties in a wide range of sectors in mathematics. In this section we will consider the most important of them in the context of our subject. We will prove that there is an one to one correspondence to set of positive type functions and the set of reproducing kernels. This result will turn to be very helpful

 https://doi.org/10.26219/heal.aueb.6279

not only on proving whether a complex valued function K is a reproducing kernel of a Hilbert space H, but also in constructing RKHS.

Let us now give the definition of a positive type function.

Definition 2.3.1. *A function $K : E \times E \to \mathbb{C}$ is called a positive type type function (or positive definite function) if*

$$\forall n \geq 1,\ \forall(a_1, ..., a_n) \in \mathbb{C},\ \forall(x_1, ..., x_n) \in E^n,$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\bar{a}_j K(x_i, x_j) \geq 0 \tag{2.4}$$

Moreover with the assumptions made in Example 2.1.1 is not hard to see that since

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\bar{a}_j K(x_i, x_j) = [a]'_n[K(x_i, x_j)]_{ij}[a]_n$$

the above condition in Definition 2.3.1 is equivalent to show that the matrix

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

is positive definite.
We will now present some examples of positive type functions

Example 2.3.2. *Any constant non negative function on $E \times E$ is positive type since*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\bar{a}_j \geq 0\ = \left|\sum_{i=1}^{n} a_i\right|^2 \geq 0$$

Example 2.3.3. *the delta function* $(x, y) \to\ \delta_{xy} = \begin{cases} 1\ if\ i = j \\ 0\ else \end{cases}$ *is of positive type*

Proof. Let $n \geq 1$ $(a_1, ..., a_n) \in \mathbb{C}^n$, $(x_1, ..., x_n) \in E^n$ and $\{a_1, ..., a_p\}$ the set of different values among $x_1, ..., x_n$. Then

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\bar{a}_j\delta_{x_i,x_j} &= \sum_{i=1}^{n}\sum_{x_i=x_j} a_i\bar{a}_j \\ &= \sum_{k=1}^{p}\sum_{x_i=x_j=a_k} a_i\bar{a}_j \\ &= \sum_{k=1}^{p}\left|\sum_{x_i=a_k} a_i\right|^2 \geq 0\end{aligned}$$

□

https://doi.org/10.26219/heal.aueb.6279

It also clear that if K and a is a non negative constant then the product aK is of a positive type function.

In contrast to the above examples, in most of our cases, it is difficult to prove through (2.4), that a given function is of positive type. Therefore we will present a very useful lemma for this case.

Lemma 2.3.4. *Let H be some Hilbert space with inner product$< .,. >_H$ and let $\phi : E \to H$. Then, the function K*

$$E \times E \to \mathbb{C}$$

$$(x, y) \to K(x, y) =< \phi(x), \phi(y) >_H$$

is of positive type.

Proof. The conclusion easily follows from the following equalities

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \bar{a}_j K(x_i, x_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} < a_i \phi(x_i), a_j \phi(x_j) >_H$$
$$= \left\| \sum_{j=1}^{n} a_i \phi(x_i) \right\|_H^2$$

□

Lemma 2.3.4 tells us that writing $K(x, y) =< \phi(x), \phi(y) >_H$, is sufficient to prove that K is of positive type.
A commonly used space in Probability theory is the space $\mathscr{L}^2(\Omega, A, P)$. This particular space provides us a useful consequence.

Example 2.3.5. *Consider the space $\mathscr{L}^2(\Omega, A, P)$ of square integrable functions in the probability space (Ω, A, P) as in Example 2.1.4. Then the covariance function K*

$$E \times E \to \mathbb{C}$$

$$E(X_t \bar{X}_s) =< X_t, X_s >_{\mathscr{L}^2(\Omega, A, P)}$$

of some complex valued zero mean stochastic process $(X_t)_{t \in E}$ is of positive type since

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

https://doi.org/10.26219/heal.aueb.6279

The above example tells us if we want to prove that the matrix $[K(t_i, t_j)]_{1\leq i,j\leq n}$ is of positive type, it is sufficient to prove that K is the covariance matrix of some zero mean random vector in $\mathscr{L}^2(\Omega, A, P)$.

Before we proceed to the main result of this chapter we will present some properties of positive type functions.

Lemma 2.3.6. *Let L be any positive type function on $E \times E$. Then,*

a) $\forall x \in E \quad L(x,x) \geq 0$

b) $\forall (x,y) \in E \times E \quad L(x,y) = \overline{L(y,x)}$

c) $\overline{L}$ *is a positive type function*

d)$|L(x,y)| \leq L(x,x)L(y,y)$

Proof. a)If we take $n = 1$ and $a_1 = 1$ then by Definition 2.3.1 we get $L(x,x) \geq 0$
b)Let $(x,y) \in \mathbb{E} \times \mathbb{E}$. From (2.4) and for any $(a,b) \in \mathbb{C}^2$ the number

$$C(a,b) = |a|^2 L(x,x) + a\bar{b}L(x,y) + b\bar{a}L(y,x) + |b|^2 L(y,y)$$

is a nonnegative real number. Moreover for $a = b = 1$ we get

$$L(x,y) + L(y,x) = C(1,1) - L(x,x) - L(y,y) = A$$

similarly, for $a = i$ and $b = 1$ we obtain

$$iL(x,y) - iL(y,x) = C(i,1) - L(x,x) - L(y,y) = B$$

Thus,

$$L(x,y) + L(y,x) = A \in \mathbb{R}$$

and

$$iL(x,y) - iL(y,x) = B \in \mathbb{R}$$

It follows that

$$A + iB = 2L(y,x)$$

and

$$A - iB = 2L(x,y)$$

https://doi.org/10.26219/heal.aueb.6279

hence $L(y,x)$ is the conjugate of $L(x,y)$ c) Taking the conjugate in equation 2.4 we obtain

$$\sum_{i=1}^{n}\sum_{i=1}^{n} a_i \bar{\bar{a}}_j L(x_i,x_j) \geq 0.$$

d) From b) we have, for any $a \in \mathbb{R}$,

$$0 \leq C(a, L(x,y)) = a^2 L(x,x) + 2a|L(x,y)|^2 L(x,x)L(y,y)$$

If $L(x,y) \neq 0$, the conclusion follows. Otherwise it is clear from a) $\square$

Lemma 2.3.6 clearly reveals us the existence of some similarities between positive type functions and reproducing kernels. In the next sections, we will see that this looming connection is fundamental in the context of reproducing kernels.

The following lemma proves the direct part of the equivalence between positive definite functions and Reproducing Kernels.

Lemma 2.3.7. *Any reproducing kernel is of positive type*

Proof. If K is a reproducing of H we have,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \bar{a}_j K(x_i,x_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} < a_i K(.,x_i), a_j K(.,x_j) >_H \\ &= \left\| \sum_{i=1}^{n} a_i K(.,x_i) \right\|_H^2 \geq 0\end{aligned}$$

$\square$

2.4 Construction of Reproducing Kernel

The next theorem is of great importance in constructing RKHS. Particularly, it will provide us the necessary and sufficient conditions required in order to construct a space of functions (RKHS) with the desirable properties(i.e.continuity of the evaluation functionals and convergence in the norm sense implies pointwise convergence). Furthermore will rely on Theorem 2.4.1 in order to prove the converse of Lemma 2.3.7 which is actually a way to fashion Reproducing Kernel Hilbert Spaces.

Theorem 2.4.1. *Let H_0 be any subspace of $\mathbb{C}^E$ the space of complex functions on E ,on which an inner product $< .,. >_{H0}$ is defined, with associated norm $\|.\|_{H_0}$. In order that there exists a Hilbert space H such that.*

 https://doi.org/10.26219/heal.aueb.6279

a)$H_0 \subset H \subset \mathbb{C}^E$ and the topology defined on H_0 by the inner product $< .,. >_{H0}$ coincides with the topology induced on H_0 by H.
b)H has a reproducing kernel K

it is necessary and sufficient that

c)the evaluation functionals $(e_t)_{t\in E}$ are continuous on H_0
d)any Cauchy sequence (f_n) in H_0 converging pointwise to 0 converges also in the norm sense.
(d) is equivalent to show that for any function f in H_0 converging pointwise to f, (f_n) converges to f in the norm sense)

Proof. **Direct part**
If H exists with conditions a), b) the evaluation functionals are by Theorem 2.2.5 continuous on H and therefore on H_0.
Let (f_n) be a Cauchy sequence in H_0 converges pointwise to 0. As H is complete, (f_n) converges in the norm sense to some $f \in H$.Thus we have

$$\forall x \in E \quad f(x) = e_x(f) = \lim_{n\to\infty} e_x(f_n) = \lim_{n\to\infty} f_n(x) = 0$$

and so $f \equiv 0$
converse
suppose c) and d) hold. Define $f \in \mathbb{C}^E$ for which there exists Cauchy sequence f_n in H_0 converging pointwise to f.It is easy to see that,

$$H_0 \subset H \subset \mathbb{C}^E$$

The proof of Theorem 2.4.1 will be completed by the Lemmas below. □

Lemma 2.4.2. *Let f , $g \in H$ and (f_n , g_n) be two Cauchy sequences inH_0 converging pointwise to f and g respectively. Then the sequence$< f_n, g_n >_{H_0}$ is convergent and its limit only depends on f,g.*

 https://doi.org/10.26219/heal.aueb.6279

Proof. f,g are Cauchy sequences in H_0, thus there exists $M, N > 0$ such that

$$\|f_n - f_m\|_{H_0} \leq 0 \;,\; \|g_n - g_m\|_{H_0} \leq 0 \tag{2.5}$$

$\forall (n,m) \in \mathbb{N}$,
$| < f_n, g_n >_{H_0} - < f_m, g_m >_{H_0} | = | < f_n - f_m, g_n > + < f_n, g_n - g_m > |_{H_0} \leq$
$\leq \|f_n - f_m\|_{H_0} \|g_n\|_{H_0} + \|f_m\|_{H_0} \| \|g_n - g_m\|_{H_0}$
by Cauchy-Schwartz inequality. This shows that $(< f_n, g_n >_{H_0})$ is a Cauchy sequence in $\mathbb{C}$ and therefore convergent. In the same way if f'_n, g'_n are two other Cauchy sequences in H_0 converging pointwise to f, g respectively we have,
$| < f_n, g_n >_{H_0} - < f'_n, g'_n >_{H_0} | \leq \|f_n - f'_n\|_{H_0} \|g_n\|_{H_0} + \|f'_n\|_{H_0} \|g_n - g'_n\|_{H_0}$.
Thus,$(f_n - f'_n)$ and $(g_n - g'_n)$ are Cauchy sequences in H_0 converging pointwise to 0. From assumption d) they also converge to 0 in the norm sense. It follows that $| < f_n, g_n >_{H_0} |$ and $< f'_n, g'_n >_{H_0}$ share the same limit. □

Lemma 2.4.3. *Suppose that* (f_n) *is a Cauchy sequence in* H_0 *converging pointwise to* f *and that* $\lim_{n\to\infty} < f_n, f_n >_{H_0} = 0$ *(*f_n *tends to 0 int the norm sense). Then* $f \equiv 0$

Proof.

$$\begin{aligned} \forall x \in E, \quad , f(x) &= \lim_{n\to\infty} f_n(x) \\ &= \lim_{n\to\infty} e_x(f_n) = 0 \; (by\ assumption\ c) \end{aligned}$$

Thus we can define an inner product on H by setting

$$< f, g >_H = \lim_{n\to\infty} < f_n, g_n >_{H_0}$$

where f_n , g_n are Cauchy sequences in H_0 converging pointwise to f , g respectively. It is clear that the $< f, g >_H$ satisfies the properties of a well defined inner product. The positivity, the hermitian symmetry and the linearity of $< .,. >_H$ arises from the properties of $< .,. >_{H_0}$. Lemma 2.4.3 tells us that if $< f, f >= 0$ then $f \equiv 0$. □

Lemma 2.4.4. *Let* $f \in H$ *and* (f_n) *be a Cauchy sequence in* H_0 *converging pointwise to* f*. Then* (f_n) *converges to* f *in the norm sense.*

Proof. Let $\epsilon > 0$ and let $N(\epsilon)$ be such that

$$(m\ ,\ n > N(\epsilon)) \implies \|f_n - f_m\|_{H_0} < \epsilon$$

Fix $n > N(\epsilon)$.The sequence $(f_n - f_m)_{m\in\mathbb{N}}$ is a Cauchy sequence in H_0 converging pointwise to $(f - f_n)$.Therefore

$$\|f - f_n\|_H \ = \lim_{n\to\infty} \|f_n - f_m\|_{H_0} \leq \epsilon$$

Thus f_n converges to f in the norm sense. □

Corollary 2.4.5. *H_0 is dense in H.*

Proof. By definition for any $f \in H$ there exists a Cauchy sequence (f_n) in H_0 converging pointwise to f. By Lemma 2.4.4 (f_n) converges to f in the norm sense. The Corollary follows. □

Lemma 2.4.6. *The evaluation functionals are continuous on H*

Proof. It is sufficient to show that the evaluation functionals are continuous at 0. (as long as linearity holds)
Let $x \in E$. By assumption c) in Theorem 2.4.1, the evaluation functional e_x is continuous on H_0. Fix $\epsilon > 0$ and let n such that

$$(f \in H_0 \ and \ \|f\|_{H_0} < n) \implies |f(x)| < \frac{\epsilon}{2}$$

For any function ϕ in H with $|\phi|_H < \frac{n}{2}$ there exists by Lemma 2.4.4 a function g in H_0 such that

$$|g(x) - \phi(x)| < \frac{\epsilon}{2} \qquad and \qquad |g - \phi|_H < \frac{n}{2}.$$

This entails

$$|g|_{H_0} \ = \ \|g\|_H \ \leq \|g - \phi\|_H + \|\phi\|_H < \ n$$

. Hence $|g(x)| < \frac{n}{2}$ and $|g(x)| < \epsilon$ .Thus e_x is continuous on H

□

Lemma 2.4.7. *H is a reproducing kernel Hilbert space*

 https://doi.org/10.26219/heal.aueb.6279

Proof. It has already been proved that all the evaluation functionals are continuous on H(Lemma 8). Therefore it suffices to show that H is complete. Let f_n be a Cauchy sequence in $\mathbb{C}$. Then by Lemma 2.4.6, the evaluation functional e_x is continuous, thus (f_n) is also a Cauchy sequence in $\mathbb{C}$ and converges to some f. One has to prove that this f belongs to H. Let $\epsilon(n)$ be any sequence of positive numbers tending to 0 as n tends to infinity.

As H_0 is dense in H

$$\forall i \in \mathbb{N}^\star \quad \text{there exists} \quad g_i \in \mathbb{N} \text{ such that } \quad \|f_i - g_i\|_H < \epsilon_i.$$

From the inequalities

$$\begin{aligned} |g_i(x) - f(x)| &\leq |g_i(x) - f_i(x)| \ + \ |f_i(x) - f(x)| \\ &\leq |e_x(g_i - f_i)| \ + \ |f_i(x) - f(x)| \end{aligned}$$

and from the properties of e_x (Lemma 2.4.6) it follows that(g_n) tends to 0 as n tends to ∞.

We have,

$$\begin{aligned} \|f_i - g_j\|_{H_0} = \|f_i - g_i\|_H &\leq \|f_i - g_i\|_H + \|f_i - f_j\|_H + \|f_j - g_j\|_H \\ &\leq \epsilon_i + \epsilon_j + \|f_i - f_j\|_H \end{aligned}$$

Thus (g_n) is a Cauchy sequence in H_0 tending pointwise to f, and so $f \in H$. By Lemma 2.4.4 g_n tends to f in the norm sense. Now,

$$\|f_i - f\|_H \ = \ \|f_i - g_i\|_H + \|g_i - f\|_H$$

Therefore (f_n) converges to f in the norm sense and H is complete. (We call H the functional completion of H_0). □

Now, with help of Theorem 2.4.1, we will prove the converse of Lemma 2.3.7 which plays fundamental role in RKHS theory.

Theorem 2.4.8. *MOORE-ARONSJAJN Let K be a positive type function on $E \times E$. There exists only one Hilbert space of functions on E with K as a reproducing kernel.The subspace H_0 of H spanned by the functions $(K(.,x)_x \in E$ is dense in H and*

https://doi.org/10.26219/heal.aueb.6279

H is the set of functions on E which are pointwise limits of Cauchy sequences in H_0 with the inner product

$$< f, g >_{H_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \bar{b}_j K(y_j, x_i)$$

where

$$f = \sum_{i=1}^{n} a_i K(., xi) \quad \textit{and} \quad g = \sum_{j=1}^{m} b_j K(., y_j) \tag{2.6}$$

Proof. First remark that the complex number $< f, g >$ defined by (2.6) does not depend on the representations not necessarily unique of f and g :

$$< f, g >_{H_0} = \sum_{i=1}^{n} a_i \overline{g(x_i)} = \sum_{j=1}^{m} \bar{b}_j f(y_j)$$

this shows that $< f, g >_{H_0}$ depends on f and g only through their values.

Then taking

$$f = \sum_{i=1}^{n} a_i K(., x_i) \quad \text{and} \quad g = K(., x)$$

we get

$$< f, K(., x) > = \sum_{i=1}^{n} a_i \overline{g(x_i)} = \sum_{i=1}^{n} a_i K(x, x_i) = f(x).$$

Thus the inner product with $K(., x)$ ”reproduces” the values of functions in H_0. In particular

$$\||K(., x)\|_{H_0}{}^2 = < K(., x), K(., x) > = K(x, x).$$

As K is a positive type function, $< ., . >_{H_0}$ is a semi-positive hermitian form on $H_0 \times H_0$. From the Cauchy-Schwartz inequality we have

$$\forall x \in E \quad |f(x)| = | < f, K(., x) >_{H_0} | \leq < f, f >_{H_0}^{1/2} [K(x, x)]^{1/2} = 0$$

and $f \equiv 0$

Let us consider H_0 endowed with the topology associated with the inner product $< ., . >_{H_0}$ and check conditions c),d) in theorem 2.4.1. Let f and g in H_0

$$\begin{aligned} \forall x \in E \quad |e_x(f) - e_x(g)| &= | < f - g, K(., x) >_{H_0} | \\ &\leq \|f - g\|_{H_0} [K(x, x)]^2 . \end{aligned}$$

https://doi.org/10.26219/heal.aueb.6279

Therefore the evaluation functionals are continuous on H_0 and c) is satisfied. Let (f_n) be a Cauchy sequence in H_0 converging pointwise to 0. Hence (f_n) is bounded, thus there exists $A > 0$ such that $\|f_n\| \leq A$. Now let $\epsilon > 0$ and $N(\epsilon)$ such that

$$n > N(\epsilon) \implies \|f_{N(\epsilon)-f_n}\|_{H_0} < \frac{\epsilon}{A}$$

Fix $a_1, ..., a_k$ and $x_1, ..., x_k$ such that

$$f_{N(\epsilon)} = \sum_{i=1}^{k} a_i K(., x_i).$$

As

$$\|f_n\|^2_{H_0} =< f_n - f_{N(\epsilon)}, f_n, >_{H_0} + < f_{N(\epsilon)}, f_n, >_{H_0}$$

we have for $n > N(\epsilon)$

$$\|f_n\|^2_{H_0} < \epsilon + \sum_{i=1}^{k} a_i f_n(x_i),$$

hence $\limsup_{n\to\infty} \|f_n\|^2 \leq \epsilon$ .Therefore condition d) satisfied. Combining the above results and applying Theorem 2.4.1, there exists a Hilbert space H of functions on E satisfying a) b) in theorem 2.4.1. H is the set of functions for which there exists a Cauchy sequence (f_n) in H_0 converging pointwise to f. By Lemma 2.4.4 f_n is also converging to f in the norm sense, thus H_0 is dense in H.Therefore H is unique and

$$\begin{aligned}\forall x \in E \quad f(x) &= \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} < f_n, K(., x) > \\ &=< f_n, K(., x) >_H\end{aligned}$$

Thus K is the reproducing kernel of H □

Essentially, MOORE-ARONSJAJN ensures us that if K is a positive type function(or equivalently a Reproducing Kernel), then there exists a unique Hilbert space H_K which possesses K as it's unique Reproducing Kernel. H_K can be constructed just by adding all finite combinations of the form $\sum a_i K(x_i,)$ and their limits under the norm induced by the inner product defined above. Lemma 2.2.6 ensures ue that norm convergence implies pointwise convergence and therefore these limits of functions are well defined in their pointwise sense.

Theorem 2.4.9 states an important conclusion. We will prove that the definition of positive type function (or equivalently of a reproducing kernel on $E \times E$) is equivalent of a mapping on E with values in some space $l^2(X)$ where $l^2(X)$ is the set of all complex valued sequences

$$\{x_a \ , \ a \in X\}$$

https://doi.org/10.26219/heal.aueb.6279

endowed with an inner product

$$< (x_a), (y_a) >= \sum_{a \in X} x_a \bar{y}_a$$

.

Theorem 2.4.9. *A complex valued function K defined on* $E \times E$ *is a reproducing kernel or a positive type function if and only if there exists a mapping T form E to some space* $l^2(X)$ *such that*

$$\begin{aligned} \forall (x,y) \in E \quad K(x,y) &= < T(x), T(y) >_{l^2(X)} \\ &= \sum_{a \in X} (T(x))_a \, (T(y))_a \end{aligned}$$

Proof. Let H be a RKHS of functions on a set E with kernel K. Consider the mapping

$$\begin{aligned} \Psi_K \quad : \quad & E \to H \\ & x \to K(., x) \end{aligned}$$

□

Like any Hilbert space, H is isometric to some $l^2(X)$. If ϕ denotes any isometry from H to $l^2(X)$, the mapping $T = \phi \circ \Psi_K$ meet requirements. Conversely, the mapping

$$T \quad : E \to l^2(X)$$

being given from a set E to some space l^2 the mapping

$$\begin{aligned} K : \quad & E \times E \to \mathbb{C} \\ & (x,y) \to < T(x), T(y) >_{l^2(X)} \end{aligned}$$

is by Lemma 1 a positive type function.

This characterization provides us an effective way of constructing Reproducing Kernels (or proving that a given function is a Reproducing Kernel). Particularly, since any pre-Hilbert space can be considered as a suitable isomorphism of $l^2(X)$, the only thing needed in order to construct a reproducing kernel is a mapping T as described in Theorem 2.4.9.

The following examples are a good sample of this useful theorem

https://doi.org/10.26219/heal.aueb.6279

Example 2.4.10. *Let* $E = [0, 1]$, $H = \mathcal{L}^2(-1, 1)$ *and* $T(x) = \cos(x.)$. *By theorem 4 we get that K defined on* $E \times E$ *by*

$$K(x, y) = < T(x), T(y) >_H = \int_{-1}^{1} \cos(xt) \cos(yt) d\lambda(t)$$
$$= \frac{\sin(x-y)}{x-y} + \frac{\sin(x+y)}{x+y} \quad if\ x \neq y$$

$$K(x, x) = 1 + \frac{\sin(2x)}{2x} \quad if\ x \neq 0$$

and

$$K(0, 0) = 2$$

is a reproducing kernel

2.5 Eigen-Structure of RKHS and Mercers theorem

As we described in the previous section, the basic tool needed for the construction of a RKHS, is a function with specific properties such as positive definiteness and symmetry. In this section we will present an important theorem which will turn to be helpful especially when problems that involve integral operators are considered. Since integral operators are strictly related with their eigenfunctions, it could be useful of finding a way to express a kernel of an operator in terms of their corresponding eigenfunctions and their eigen-values. Moreover, we will demonstrate a new direction of constructing RKHS through the eigenvalues of an operator. To be precise we will show that under the assumption that a function K symmetric and positive definite, we can define an integral operator L in order to use its corresponding eigenfunctions for the generation of a RKHS with reproducing kernel K. But before we proceed to that we thought that it could be useful to present the basic tools needed for this purpose.

Definition 2.5.1. *Let* $L : H \to H$ *be a compact linear operator then, by the term eigenvalue we mean all the functions* ϕ_i *that satisfy* $L\phi_i = \lambda_i \phi_i$. λ_i *is the i-th eigenvalue of* L.

Theorem 2.5.2. *SPECTRAL THEOREM Let* L *be a compact linear operator on an infinite dimensional Hilbert space* H. *Then there exists in* H *a complete orthonormal system* $\{\phi_1, \phi_2 ..., \}$ *consisting of the eigenvectors of* L. *If* λ_k *is the eigenvalue corresponding to* ϕ_k, *then the set* $\{\lambda_k\}$ *is either finite or* $\lambda_k \to 0$, *when* $k \to \infty$. *In addition,* $\max_{k \geq 1} |\lambda_k| = \|L\|$. *The eigenvalues are real if* L *is self adjoint. If in addition,* L

 https://doi.org/10.26219/heal.aueb.6279

is positive, then $\lambda_k \geq 0$ for all $k \geq 1$,and if L is strictly positive, then $\lambda_k > 0$ for all $k \geq 1$

if L is a strictly positive operator, then L^τ is defined for all $tau \geq 0$, by

$$L^\tau(\sum a_k\phi(k)) = \sum \lambda^\tau a_k \phi_k$$

.If $\tau < 0, L^\tau$ is defined by the same formula on the subspace

$$S_\tau \ = \{\sum a_k\phi_k \mid \sum (a_k\lambda_k^\tau)^2 \textit{ is convergent}\}$$

for $\tau < 0$ the expression $\|L^\tau a\|$ must be understood as ∞ if $a \notin S_\tau$

Let X be a compact domain or manifold in Euclidean space with dim$X$=n. Let λ be a Borel measure and $\mathscr{L}^2(X)$ be the Hilbert space of square integrable functions. Let $K \ : \ X \times X \to \mathbb{R}$. Then the mapping

$$L_K \ : \mathscr{L}^2(X) \to \mathscr{L}^2(X)$$

is continuous. K is said to be the kernel of L_K. Kernels of operators in the context of RKHS will be further discussed in chapter 2

Proposition 2.5.3. *If K is continuous then L_K is well defined and compact. Moreover $|Ł_K\| \leq \sqrt[2]{\lambda(X)}C_K$ where C_K is defined as $\sup_{x,t\in X}|K(x,t)|$*

Proof. Let $f \in \mathscr{L}^2(X)$ and $x_1, \ x_2 \ \in X$.Then

$$\begin{aligned}|(L_K(f))(x_1) - (L_K(f))(x_2)| &= |\int K(x_1,t) - K(x_2,t)f(t)| \\ &\leq \|K(.,x_1) - K(.,x_2)\|\|f\| \\ &\leq \sqrt[2]{\lambda(X)} \max_{t\in X}|K(x_1,t) - K(x_2,t)|\|f\|\end{aligned}$$

□

Where the first inequality comes from the Cauchy Schwartz inequality. Since K is continuous and X is compact, K is uniformly continuous. This implies continuity of $L_K f$.
Moreover one can similarly prove that $\|L_K\| \leq C_K$

Proposition 2.5.4. *a) If K is symmetric, then $L_K \ : \mathscr{L}^2(X) \to \mathscr{L}^2(X)$ is self adjoint (i.e. a linear operator $A : H \to H$ that satisfies $< Af, g >=< f, Ag >$ for all f,g in H.).*

b)If, in addition, K is positive definite, then L_K is positive definite.

https://doi.org/10.26219/heal.aueb.6279

Proof. part (a) follows easily from Fubini's theorem and the symmetry of K. For (b) just note that

$$< L_K f, g >_{\mathscr{L}^2} = \int\int K(x,t)f(x)f(t) = \lim_{k\to\infty} \frac{\lambda(X)}{k^2} \sum_{i,j=1}^{k} K(x_i,x_j)f(x_i)f(x_j)$$
$$= \lim_{k\to\infty} \frac{\lambda(X)}{k^2} f_X^T K[x] f_x$$

□

where for all $k \geq 1$, $x_1, ..., x_k \in X$ is a set of point conveniently chosen an $K[x]$ is the Gram matrix. The result follows So through a reproducing kernel one can define an integral operator which satisfies the conditions of the spectral theorem.

Corollary 2.5.5. *Let $\lambda_k, k \geq 1$, be the eigenvalues of L_K and ϕ_K the corresponding eigenfunctions. If $\lambda_k \neq 0$, then ϕ_k os continuous on X*

Proof. Let $s_n, s \in X$ such that,

$$s_n \to s \text{ as } n \to \infty.$$

If $\|L\| = \max_{k\geq 1 \lambda_k}$ with $\lambda_k \geq \lambda_{k+1}$ then, by writing $\phi_k = \frac{1}{\lambda_k} L_K \phi_k$,

$$|\phi_k(s_n) - \phi_k(t)| = \left| \frac{1}{\lambda_k}(L_K\phi_k)(s_n) - \frac{1}{\lambda_k}(L_K\phi_k)(s) \right| \leq$$
$$\leq \frac{1}{\|L\|} \int |K(s_n,t) - K(s,t)||\phi_k(s)|$$

which tends to 0 by continuity of K. This proves the continuity of ϕ □

We are now in position to state the main theorem of this section.

Theorem 2.5.6. MERCER *Let X be a compact domain or a manifold, λ a Borel measure on X, and $K : X \times X \to \mathbb{R}$ a Mercer Kernel. Let λ_k be the k-th eigenvalue of L_K and $\{\phi_k\}$ the corresponding eigenvectors. For all $x, t \in X$, $K(x,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x)\phi_k(t)$ where the convergence (for each x,y $X \times X$) and uniform (on $X \times X$).*

The readers are referred to H. Hochstadt, Integral equations, John Wiley and sons for the proof of the above theorem.

https://doi.org/10.26219/heal.aueb.6279

Corollary 2.5.7. *The sum* $\sum \lambda_k$ *is convergent and*

$$\sum_{k=1}^{\infty} \lambda_k \;=\; \int_X K(x,x) \le \lambda(X)C_K$$

therefore for all $k \le 1$, $\lambda_k \le \frac{\lambda(X)C_K}{k}$

Proof. By taking $x = t$ in Theorem 2.5.6 we get

$$K(x,x) = \sum_{k=1}^{\infty} \lambda_k \phi^2(x).$$

Integrating on both sides of this equality, we get

$$\sum_{k=1}^{\infty} \lambda_k \int_X \phi^2(x) \;=\; \int_X K(x,x) \le \lambda(X)C_K$$

But since $\{\phi_1, \phi_2 ....\}$ is a Hilbert basis, for all $k \ge 1$ we have $\int \phi_k = 1$ and the first statement follows. The second statement follows from the assumption $\lambda_k \ge \lambda_j$ for $j > k$ □

Theorem 2.5.8. *The map*

$$\begin{aligned} \Phi :& X \to l^2 \\ & x \to (\sqrt[2]{\lambda_k}\phi_k(x))_{k\in\mathbb{N}} \end{aligned}$$

is well defined, continuous, and satisfies

$$K(x,t) =< \Phi(x), \Phi(t) >$$

Proof. For every $x \in X$, by Mercer's theorem, $\sum \lambda_k \phi^2(k)$ converges to $K(x,x)$. This proves that $\Phi(x) \in l^2$.

Also be Mercers Theorem, for every $x, t \in X$,

$$K(x,t) \;=\; \sum_{k=1}^{\infty} \lambda_k \phi_k(x)\phi_k(t) =< \Phi(x), \Phi(t) >$$

It remains to prove that $\Phi \;:\; X \to l^2$. But for any $x, t \in X$,

$$\begin{aligned} \|\Phi(x) - \Phi(t)\| &=< \Phi(x), \Phi(x) > \;+\; < \Phi(t), \Phi(t) > \;- 2 < \Phi(x), \Phi(t) > \\ &= K(x,x) + K(t,t) - 2K(x,t) \end{aligned}$$

whch tends to 0 as x tends to t. The result follows. □

Let K be a reproducing kernel. As we mentioned in Proposition 2.5.3 and Proposition 2.5.4 one can define an operator L_K with the properties needed in order to apply the spectral theorem ($\lambda_k \geq 0$). Without loss of generality we will assume that $\lambda_k > 0$. Our goal now is to construct a RKHS H_K through the eigenvalues that L_K ensures.

Consider the space

$$H_K = \left\{ f \in \mathscr{L}^2(X) \mid f = \sum_{k=1}^{\infty} a_k \phi_k \text{ with } \frac{a_k}{\sqrt[2]{\lambda_k}} \in l^2 \right\} \tag{2.7}$$

Then H_K endowed with the inner product

$$< f, g >_K = \sum_{k=1}^{\infty} \frac{a_k b_k}{\lambda_k} \tag{2.8}$$

is a Hilbert space of functions for every $f = \sum a_k \phi_k$ and $g = \sum b_k \phi_k$.
An interesting result arises directly from the definition of H_K

Lemma 2.5.9. *The map*

$$\begin{aligned} L_K^{1/2} &: \mathscr{L}^2(X) \to H_K \\ \sum a_k \phi_k &\to \sum a_k \sqrt[2]{\lambda_k} \phi_k \end{aligned}$$

defines a Hilbert is a Hilbert isomorphism where $\mathscr{L}^2(X)$ is the unique operator such that $\mathscr{L}^2(X) \circ \mathscr{L}^2(X) = L_K$

Proof. It suffices to show that $< \mathscr{L}^{1/2} f, \mathscr{L}^{1/2} g, >_K = < f, g >_{\mathscr{L}^2}$. Let $f = \sum a_k \phi_k$ and $g = \sum b_k \phi_k$. Then

$$\mathscr{L}^{1/2} f = \sum a_k \sqrt[2]{\lambda_k} \phi_k \quad \text{and} \quad \mathscr{L}^{1/2} f = \sum b_k \sqrt[2]{\lambda_k} \phi_k$$

Also by (2.8) we have that

$$< \mathscr{L}^{1/2} f, \mathscr{L}^{1/2} g, >_K = \sum_{i=1}^{\infty} \frac{\lambda_i a_i b_i}{\lambda_i} = \sum_{i=1}^{\infty} a_i b_i \tag{2.9}$$

But

$$\begin{aligned} < f, g >_{\mathscr{L}^2} &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i b_j < \phi_i, \phi_j >_{\mathscr{L}^2} \\ &= \sum_{i=1}^{\infty} a_i b_i = < \mathscr{L}^{1/2} f, \mathscr{L}^{1/2} g, >_K \end{aligned} \tag{2.10}$$

The conclusion follows. ☐

https://doi.org/10.26219/heal.aueb.6279

It is worth noting that similarly, with the help of Theorem 2.5.2 one can prove that

(2.11) $$< f, g >_K = < \mathscr{L}^{-1/2} f, \mathscr{L}^{-1/2} g >_{\mathscr{L}^2}$$

Proposition 2.5.10. *The elements of H_K are continuous functions on X. In addition, for $f \in H_K$, if $f = \sum a_k \phi_k$, then the series converges absolutely and uniformly to f.*

Proof. Let $g \in H_K$, $g = \sum g_k \phi_k$ and $x \in X$. Then

$$g(x) = \left|\sum_{k=1}^{\infty} g_k \phi(k)\right| = \left|\sum_{k=1}^{\infty} \frac{g_k}{\sqrt[2]{\lambda_k}} \sqrt[2]{\lambda_k}\phi_k(x)\right| \leq \|g\|_K \|\Phi(x)\| = \|g\|_K K(x,x)^{1/2}$$

where the inequality follows form the Cauchy-Schwartz and the last inequality by Mercer's Theorem.

Now since the last equality holds true for all $x \in X$ we have $\|g\|_\infty \leq \sqrt[2]{C_K}\|g\|_K$.
Applying the previous inequality to the series $g_N = f - \sum a_k \phi_k$, proves the statement of the uniform convergence. The continuity of f follows that of the ϕ_k ($\phi_k = \frac{1}{\lambda_k}$). The absolute convergence comes follows form the inequality $\sum |g_k \phi_k| \leq \|g\|_K \|\Phi(x)\|$

□

Lemma 2.5.11. *Let $x \in X$. The function $\phi : X \to \mathbb{R}$, $\phi_x(t) =< \Phi(x), \Phi(t) >$ belongs to H_K*

Proof. Let $x \in X$. Then for any $t \in X$

$$\phi_x(t) =< \Phi(x), \Phi(t) > = \sum_{k=1}^{\infty} (\lambda_k \phi_k(x)) \phi_k(t) = \sum_{k=1}^{\infty} a_k \phi_k(t)$$

where $a_k = \lambda_k \phi_k(t)$.
But the quantity $\frac{a_k}{\sqrt[2]{\lambda_k}} = \sqrt[2]{\lambda_k}\phi_k(t)$ belongs to l^2 by Theorem 2.5.8. The result follows

□

Proposition 2.5.12. *For all $f \in H_K$ and all $x \in X$, $f(x) =< f, K(.,x) >$*

Proof. For $f \in H_K$, $f = \sum w_k \phi_k$,

$$\begin{aligned} < f, K(.,x) >_K &= \sum_{k=1}^{\infty} w_k < \phi_k, K(.,x) >_K = \sum_{k=1}^{\infty} \frac{w_k}{\lambda_k} < \phi_k, K(.,x) > \\ &= \sum_{k=1}^{\infty} \frac{w_k}{\lambda_k} \int \phi_k(t) K(x,t) = \sum_{k=1}^{\infty} \frac{w_k}{\lambda_k} (Ł_K \phi_k)(x) \\ &= \sum_{k=1}^{\infty} \frac{w_k}{\lambda_k} \lambda_k \phi_k(x) = f(x) \end{aligned}$$

https://doi.org/10.26219/heal.aueb.6279

□

Thus by H_K is an RKHS. So far, we have demonstrated two methods form which, a RKHS could be defined. An obvious question is whether the RKHS $\mathcal{H}_K$ generated from Theorem 2.4.8 and the RKHS generated by 2.7 are equivalent. The answer is true.

Theorem 2.5.13. *The Hilbert spaces $\mathcal{H}_K$ and H_K are the same space of functions on X with the same inner product.*

Proof. For any $x \in X$, the function $K(.,x)$ coincides, by Theorem 2.5.8 with the function ϕ_x in statement of Lemma 2.5.11.This proves that $\phi_x \in H_K$. Now by Proposition 2.5.12 we have that for all $t \in X$, $< f, K(.,t) >_K = 0$, $f \in H_K$. But $< f, K(.,t) >_K = 0$, thus $f = 0$ on X. The statement follows from Theorem 2.4.8. □

For more details about this subject see Poggio and Shelton (2002)

https://doi.org/10.26219/heal.aueb.6279

Chapter 3

Basic properties of Reproducing Kernels

So far, we have studied the basic properties of a notion named by Reproducing kernel. Furthermore we demonstrated several ways for the construction of a new space of functions named by Reproducing Kernel Hilbert Spaces. In this chapter we will try to showcase some useful properties of those spaces and also to examine the behavior of their elements under conditions that are commonly exist in several fields of applied mathematics and especially in Statistics (for example seperability, continuity).

3.1 Sum of Reproducing Kernels

Theorem 3.1.1. *Let K_1 and K_2 be the reproducing kernels of spaces H_1 and H_2 of functions on E with respective norms $\|.\|_{H_1}$ and $\|.\|_{H_2}$. Then the function $K = K_1 + K_2$ is the reproducing Kernel of the space $H = H_1 \oplus H_2 = \{f | f_1 + f_2, f_1 \in H_1 , f_2 \in H_2\}$ with the norm $\|.\|_H$ defined by*

$$\forall f \in H \quad \|f\|_H^2 = \min_{f | f_1 + f_2, f_1 \in H_1 \;, f_2 \in H_2} (\|f_1\|_{H_1}^2 + \|f_2\|_{H_2}^2)$$

Proof. Let $\mathcal{F} = H_1 \oplus H_2$ be the Hilbert sum of H_1 and H_2 $\mathcal{F}$ is the set $H_1 \times H_2$ endowed with the norm defined by

$$\|(f_1, f_2)\|_{\mathcal{F}}^2 = \|f_1\|_{H_1}^2 + \|f_2\|_{H_2}^2.$$

 https://doi.org/10.26219/heal.aueb.6279

Let

$$u \ : \mathcal{F} \to H$$
$$(f_1, f_2) \to f_1 + f_2$$

and let $N = \{(f_1, f_2) \ \in \mathcal{F} \ : f_1 + f_2 = 0\}$. Then the map u is linear ,onto and also a subspace of $\mathcal{F}$ .Let $(f_n, -f_n)$ be a sequence in N converging to(f_1, f_2). then f_1 , f_2 converges to $f_1 \in H_1$, $f_2 \in H_2$ respectively. Therefore $f_1 = -f_2$ and thus N is a closed subspace of $\mathcal{F}$.

Let $N^{\perp}$ be the orthogonal complement of N in $\mathcal{F}$ and let $v \ : u|_{N^{\perp}}$. The the map is 1-1 hence one can define an inner product on H by setting

$$< f, g >_H =< v^{-1}(f), v^{-1}(g) >_{\mathcal{F}} .$$

Endowed with this inner H is a Hilbert space of functions. Let us now check the reproducing property of K. Let $f \in H$ $(f', f'') = v^{-1}(f)$ and $(K'(., y), K''(., y)) = v^{-1}(K(., y)),\ y \in E$.As

$$K'(., y) - K_1(., y) + K''(., y) - K_2(., y) = 0$$

we have that $(K'(., y) - K_1(., y), K''(., y) - K_2(., y))$ belongs to N and its inner product in $\mathcal{F}$ with (f', f'') is 0. Thus

$$< f', K'(., y) >_{H_1} + < f'', K''(., y) >_{H_2} =< f', K_1(., y) >_{H_1} + < f'', K_2(., y) >_{H_2}$$

and

$$\begin{aligned} < f, K_{(.}, y) >_H &=< v^{-1}(f), v^{-1}(K(., y)) >_{\mathcal{F}} \\ &=< (f', f''), (K'(., y), K''(., y)) >_{\mathcal{F}} \\ &= f'(y) + f''(y) \end{aligned}$$

thus the reproducing property holds. It remains to express the $\|.\|_H$ as terms of $\|.\|_{H_1}$ and $\|.\|_{H_2}$.Now let $(f_1, f_2) \in \mathcal{F}^2$,$f = f_1 + f_2$ and $(g_1, g_2) = (f_1, f_2) - v^{-1}(f)$. On the other hand as (g_1, g_2) belongs to N and $v^{-1}(f)$ belongs to $N^{\perp}$. We have

$$\begin{aligned} \|(f_1, f_2)\|^2_{\mathcal{F}} &= \|v^{-1}(f)\|^2_{\mathcal{F}} + \|(g_1, g_2)\|^2_{\mathcal{F}} \\ &= \|v^{-1}(f)\|^2_{\mathcal{F}} + \|g_1\|^2_{\mathcal{F}} + \|g_2\|^2_{\mathcal{F}} \end{aligned}$$

 https://doi.org/10.26219/heal.aueb.6279

Therefore for $f = f_1 + f_2$ we always have

$$\|f\|_H^2 = \|v^{-1}(f)\|_{\mathcal{F}}^2 \leq \|f_1\|_{H_1}^2 + \|f_2\|_{H_2}^2 \tag{3.1}$$

and the equality holds if and only if $(f_1, f_2) = v^{-1}(f)$

□

Theorem 3.1.1 claims that if a direct sum ($H = H_1 \oplus H_2$) of two or more well defined RKHS exists, one can produce as many RKHS as desired just by adding their corresponding kernels of H_1 and H_2 or a linear combination of them (aK is also a reproducing Kernel).

Similarly, just by following the same technique for the structure of the mapping v whereby an inner product was defined for the Hilbert sum of H_1 and H_2, one can prove the following theorem.

3.2 Restriction of the index set

Theorem 3.2.1. *Let H be a Hilbert space of functions defined on E with reproducing kernel K and norm $\|.\|_H$ and let E_1 be a non empty subset of E. The restriction K_1 of K to $E_1 \times E_1$ is the reproducing kernel of the space H_1 of restrictions of elements of H to E_1 endowed with the norm $\|.\|_{H_1}$ defined by*

$$\forall f_1 \in H_1 \quad \|f_1\|_{H_1} = \min_{f \in H : f|_{E_1} = f_1} \|f\|_H$$

where $f|_{H_1}$ stands for the restriction of f to the subset E_1.

Proof. Let $u : H \to H_1$ such that $f \to f|_{E_1}$ and let $N = u^{-1}(\{0\})$. It is clear that u is linear and onto and N is subspace of H. If a sequence (f_n) in N tends to $f \in H$, we have

$$\forall x \in E_1, \quad \forall n \geq 1, \quad , f_n(x) = 0,$$

and, as the convergence in norm implies the pointwise convergence ,the limit f satisfies $f|_{E_1} \equiv 0$. Therefore N is closed. Let $N^\perp$ be its orthogonal complement in H and let $v := u|_{N^\perp}$ i.e. $v : \; N^\perp \to H_1$.The map is one-to-one and therefore we can define an inner product on H_1 by setting

$$< f, g >_{H_1} = < v^{-1}(f), v^{-1}(f) >_H$$

https://doi.org/10.26219/heal.aueb.6279

Endowed with this inner product H_1 is a Hilbert space of functions. It is clear that for any $y \in E_1$ we have that $K_1(.,y) \in H_1$ and that$[K(.,y) - v^{-1}(K_1(.,y))] \in N$. Thus for all $y \in E_1$ and for all $f \in H_1$ we have,

$$\begin{aligned} < f, K_1(.,y) >_{H_1} = < v^{-1}(f), v^{-1}(K_1(.,y)) >_H &= < v^{-1}(f), v^{-1}(K_{(.,y)}) >_H \\ &= v^{-1}(f)(y) = f(y) \end{aligned}$$

This shows that K_1 is the reproducing kernel of H_1.
Let $g \in H$ .If $g|_{E_1} = f_1$ then $(g - v^{-1}(f_1)) \in N$ and $v^{-1}(f_1) \in N^{\perp}$. By the Pythagorean identity ,

$$\|g\|_H^2 = \|g - v^{-1}(f_1)\|_H^2 + \|v^{-1}(f_1)\|_H^2$$

therefore

$$\|v^{-1}\|_{H_1} \leq \|g\|_H$$

and the equality holds if and only if $g = v^{-1}(f_1)$.The conclusion follows. □

3.3 Tensor product of RKHS

Products of functions and kernels play an important role especially when higher dimension problems are considered. Let H_1 and H_2 be two vector spaces of complex functions defined on E_1 and E_2 respectively. The tensor product $H_1 \tilde{\otimes} H_2$ is defined as the vector space generated by the function

$$\begin{aligned} f_1 \otimes f_2 \quad & E_1 \times E_2 \to \mathbb{C} \\ & (x_1, x_2) \to f_1(x_1) f_2(x_2) \end{aligned}$$

where $f_1 \in H_1$ and $f_2 \in H_2$

Let $< .,. >_1$ and $< .,. >_2$ be the inner products on H_1 and H_2 respectively. It can be proved that the mapping

$$\begin{aligned} & H_1 \tilde{\otimes} H_2 \to \mathbb{C} \\ & (f_1 \times f_2, f_1' \times f_2') \to < f_1, f_1' >_1 < f_2, f_2' >_2 \end{aligned}$$

is an inner product on $H_1 \tilde{\otimes} H_2$ which is therefore a pre-Hilbert space. Its completion is called the tensor product of H_1 and H_2 and is denoted by $H_1 \otimes H_2$.

Let H_1 and H_2 be two RKHS of functions in $\mathbb{C}$ with $K_1(x_1, y_1)$ and $K_2(x_2, y_2)$ be their

 https://doi.org/10.26219/heal.aueb.6279

corresponding reproducing kernels. Then $K_1(.,x_1), K_1(.,y_1)$ and $K_2(.,x_2), K_2(.,y_2)$ belong to H_1 and H_2 respectively. It is not hard to see that

$$(K_1(.,x_1) \times K_2(.,x_2), K_1(.,y_1) \times K_2(.,y_2)) \to K_1(x_1,y_1)K_2(x_2,y_2)$$

Our goal is to define a function $K \in H_1 \tilde{\otimes} H_2$ through which we can define the RKHS $H_K = H_1 \otimes H_2$. We will prove that the function $K_1 \times K_2 = K_1(x_1,y_1)K_2(x_2,y_2)$ meets the requirements (i.e.the reproducing kernel of the tensor product $H_1 \otimes H_2$). But before we proceed to that, we are will present a theorem will turn to be helpful in our proof.

Theorem 3.3.1. Horn and Johnson (2012)*Let A a Hermitian $n \times n$ matrix over the scalar $\mathbb{C}^{n \times n}$.*

a) K is positive definite if and only if there is a $n \times m$ matrix R such that $K = R'R$

b) If $K = R'R$ with R as in (a) and if $x \in \mathbb{C}^n$, then $Kx = 0$ if and only if $Rx = 0$, so $N(K) = N(R)$ and rank(A)=rank(B).

It is worth noting that the matrix R is not necessarily unique. As an example take the square root of K i.e. $R = K^{1/2}$. This particular case has many applications and we will study some of them in the next sections. We are now in position to prove the following Lemma.

Lemma 3.3.2. *H_1 and H_2 RKHS with kernels K_1 and K_2 respectively, the mapping*

$$K_1 \times K_2 \; (E_1 \times E_2)^2 \to \mathbb{C}$$

$$((x_1,x_2),(y_1,y_2)) \to K(x_1,y_1)K(x_2,y_2)$$

is of positive type function on $(E_1 \times E_2)^2$.

Proof. From Definition 2.3.1 it suffices to show that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i,x_j) \geq 0 \tag{3.2}$$

for all finite combinations of $c_1, ..., c_n$. Since K_1 is the reproducing kernel of H_1, it is Hermitian and positive definite. Thus by Theorem 3.3.1 there exists R such that $K_1 = R'R$ with

$$K_{1_{ij}} = \sum_{k=1}^{n} R_{ik}R_{kj} \tag{3.3}$$

https://doi.org/10.26219/heal.aueb.6279

Then, substituting (3.2) to (3.3) we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sum_{k=1}^{n} R_{ik} R_{kj} K_2(x_i, x_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} (c_i R_{ik})(R_{kj} c_j) K(x_i, x_j) \tag{3.4}$$

If we take $a = c'R$ then, it is clear that (3.4) takes the form

$$a'[K_2]a \geq 0 \tag{3.5}$$

since K_2 is the reproducing kernel of H_2 and therefore positive definite an Hermitian. The result follows. □

Recall that H_1 and H_2 are RKHS of function in $E \subset \mathbb{C}$. Then a natural way to define the algebraic tensor product $u = \sum_{i=1}^{n} h_i \otimes f_i$ is by setting $\hat{u}(x, s) = \sum_{i=1}^{n} h_i(x) f_i(x)$. The following theorem shows that this mapping is well defined and extends to the completed tensor product

Theorem 3.3.3. Paulsen (2009)*Let H_1 and H_2 be two RKHS with respective reproducing Kernels K_1 and K_2. Then $K((x, s), (y, t)) = K_1(x, y)K_2(s, t)$ is a reproducing kernel on $E_1 \times E_2$ and the map $u \to \hat{u}$ extends to a well defined linear isometry from $H_1 \times H_2$ onto the reproducing Kernel Hilbert Space $\mathcal{H}$*

Proof. Let $K_y^1(x) = K_1(x, y)$ and $K_t^2(s) = K_2(s, t)$. Note that if $u = \sum_{i=1}^{n} h_i \otimes f_i$, then

$$< u, K_y^1 \otimes K_t^2 >_{H_1 \otimes H_2} = \sum_{i=1}^{n} < h_i, K_y^1 >_{H_1} < f_i, K_t^2 >_{H_2} = \hat{u}(y, t) \tag{3.6}$$

Thus we may extend the mapping $u \to \hat{u}$ from an algebraic tensor product to the completed tensor product as follows. Let $u \in H_1 \otimes H_2$. We define a function $\hat{u}$ on $H_1 \times H_2$ as

$$\hat{u}(y, t) = < u, K_y^1 \otimes K_t^2 >_{H_1 \otimes H_2} \tag{3.7}$$

Let $\mathcal{H} = \{\hat{u} : u \in H_1 \otimes H_2\}$ be a vector space of functions on $H_1 \times H_2$. The map $u \to \hat{u}$ will be one to one unless there exists a non-zero $u \in H_1 \otimes H_2$ such that $\hat{u}(y, t) = 0$ for all (y, t) in $E_1 \times E_2$. But this condition would imply that u is

https://doi.org/10.26219/heal.aueb.6279

orthogonal to the span of $\{K^1_y \otimes K^2_t : (y,t) \in E_1 \times E_2\}$. But since by Theorem 2.4.8 the span of $\{K^1_y : y \in E_1\}$ is dense in H_1 and the span of $\{K^2_t : t \in E_2\}$ is dense in H_2, it follows that the span of $\{K^1_y \otimes K^2_t : \ (y,t) \in E_1 \times E_2\}$ is dense in $H_1 \otimes H_2$. Hence if $\hat{u} = 0$, then u is orthogonal to a dense subset and so $u = 0$

Thus, we have the map $u \to \hat{u}$ is one to one from $H_1 \otimes H_2$ onto $\mathcal{H}$ and we may use this identification to give $\mathcal{H}$ the structure of Hilbert space.by setting

$$< \hat{u}, \hat{v} >_{\mathcal{H}} = < u, v >_{H_1 \otimes H_2} \tag{3.8}$$

for $u, v \in H_1 \otimes H_2$. Finally, since for any $(y,t) \in E_1 \times E_2$ we have that $\hat{u}(y,t) =< u, \widehat{K^1_y \otimes K^2_t} >$ we see that $\mathcal{H}$ is a Reproducing Kernel Hilbert Space with kernel

$$\begin{aligned} K((x,s),(y,t)) &=< \widehat{K^1_y \otimes K^2_t}, \widehat{K^1_t \otimes K^2_s} >_{\mathcal{H}} =< K^1_y \otimes K^2_t, K^1_t \otimes K^2_s >_{H_1 \times H_2} \\ &=< K^1_y \otimes K^1_x, >_{H_1} < K^2_t \otimes K^2_s >_{H_2} = K_(x,y) K_2(s,t) \end{aligned} \tag{3.9}$$

and so K is a reproducing kernel.

By the uniqueness of RKHS as a Hilbert spaces, the map $u \to \hat{u}$ is an isometric linear map from $H_1 \otimes H_2$ onto $\mathcal{H}$ □

3.4 Support of a reproducing Kernel

We will now discuss a notion, first introduced by Duc-Jacquet(1973) which is of great importance in searching for bases in RKHS. It is the notion of support of a function. But before we present that, it is necessary to introduce some basic facts and definitions that will be useful in its definition.

Definition 3.4.1. *Let K be a non null complex function defined on $E \times E$. A subset A of E is said to be binding for K if and only if there exist elements $x_1, ..., x_n$ in A such that the functions $K(., x_1), ..., K(., x_n)$ are linearly dependent in the vector space $\mathbb{C}^E$.*

Theorem 3.4.2. *The set $\mathcal{F}_K$ of non-biding sets for K partial ordered by inclusion is inductive and therefore admits at least one maximal element.*

Proof. Let $\{A_i \ : \ i \in I\}$ be a set of elements of $\mathcal{F}_\mathcal{K}$ linearly ordered by inclusion. Then the set $\bigcup_{i \in I} A_i$ belongs to $\mathcal{F}_\mathcal{K}$ and is the upper bound of the chain $(A_i)_{i \in I}$. Thus

https://doi.org/10.26219/heal.aueb.6279

$\mathcal{F}_{\mathcal{K}}$ is partially ordered by inclusion is inductive. By Zorn's Lemma it has at least one maximal element. □

Definition 3.4.3. *Let K be a non null complex function defined on $E \times E$ a subset S of E is called a support of K if and only if S is a maximal element of the set $\mathcal{F}_{\mathcal{K}}$ of non-binding sets for K.*

The link between support of reproducing kernel and basis of H_0 is expressed in the following theorem.

Theorem 3.4.4. *Let H be a RKHS with kernel K on $E \times E$. Let H_0 be the subspace of H spanned by $\{K(.,x) : x \in E\}$. If a subset S of E is a support of K then $\{K(.,x) : x \in S\}$ is a basis of H_0.Conversely if $K(.,x_1),...,K(.,x_n)$ are linearly independent, there exists a support S of K containing $\{x_1,...,x_n\}$.*

Proof. The set $S = \{k(.,x : x \in E)\}$ is a set of linearly independent elements of H_0. If x_0 belongs to $X \setminus S$,consider the set $S \bigcup \{K(.,x_0)\}$ and use the maximality of S to get that $K(.,x_0)$ can be written as a linear combination of elements in S. Therefore the set S spans H_0. Now if $K(.,x_1),...,K(.,x_n))$ are linearly independent , the set $\{x_1,....,x_n\}$ is a non binding set for K, hence it is included in a support of K. □

3.5 Kernel of an operator

Definition 3.5.1. *Let $\mathcal{E}$ be a pre-Hilbert space of functions defined on E and let u be an operator in $\mathcal{E}$. A function $U : E \times E \to \mathbb{C}$ $(x,y) \to U(x,y)$ is said to be kernel of u if and only if*

$$\forall y \in E,\ U(.,y) \in \mathcal{E}$$

$$\forall y \in E\ \forall f \in \mathcal{E},\ \ u(f)(y) =< f, U(.,y) >_{\mathcal{E}}$$

Let us denote that Definition 3.5.1 is equivalent to that any Hilbert space functions H has a reproducing kernel K if and only if K is the kernel of the identity operator in H.
Also is clear that if u has two kernels U_1, U_2 one has

$$\forall y \in E\ ,\ \forall f \in \mathcal{E}\ ,\ < f, U_1(.,y) - U_2(.,y) >_{\mathcal{E}} = u(f)(y) - u(f)(y) = 0$$

So for any operator there is at most one kernel.

Theorem 3.5.2. *In a Hilbert space H of functions with reproducing kernel K any continuous operator u has a kernel U given by*

$$U(x,y) = [u^\star(K(x,y))](x) \tag{3.10}$$

where $u^\star$ denotes the adjoint operator of u.

Proof. By Riesz's theorem, in the Hilbert space H any continuous operator u has an adjoint defined by

$$\forall (x,y) \in H \times H \ \ < u(f), g >_H = < f, u^\star(g) >_H .$$

Thus we have

$$\begin{aligned} \forall y \in E \ , \ \forall f \in H, \ < f, u^\star(K(.,y)) >_H &= < u(f), K(.,y) >_H \\ &= u(f)(y). \end{aligned}$$

□

Example 3.5.3. ***THE COVARIANCE OPERATOR***
Will now consider a tool which is of great importance especially in the study of stochastic processes.
Let X be a random variable defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with values in $(H, \mathcal{B})$ where H is a RKHS of functions on a set E and $\mathcal{B}$ is the Borel σ-algebra. For any $\omega \in \Omega$, $X(\omega) = X.(\omega)$ is the function defined on E by

$$E \to \mathbb{C}$$
$$t \to X_t(\omega)$$

is called trajectory associated with ω. In other words $(X_t)_{t\in E}$ is a stochastic process on $(\Omega, \mathcal{A}, \mathcal{P})$ with trajectories in H.
Suppose that X is a second order is a second order random variable, i.e.

$$E_P(\|X\|_H)^2 < \infty.$$

https://doi.org/10.26219/heal.aueb.6279

Then the covariance operator of X is defined by

$$C_X(f) = E_P(< X, f >_H X). \tag{3.11}$$

Where the expectation is taken from the Bohner integral. It can be defined equivalently as the Unique operator C_X satisfying

$$< C_X(f), g >_H = E(< X, f >_H < X, g >_H). \tag{3.12}$$

The operator C_X is continuous self adjoint and compact. From theorem 9 its Kernel given by

$$\begin{aligned} U(t,s) &= [C_X(K(.,s))](t) \\ &= < C_X(K(.,s)), K(.,t) >_H \\ &= E(< X, K(.,s) >_H < X, K(.,t) >_H) \\ &= E(X_\bullet X_s) \end{aligned} \tag{3.13}$$

From the results proved above, it comes out that the covariance operator of a second order random variable X with values on an RKHS H of function defined on a set E has a kernel which is the second moment function of X.

3.6 Duality Between RKHS and Stochastic Processes

Let $X(t), t \in T$ be a family of random variables defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$. As we saw in the previous section, one can define a function C

$$\begin{aligned} &T \times T \to \mathbb{R} \\ &C(s,t) = EX(s)X(t), \; s,t \in T \end{aligned}$$

which is a reproducing kernel and therefore by Theorem 2.4.8, there is a unique Reproducing Kernel Hilbert Space H_C with C as its reproducing Kernel. In this section we will describe the duality between a Hilbert space spanned by a family of random variables and its associated reproducing Kernel.

Let $X(t)\, t \in T$ be a family of zero mean Gaussian variables with covariance function $C(s,t) = E(X(s)X(t))$. Let H be the Hilbert space

$$H = \overline{span\{Z \in (\Omega, \mathcal{A}, \mathcal{P}) : \; Z = \sum a_j X(t_j)\}} \tag{3.14}$$

with inner product $< Z_1, Z_2 >= EZ_1 Z_2$. Observe that

$$\|Z - Z_l\|^2 = < Z - Z_l, Z - Z_l >= E(Z - Z_l)^2 \tag{3.15}$$

https://doi.org/10.26219/heal.aueb.6279

Thus, a random variable Z belongs to H iff there exist a sequence $Z_l \in H$ such that

$$E(Z - Z_l)^2 \to 0$$

Let H_C be the RKHS associated with C. Then H_C is isometrically isomorphic to H since

$$< X(s), X(t) >_H = EX(s)X(t) = C(s,t) =< C(.,s), C(.,t) >_{H_C} . \tag{3.16}$$

So this duality between stochastic processes and RKHS, provides us lots of useful applications such as Bayesian estimation (Wahba (1990)) or in regularization methods which will be discussed in later sections.

3.7 Kernel of a closed subspace

Theorem 3.7.1. *Let V be a closed subspace of a Hilbert space H with reproducing kernel K. Then V is a reproducing kernel Hilbert space and its kernel K_V is given by*

$$K_V(x,y) \;=\; [\Pi_V(K(.,y))](x)$$

where Π_V denotes the orthogonal projection onto the space V

Proof. As Π_V is a self-adjoint operator $(\Pi_V = \Pi_V^\star)$,by Theorem 9 K_V is the reproducing kernel of Π_V. Now the restriction of Π_V into the subspace V is the identity operator of V. Therefore K_V is the reproducing kernel of V. □

By Riesz's representation theorem, for any continuous linear functional u

$$H \to \mathbb{C}$$
$$f \to u(f)$$

there exists unique $\tilde{u} \in H$ such that u can be represented as

$$u(f) =< f, \tilde{u} > \text{for all} f \in H \tag{3.17}$$

Particularizing to the case of a RKHS $\tilde{u}$ can be easily be expressed through its kernel K

Lemma 3.7.2. *In a Hilbert space H of functions with reproducing kernel K any continuous linear form $u \;:\; H \to C$ has a Riesz representer $\tilde{u}$ given by*

$$\tilde{u}(x) = u(K(.,x)) \tag{3.18}$$

https://doi.org/10.26219/heal.aueb.6279

Proof. If we put f(.)=K(.,x) in (3.17) we obtain

$$u(K(.,x)) =< K(.,x), \tilde{u} >= \tilde{u}(x)$$

by the reproducing property of H. Thus, $\tilde{u}(x) = u(K(.,x))$ is Riesz the representer of u. □

3.8 Condition $H_K \subset H_R$

Denote by H_k the RKHS corresponding to a given reproducing kernel K ,as given by the Moore-Aronszajn theorem. When the index set is a seperable metric space Aronszajn(1950) proves the following.

Theorem 3.8.1. *Let K be a continuous non negative kernel on $T \times T$ and R be a continuous positive kernel on $T \times T$. The following statements are equivalent*

i)$H_K \subset H_R$

ii) There exists a constant such that $B^2R - K$ is a non negative kernel.

Ylvisaker(1962) gives an alternative condition which is that if $\sum_{j=1}^{N(n)} c_{j_n} R(.,t_{j_n})$ is a Cauchy sequence in h_R then $\sum_{j=1}^{N(n)} c_{j_n} K(.,t_{j_n})$ must be a Cauchy sequence in H_K.
Discroll(1973) proves that any of these conditions is equivalent to :There exists an operator $T : H_R \to H_K$ such that $\|L\| \leq B$ and $LR(t,.) = K(t,.)$ for all $T \in T_0$ where T_0 is a countable dense subset of T. Moreover i) implies that there exists a constant B such that

$$\forall g \in H_K, \quad \|g\|_R \leq \|g\|_K,$$

and either of these conditions implies that there exists a self adjoint operator $L : H_R \to H_K$ such that $\|L\| \leq B$ and

$$\forall t \in T, \quad LR(T,.) = K(t,.)$$

3.9 Separability Continuity

Seperable Hilbert spaces provides us many useful properties especially when we are considering the representation of an element in a such a space H. We will see that most of these properties can be particularized on an RKHS.

We know that in a separable metric space $\mathcal{E}$ any dense subset contains a countable subset which is dense in $\mathcal{E}$. Lemma 3.9.1 implies that something analogous holds also for an RKHS.

https://doi.org/10.26219/heal.aueb.6279

Lemma 3.9.1. *In a seperable RKHS H there is a countable set D_0 of finite linear combinations of functions $K(.,x)$, $x \in E$, which is dense in H.*

Proof. By theorem 3 the subspace H_0 of H spanned by the functions K(.,x), $x \in E$,is dense in H. Let $\{x_p \ , \ x \in E\}$ be a countable subset dense in H and n be a positive integer. As $\bar{H}_0 = H$, for any p in $\mathbb{N}$ there exists y_p^n in H_0 such that

$$\|y_p^n - x_p\| < \frac{1}{n}$$

. The the countable set $\bigcup_{n>0}\{y_n^p \ , \ p \in \mathbb{N}\}$ satisfies the requirement. To see this, consider y in H, $\epsilon > 0$ and $n > \frac{2}{\epsilon}$. There exists $p \in \mathbb{N}$ such that

$$\|y - x_p\| < \frac{\epsilon}{2}$$

. Therefore

$$\|x_p - y_p^n\| < \frac{1}{n} < \frac{\epsilon}{2} \text{ and } \|y - y_p^n\| < \epsilon.$$

We can conclude that D_0 is dense in H. □

As we mentioned before, any element $f \in H$ can be represented as an infinite sum via an orthonormal system. Theorem 3.9.2 presented below tells us that a reproducing kernel can be expressed in such manner.

Theorem 3.9.2. *Let $H \subset \mathbb{C}^E$ be a seperable Hilbert space with reproducing kernel K. For any complete orthonormal system $(e_i)_{i\in\mathbb{N}}$ in H we have*

$$\forall t \in E, \quad K(.,t) = \sum_{i=0}^{\infty} \bar{e}_i(t) e_i(.) \quad \textit{(convergence in H)} \tag{3.19}$$

Conversely if (3.19) holds for an orthonormal system $(e_i)_{i\in\mathbb{N}}$ then this system is complete and H is seperable. Moreover, (3.19) implies that

$$\forall s \in E, \quad \forall t \in E, \quad K(.,t) = \sum_{i=0}^{\infty} \bar{e}_i(t) e_i(.) \quad \textit{(convergence in } \mathbb{C}).$$

Proof. Fix t in E. The function $K(.,t)$, as an element of H has a Fourier series expansion

$$\sum_{i=0}^{\infty} c_i e_i(.)$$

https://doi.org/10.26219/heal.aueb.6279

where

$$c_i =< K(.,t), e_i >= \overline{< e_i, K(.,t) >} = \bar{e}_i(t)$$

Conversely if (3.19) holds for an orthonormal system $(e_i)_{i\in\mathbb{N}}$ then

$$\forall \phi \in H, \quad \forall t \in E, \quad \phi(t) =< \phi, K(..,t) >= \sum_{i=0}^{\infty} e_i(t) < \phi, e_i >$$

thus

$$\forall \phi \in H, \quad \phi = \sum_{i=0}^{\infty} < \phi, e_i > e_i(t) (\text{convergence in H})$$

Therefore the system $(e_i)_{i\in\mathbb{N}}$ is complete and H is seperable. The last property follows from (3.19) by computing $K(s,t) =< K(.,s), K(.,t) >$. □

We will now present a useful Theorem which can be used as criterion for the seperability of a Hilbert space with reproducing kernel.

Theorem 3.9.3. *Let H be a Hilbert space of functions on E with reproducing kernel K.Suppose that E contains a countable subset E_0 such that*

$$\forall g \in H, \quad (g|_{E_0} = 0 \Leftrightarrow g = 0)$$

Then H is seperable.

Proof. Consider an element g in H orthogonal to the family $(K(.,y))_{y\in E_0}$. For any $y \in E_0$, one has

$$g(y) =< g, K(.,y) >_H = 0$$

thus $g|_{E_0} = 0$ and $g = 0$ by the hypothesis.It follows that the subspace V generated by the family $(K(.,y))_{y\in E_0}$ is equal to 0. Hence V is equal to H. The countable family $(K(.,y))_{y\in E_0}$ is total in H and therefore H is separable. □

Corollary 3.9.4 provides us a useful example in order to understand the utility of the seperability in a Hilbert space.

Corollary 3.9.4. *A non separable Hilbert space of continuous functions on a separable topological space E has no reproducing kernel.*

Theorem 3.9.5. *(Fortet(1973)) A RKHS with kernel K is separable if and only if for any $\epsilon > 0$,there exists a countable partition B_j, $j \in \mathbb{N}$ of E such that*

(3.20) $$\forall j, \forall t_1, t_2 \in B_j, K(t_1,t_1) + K(t_2,t_2) - K(t_1,t_2) - K(t_2,t_1) < \epsilon$$

https://doi.org/10.26219/heal.aueb.6279

Let us now turn to a characterization of RKHS of continuous functions

Theorem 3.9.6. *Let H be a Hilbert space of functions defined on a metric space (E,d) with reproducing kernel K. Then any element of H is continuous if and only if K satisfies the following conditions.*
a)$\forall y \in E$,$K(.,y)$ is continuous
b)$\forall x \in E$ there exists $r > 0$ such that the function

$$E \to \mathbb{R}^+$$
$$y \to K(y,y)$$

is bounded on the open ball $B(x,r)$.

Proof. If any element of H is continuous a) is clearly satisfied. Suppose that b) does not hold true. Then there exists $x \in E$ such that

$$\forall n \in \mathbb{N} \text{ , } \exists x_n \in B(x, 1/n) \text{ , such that} K(x_n, x_n) \geq n$$

As the sequence x_n converges to x we have for any continuous function ϕ in H

$$< \phi, K(., x) >= \phi(x) = \lim_{n\to\infty} \phi(x_n) = \lim_{n\to\infty} < \phi, K(., x_n) >.$$

Therefore the sequence $(K(., x_n))$ converges weakly to $K(., x)$ whereas

$$\|K(., x_n)\|^2 = K(x_n, x_n)$$

tends to infinity. This is a contradiction since any weakly convergent sequence in a Hilbert space is bounded .Hence b) is satisfied.
Conversely suppose that a),b) hold true. Let (x_n) be a convergent sequence in E with limit x, let ϕ be an element of H and let (r, M) such that

$$\sup_{y\in B(x,r)} K(y, y) \leq M.$$

For n large enough x_n belongs to $B(x, r)$ hence we have

$$\|K(., x_n)\|^2 \leq M$$

https://doi.org/10.26219/heal.aueb.6279

Let H_0 be the dense subspace of H spanned by the functions $(K(.,y))_{y\in E}$. Any element of H_0 con be written as a finite linear combination

$$\sum_{i=1}^{k} a_i K(.,y_i)$$

so it is by a), a continuous function. Let(ϕ_m) be a sequence in H_0 converging to ϕ in the norm sense. By Lemma 2.2.6 (ϕ_m) also converges pointwise to ϕ. Let $\epsilon > 0$. Fix m large enough to have

$$|\phi_m(x_n) - \phi_m(x)| < \epsilon$$

and

$$\|\phi_m - \phi\|_H < \epsilon$$

as ϕ_m is continuous, for n large enough we have

$$|\phi_m(x_n) - \phi_m(x)| < \epsilon.$$

Therefore for n large enough,

$$\begin{aligned}
|\phi(x_n) - \phi(x)| &\leq |\phi(x_n) - \phi_m(x_n)| + |\phi_m(x_n) - \phi_m(x)| \\
&\quad + |\phi_m(x_n) - \phi(x)| \\
&\leq | < K(.,x_n), \phi - \phi_m >_H | + 2\epsilon \\
&\leq (K(x_n,x_n))^{1/2} \|\phi - \phi_m\|_H + 2\epsilon \\
&\leq (M+2)\epsilon
\end{aligned}$$

This shows that ϕ is continuous at x. □

Corollary 3.9.7. *Let H be a Hilbert space of functions defined on a metric space (E,d) with Reproducing Kernel K. If K is bounded and if, for any $y \in E$, $K(.,y)$ is continuous (this implies, by symmetry, that for any $x \in E$, $K(.,x)$ is continuous) then H is a space of continuous functions. If moreover, E is seperable, then H is seperable and*

$$\forall s \in E, \quad \forall t \in E \quad K(s,t) = \sum_{i=0}^{\infty} \bar{e}_i(t) e_i(s), \tag{3.21}$$

where e_i is any orthonormal system in H.

https://doi.org/10.26219/heal.aueb.6279

Corollary 3.9.8. *Let H be a Hilbert space of functions defined on a compact metric space (E,d) with reproducing kernel K. If K is continuous then H is a seperable space of continuous functions and*

$$\forall s \in E, \quad \forall t \in E, \quad K(s,t) = \sum_{i=0}^{\infty} \bar{e}_i(t) e_i(s) \tag{3.22}$$

where the convergence is uniform on $E \times E$ and (e_i) is an orthonormal system in H the functions e_i are uniformly continuous and bounded by $(\sup_t K(t,t))^{1/2}$.

Proof. E is compact,hence is separable and K is continuous and therefore bounded. Thus Corollary 3.9.7 applies. H is a seperable space of continuous functions and

$$\forall s \in E, \quad \forall t \in E, \quad K(s,t) = \sum_{i=0}^{\infty} \bar{e}_i(t) e_i(s)$$

where the functions (e_i) are continuous on H (therefore uniformly continuous) and orthonormal in in H. For any $t \in E$,

$$|e_i(t)| \ = \ | < e_i, K(.,t) >_H | \leq \|e_i\| \|K(.,t)\| = [K(t,t)]^{1/2}$$

It remains to show that the convergence in (3.22) is also uniform. The sequence

$$\{\sum_{i=0}^{n} |e_i|^2(t) \ : \ n \in \mathbb{N}\}$$

is an increasing sequence of continuous functions of the variable t converging pointwise to a continuous function

$$K(t,t) = \sum_{i=0}^{n} |e_i|^2(t)$$

on the compact set E. By Dini's theorem the convergence is uniform. As we have

$$\left| \sum_{i=n}^{\infty} \bar{e}_i(t) e_i(s) \right|^2 \leq \left| \sum_{i=n}^{\infty} |e_i| \right|^2 (t) \left| \sum_{i=n}^{\infty} |e_i| \right|^2 (s)$$

so the convergence of

$$\sum_{i=0}^{\infty} \bar{e}_i(t) e_i(s) \text{ to } K(s,t)$$

is uniform on $E \times E$ □

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

Chapter 4

Representer Theorem for Reproducing Kernel Hilbert spaces

4.1 Introduction

Let X and Y be real Hilbert Spaces. In this chapter we are going to deal with problems that can be written as

$$Ax = y \tag{4.1}$$

where A is a linear operator with domain $D(A) \subset X$ and Range $R(A) \subset Y$.
Before we proceed to the quest the solution of aforementioned problems with the form of (4.1), we need to answer a few theoretical questions. The most obvious question that is raised directly from the formulation of this problem is whether there exists a solution that lies in X for a given $y \in Y$. And if there exists such a solution, is it unique? More importantly, this solution depends continuously on the data y? Equations that all of the above questions can be answered positively are said to be well-posed. Otherwise they are called ill-posed equations. In this chapter we will discuss about finding a solution from ill-posed problems

Practically, in the most cases such equations can be solved approximately using methods that guaranties the existence of an element with "similar" properties. For example instead of solving a problem with the form of (4.1) we can try to find an element $u \in D(A)$ such that for given $y \in Y$ $\inf\{\|Ax - y\| \ : \ x \in D(A)\} = \|Au - y\|$. Such a solution is called Least squares solution of the equation $Ax = y$. This topic will further be examined in another section of this chapter. For further details about well posed and ill posed equations see Nair (2009)

Our main target of this chapter is to show that when X and Y are $\mathcal{L}^2$ functions, that is the topology of Reproducing Kernel Hilbert spaces is an appropriate topology for the

https://doi.org/10.26219/heal.aueb.6279

regularization of ill-posed linear operator equations and also providing an approach to optimal approximations of linear operator equations in the context of RKHS. Furthermore we will demonstrate the relation between the regularization operator of the equation $Af = g$ and the generalized inverse of A in an appropriate RKHS.

Let us now examine the main parts that will be useful in presenting the main part of this chapter, The representer theorem.

4.2 Preliminaries

Let X and Y be linear spaces over the scalar field $\mathbb{C}$ and let $T : X \to Y$ be linear operator. We define its null space $N(T)$ as

$$N(T) := \{x \in X : Tx = 0\}.$$

The range $R(T)$ of T as

$$R(T) := \{Tx : x \in X\}.$$

The domain $D(T)$ of T as

$$D(T) := \{x \in X \; \exists y \in Y : Tx = y\}.$$

.

The set of all bounded linear operators T on some set H will be denoted by $B(H)$

A linear operator T is injective if and only if for every x_1, x_2 with $x_1 \neq x_2 \Rightarrow T(x_1) \neq T(x_2)$

A linear operator T is called bijective if $D(T) = X$ and $R(T) = Y$

Definition 4.2.1. *Let H be a Hilbert space of functions numbers with inner product $< .,. >_H$ and let A be a subspace of H with $A \neq \emptyset$. We define the orthogonal complement $A^{\perp}$ of A as*

$$A^{\perp} := \{x \in H : \forall a \in A, < x, a >_H = 0\}$$

Definition 4.2.2. *Let $B(H, K)$ be the set of all bounded operators where H and K are complex Hilbert spaces and let $T \in B(H, K)$.There exists a unique operator T^* such that*

$$< Tx, y > = < x, T^* y >$$

for all $x \in H$ and all $y \in K$.We call T^* the adjoint operator of T. An operator $T \in B(H)$ is called self adjoint if $T = T^*$

https://doi.org/10.26219/heal.aueb.6279

Theorem 4.2.3. *Let X be a normed linear space, Y be a Banach space, Λ be a subset of $\mathbb{R}$ having a limit point a_0 and $\{T_a\}_{a\in\Lambda}$ be a uniformly bounded family of operators in $B(X,Y)$. Suppose D is a dense subset of X such that $(T_a x)$ converges as $a \to a_0$ for every $x \in D$. Then $(T_a x)$ converges as $a \to a_0$ and the operator $T : X \to Y$ defined by*

$$Tx := \lim_{a\to a_0} T_a x, \; x \in X$$

belongs to $B(X,Y)$

Theorem 4.2.4. *Let H be a Hilbert space and M be a closed subspace of H(or a finite dimensional space). Then $H = M \oplus M^{\perp}$.More precisely every $x \in H$ can uniquely be written as $x = x_1 + x_2$ where $x_1 \in M$ and $x_2 \in M^{\perp}$*

4.3 Well-Posed and Ill-posed Operator Equations

Let X and Y be linear spaces over the scalar field and let $T : X \to Y$ be linear operator.

We say that the equation (4.1) is well posed if

a)for every $y \in Y$, there exists a unique $x \in X$ such that $Tx = y$
b)for every $y \in Y$ and for every $\epsilon > 0$, there exists $\delta > 0$ with the following properties: If $\tilde{y} \in Y$ with $\|\tilde{y} - y\| \le \delta$ and if $x, \tilde{x} \in X$ are such that $Tx = y$ and $T\tilde{X} = \tilde{y}$ then $\|\tilde{x} - x\|_X \le \epsilon$

It is worth noting that condition a) in the above definition represents the existence and the uniqueness of the solution (4.1) and condition b) is actually the assertion of the continuous dependence of the solution on the data y.
Another definition about the well-posedeness of an equation is that equation (4.1) is well posed if and only if the operator T is bijective and the inverse operator $T^{-1} : Y \to X$ is continuous.
Particularizing this to Banach spaces and hence to Hilbert spaces one can prove that if T is a continuous linear operator then continuity of T^{-1} is a consequence of the fact that T is bijective.

If the equation (4.1) is not well-posed, then it is called ill-posed.

As we mentioned, there are many ways of solving equation (4.1) under the condition that it is well posed. But if it has no solution, then the best thing that one can think of is to look for a unique element with some prescribed properties that solves

 https://doi.org/10.26219/heal.aueb.6279

a "similar" problem which is well posed. This syllogism derives us to the following definition.

Definition 4.3.1. *Let $T : X \to Y$ be a bounded linear operator.*

a)The $u \in X$ is called least squares solution of $Tx = y$ if $\inf\{\|Ax - y\| : x \in D(A)\} = \|Au - y\|$

b)$u \in X$ is called best approximate solution of $Tx = y$ if x is least squares solution and $\|u\| = \inf\{\|x\| :$ x is least squares solution of Tx=y$\}$

The best approximate solution is also called a pseudosolution of an Operator equation.
An equation of the form of (4.1) which possesses a unique pseudosolution is also a well-posed equation.Otherwise as we mentioned before is an ill-posed equation.

The peudosolution is closely related to the Moore-Penrose inverse $T^\dagger$ of T. This topic will be discussed in our next section.

4.4 Generalized inverses in RKHS

In this section, our main target is to verify the necessary and sufficient conditions for the existence of a solution under the topology that an RKHS provides.
But before we proceed to that, we will present some useful tools, which are strictly weaved to inverse problems.

A large variety of inverse problems can be formulated by integral operators and considering that integral operators are compact we thought that it will be useful to give the definition of compact operator.

Definition 4.4.1. *Let X and Y be normed spaces.A linear transformation $T \in L(X, Y)$ is compact if, for any bounded sequence $\{x_n\}$ in X, the sequence $\{Tx_n\}$ in Y contains a convergence subsequence.For simplicity a transformation T is compact if it takes bounded sets in X into pre-compact sets in Y*

Our next step before proceeding to the main result of this section to introduce a class of operators named by Hilbert-Schmidt operators with many helpful properties and applications.

Definition 4.4.2. *Let H be an infinite dimensional Hilbert space with an orthonormal basis $\{e_n\}$ and let $T \in B(H)$.If the condition*

$$\sum_{i=1}^{\infty} \|Te_n\|^2 < \infty$$

https://doi.org/10.26219/heal.aueb.6279

holds then T is a Hilbert-Schmidt operator.

The following theorem will provide us some useful consequences. The first one is that the above definition does not depend on the choice of the orthonormal basis and secondly and more importantly that there is a connection between Hilbert-Schmidt and compact operators, which are playing an important role especially when problems of the form of (4.1) are considered.

Theorem 4.4.3. *Let H be an infinite dimensional Hilbert space and let $\{e_n\}$ and f_n be orthonormal bases for H.Let $T \in B(H)$.*

a)$\sum_{i=1}^{\infty} \|Te_n\|^2 = \sum_{i=1}^{\infty} \|T^* f_n\|^2 = \sum_{i=1}^{\infty} \|T f_n\|^2$
where the values of these sums may be either finite or ∞.

b)T is Hilbert-Schmidt if and only if T^* is Hilbert-Schmidt.

c)If T is Hilbert-Schmidt then it is compact

d)The set of Hilbert-Schmidt operators is a linear subspace of $B(H)$

Let us now introduce the notion of the generalized inverse(Moore-Penrose).

Definition 4.4.4. *The Moore-Penrose inverse $T^\dagger$ of $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ is defined as the unique linear extension of $\tilde{T}^{-1}$ to*

$$D(T^\dagger) := R(T) + R(T)^\perp \tag{4.2}$$

with

$$N(T^\dagger) = R(T)^\perp \tag{4.3}$$

where

$$\tilde{T} = T_{|N(T)^\perp} : N(T)^\perp \to R(T) \tag{4.4}$$

Proposition 4.4.5. *Let P and Q be orthogonal projectors onto $N(T)$ and $R(T)$ respectively. Then,*

a)The Moore-Penrose equations

$$TT^\dagger T = T \tag{4.5}$$

 https://doi.org/10.26219/heal.aueb.6279

(4.6) $$T^\dagger T T^\dagger = T^\dagger$$

(4.7) $$T^\dagger T = I - P$$

(4.8) $$T^\dagger T = Q|_{D(T^\dagger)}$$

hold true

(4.9) $$R(T^\dagger) = N(T)^\perp$$

Proof. From the definition of $T^\dagger$, for all for all $y \in D(T^\dagger)$,

$$T^\dagger y \;=\; \tilde{T}^{-1}Qy \;=\; T^\dagger Qy$$

so that

$$T^\dagger y \in R(\tilde{T}^{-1}) = N(T^\dagger).$$

Now, for all $x \in N(T^\dagger)$ we have that

$$T^\dagger T x \;=\; \tilde{T}^{-1}\tilde{T}x \;=\; x.$$

This proves that $R(T^\dagger = N(t)^\perp)$.

Also, for any $y \in D(T^\dagger)$, (4.9) implies that

$$TT^\dagger y \;=\; TT^\dagger Qy \;=\; T^{-1}\tilde{T}Qy = \tilde{T}^{-1}\tilde{T}Qy = Qy$$

since $\tilde{T}^{-1}Qy \in N(T)^\perp$. Consequently (4.8) holds.

By definition of $T^\dagger$ we have that for all $x \in X$:

$$T^\dagger T x \;=\; \tilde{T}^{-1}T(Px - (I-P)x) = \tilde{T}^{-1}TPx + \tilde{T}^{-1}T(I-P)x = (I-P)x$$

thus (4.7) holds. Moreover (4.5) comes out directly from (4.7):

$$TT^\dagger T = T(I-P) = T - TP = T$$

Finally, (4.9) arises directly from (4.8) and (4.6)

$$T^\dagger T T^\dagger = T^\dagger Q|_{D(T^\dagger)} = T^\dagger$$

□

https://doi.org/10.26219/heal.aueb.6279

Theorem 4.4.6. *Let $y \in D(T^\dagger)$, then the equation*

$$Tx = y, \quad T: X \to Y \tag{4.10}$$

has a unique best approximate solution which is given by $T^\dagger y$. The set of all least-squares solutions is $T^\dagger y + N(T)$

Proof. Let

$$S = \{z \in X : Tz = Qy\}$$

since $y \in D(T^\dagger) = R(T) + R(T^\dagger)$, $Qy \in R(T) \Rightarrow S \neq \emptyset$. As the orthogonal projector Q is also a metric projector, we have for all $z\imath S$ and for all $x \in X$:

$$\|Tz - y\| = \|Qy - y\| \leq \|Tx - y\|.$$

So, all elements in S ale least squares solutions of (4.10).
Conversely, let z be a l.s.s. of (4.10). Then

$$\|Qy - y\| \leq \|Tz - y\| = \inf\{\|u - y\| : u \in R(T)\} = \|Qy - y\|$$

Thus, Tz is the closest element to y in $R(T)$,i.e. $Tz = Qy$ and

$$S : \{x \in X : \text{x is least squares solution of Tx=y}\} \neq \emptyset$$

Now, let $\bar{z}$ be the element of minimal norm in the closed linear manifold $S = T^{-1}(\{Qy\})$. Since then $S = \bar{z} + N(T)$, it suffices to show that

$$\bar{z} = T^\dagger y.$$

As an element of minimal norm in $S = \bar{z} + N(T)$, $\bar{z}$ is orthogonal to $N(T)$, i.e. $\bar{z} \in N(T)$. This implies that

$$\bar{z} = (I - P)\bar{z} = T^\dagger T\bar{z} = T^\dagger Qy$$
$$= T^\dagger T T^\dagger y = T^\dagger y$$

i.e. $\bar{z} = T^\dagger y$ □

https://doi.org/10.26219/heal.aueb.6279

Theorem 4.4.7. *Let X and Y be Hilbert spaces, X_0 be a subspace of X and T : $X_0 \to Y$ be a closed linear operator. Then*

i)$T^\dagger$ is a closed linear operator, and

ii)$T^\dagger$ is continuous if and only if $R(T)$ is a closed subspace of Y.

Theorem 4.4.8. *Let $y \in D(T^\dagger)$. Then $x \in X$ is a least squares solution of $Tx = y$ if the normal equation*

$$T^*Tx = T^*y \tag{4.11}$$

holds

Proof. x is l.s.s. of (4.10) if and only if Tx is the closest element in $R(T)$ to y, which is equivalent to $Tx - y \in R(T) = N(T^*)$, i.e. to and thus to (4.11) □

From the last theorem we obtain that $T^\dagger$ is the solution of (4.11) of minimal norm and thus

$$T^\dagger = (T^*T)^\dagger T^*$$

Until now ,we have defined some basic notions and tools that provide us some very useful properties. We will try to combine these, in order to present the necessary conditions for the existence of a specific inverse solution, under the topology that an RKHS ensures.

Let H_Q denote the RKHS of real valued functions on the bounded interval S with inner product $< .,. >_Q$ and norm $\|.\|_Q$. Let $Q(s, s')$ be the Reproducing Kernel of H_Q as it has been defined in (2.2). Then the kernel Q induces a self-adjoint Hilbert-Schmidt operator on $\mathscr{L}^2(S)$, the space of square integrable functions on S, by

$$(Qf)(s) = \int_S Q(s, s')f(s')ds'.$$

Furthermore, we have $H_Q \;=\; Q^{1/2}((S)^2)$ and

$$\|f\|_Q \;=\; \inf\{\|p\|_{\mathscr{L}(S)^2} \;:\; p \in \mathscr{L}(S)^2,\; Q^{1/2}p = f\}.$$

For $f \in H_Q$, let $Q^{-1/2}f$ denote the element p of minimal $\mathscr{L}(S)^2$-norm that satisfies $Q^{1/2}p = f$. Then we have

$$< f_1, f_2 >_Q = < Q^{-1/2}f_1, Q^{-1/2}f_2 >_{\mathscr{L}(S)^2}$$

.

https://doi.org/10.26219/heal.aueb.6279

Theorem 4.4.9. *Nashed et al. (1974)Let A be a linear operator from $X = \mathscr{L}(S)^2$ to $Y = \mathscr{L}(T)^2$, where S , T are closed bounded intervals. Assume that A has the following properties :*

i)$H_Q \subset D(A) \subset X$ (throughout the $\subset$ denotes the point-set inclusion only), where H_Q is an RKHS with continuous kernels on $S \times S$
ii)$A(H_Q) \subset H_R \subset H_{\tilde{R}} \subset Y$ where H_R and $H_{\tilde{R}}$ are RKHS with continuous kernels on $T \times T$ and
iii) the null space of A in H_Q is closed in H_Q.

Let $A^{\dagger}_{Q,\tilde{R}}$ denote the generalized inverse of A when A is considered as a map from X into Y (respectively from H_Q into $H_{\tilde{R}}$).Let $y \in D(A^{\dagger}_{Q,\tilde{R}})$.Then $y \in D(Q^{1/2}AQ^{1/2})^{\dagger}_{(X,Y)}\tilde{R}^{-1/2}y$ and

$$A^{\dagger}_{(Q,\tilde{R})}y = Q^{1/2}(\tilde{R}^{-1/2}AQ^{1/2})^{\dagger}_{X,Y}\tilde{R}^{-1/2}y \tag{4.12}$$

The operators Q and $\tilde{R}$ are those induced by the RKHS H_Q and $H_{\tilde{R}}$ respectively.

It is worth noting that an operator A can represent A_Q into a RKHS but this does not guarantee that its range is closed in $\mathscr{L}^2(T)$. Take for example the case that A is a Hilbert-Schmidt integral operator on $\mathscr{L}^2(S)$

4.5 The Representer theorem for linear operator equations

So far we have discussed about the circumstances required for the existence of a generalized inverse solution of a linear operator equation.
In this section we will demonstrate how this solution takes a specific representation under some properties that an RKHS provides.

We assume that H_Q is chosen so that the linear functionals E_t defined by $E_t f = (Af)(t)$ are continuous in H_Q so Theorem 2.2.5 is satisfied.
Then there exists $n_t \in H_Q$ for $t \in T$ such that $(AF)(t) =< n_t, f >_Q$ where

$$n_t(s) =< n_t, Q(., s) >_Q = \ AQ_s(t).$$

Let H_R be the RKHS with kernel

$$R(t,t') =< n_t, n_{t'} >_Q \ , \ t, \ t' \in T.$$

https://doi.org/10.26219/heal.aueb.6279

Let V be the closure of the span of $\{n_t : t \in T\}$ in H_Q. It follows easily that the null space of A $N(A)$ is $V^{\perp}$.
Since $< n_t, n_{t'} >_Q = < R_t, R_{t'} >_R$ where $R_t(t') := R(t,t')$, there is an isometric isomorphism between the subspace V and H_R generated by the correspondence

$$n_t \in V \sim R_t \in H_R.$$

Under this isomorphism we have,

$$f \sim g \iff < n_t, f >_Q = < R_t, g >$$

i.e. $g(t) = (Af)(t)$; $P_V Q_s \sim n_s^* := AQ_s$ where P_V is the orthogonal projector from H_Q onto V.
For $g \in H_R$ let $\hat{f}$ be the element in H_Q of minimal H_Q-norm which satisfies the equation $Af = g$.Then $\hat{f} \in V$ and $g \sim f$.We have the following representations for $\hat{f}$.

Theorem 4.5.1. Nashed et al. (1974)*If all the linear functionals E_t defined by $E_t f = (Af)(t)$ are continuous on H_Q and $g \in H_R$, then $\hat{f}(s) =< Q_s, \hat{f} >_Q = < n_s^\star, g >_R$. Furthermore, if $D(A^*)$ is dense in Y where A^* is the adjoint og A considered as an operator from X to Y, and if H_Q , $H_R = A(H_Q)$ possess continuous kernels, then*

$$A^{\dagger}_{(Q,R)} g \quad = QA^*(AQA^*)^{\dagger}_{X,Y} g$$

4.6 The Representer Theorem for regularized ill-posed equations

Let us return to equation $Ax = y$. If (4.1) has no solution then one can try to minimize the quantity $\|Ax - y\|$ and then enquire if this problem is well-posed.
If not then we are in need of a method through which we can replace the original problem by a class of well posed problems depending on a specific parameter. Such techniques are called regularization methods.

Let $\{R_a\}_a > 0$ be a family of operators in $B(X,Y)$ such that

$$R_a y \to T^{\dagger} y \text{ as } a \to 0$$

for every $y \in T^{\dagger}$. Such family $\{R_a\}_a > 0$ is called regularization family and for each $a > 0$ the

$$x_a(y) \ := R_a(y)$$

is called regularized solution of the ill-posed equation.

https://doi.org/10.26219/heal.aueb.6279

Lemma 4.6.1. *Let X and Y be Hilbert spaces, X_0 be a subspace of X and T $X_0 \to Y$ be a closed linear operator. Then a regularization family $\{R_a\}_{a>0}$ for T is uniformly bounded if and only if $R(T)$ is closed in Y.*

Proof. Suppose $\{R_a\}_{a>0}$ is uniformly bounded. Since $D(T^\dagger)$ is a dense subspace of Y and since $R_a y \to T^\dagger y$ as $a \to 0$ for every $y \in D(T^\dagger)$, it follows that $(R_a y)$ converges as $a \to 0$ for every y in Y to some operator R_0 defined by

$$R_0 y \ := \lim_{a \to 0} R_a y$$

belongs to $B(X, Y)$. But

$$T^\dagger y = R_0 y \; \forall y \in D(T^\dagger)$$

so that $T^\dagger$, the restriction of the bounded operator R_0, is also bounded and closed in Y. Conversely suppose that $R(T)$ is closed in Y. Then $D(T^\dagger) = Y$ and therefore, by the hypothesis, $(R_a y)$ converges as $a \to 0$ for every $y \in Y$. Hence, the family $\{R_a\}_{a>0}$ is uniformly bounded. □

In Hilbert spaces (and so in RKHS), the most commonly used regularization method is the so called Tikhonov regularization. In Tikhonov regularization, the regularized solution

$$x_a(y) := R_a y$$

for $y \in Y$ and $a > 0$ is defined as the unique element which minimized the Tikhonov functional

$$x \to \|Tx - y\|^2 + a\|x\|^2.$$

Theorem 4.6.2. *Let X_0 and T as in Lemma 4.6.1. For each $y \in Y$ and $a > 0$, there exists a unique $x_a(y) \in X_0$ which minimizes the map*

$$x \to \|Tx - y\|^2 + a\|x\|^2, \; x \in X_0.$$

Moreover, for each $a > 0$, the map

$$R_a \ : y \to x_a(y) \ , \ y \in Y$$

is a bounded linear operator from Y to X

https://doi.org/10.26219/heal.aueb.6279

Proof. Consider the product space $X \times Y$ with inner product defined by

$$< (x_1, y_1), (x_2, y_2) >_{X\times Y} = < x_1, x_2 >_X + < y_1, y_2 >_Y$$

for $(x_i, y_i) \in X \times Y$, $i = 1, 2$. It easy seen that $X \times Y$ is a Hilbert space with respect to this inner product. For each $a > 0$, consider the function $F_a \ : X_0 \to X \times Y$ defined by

$$F_a x = (\sqrt[2]{a}x, Tx) \ , \ x \in X_0.$$

Clearly F_a is an injective linear operator. Since the graph of T is a closed subspace of $X \times Y$, it follows that F_a is a closed operator and $R(F_a)$ is a closed subspace of $X \times Y$. Therefore, by Theorem 4.4.6 and Theorem 4.4.7, for every $y \in Y$, there exists a unique $x_a \in X_0$ such that for every $x \in X_0$

$$\begin{aligned} \|Tx_a - y\|^2 + a\|x_a\|^2 &= \|F_a x_a - (0, y)\|_{X\times Y} \\ &\le \|F_a x - (0, y)\|_{X\times Y} \\ &= \|Tx - y\|^2 + a\|x\|^2 \end{aligned}$$

and the generalized inverse $F^{\dagger}a$ of F_a is continuous. In particular the map $y \to x_a \ := F^{\dagger}(0, y)$ is continuous □

Let H_Q and H_P be RKHS with norms $\|.\|_Q$ $\|.\|_P$ respectively. By a regularized pesudosolution of the equation $Af = g$, we mean a solution to the variational problem: Find $f_\lambda \in H_Q$ to minimize

$$\Phi_g(f) = \|Af - g\|_P^2 + \lambda\|f\|_Q \ , \ \lambda > 0 \tag{4.13}$$

$\Phi_g(f)$ takes the value $+\infty$ if $Af - g \notin H_P$
In this section A is a linear operator densely defined on $\mathscr{L}^2(S)$ into $\mathscr{L}^2(T)$, and H_Q must be chosen so that $A(H_Q) = H_R$, where H_R is some RKHS contained as a set in $\mathscr{L}^2(T)$.
Assume $g = g_0 + \xi$ for some $\xi \in H_P$.For $\lambda > 0$, Let $H_{\lambda P}$ be the RKHS with kernel $\lambda P(t, t')$, where $P(t, t')$ is a continuous kernel associated with H_P. We have $H_P = H_{\lambda P}$ and $\|.\|_P^2 = \lambda H_P$.
Let $R(\lambda) = R + \lambda P$ and let $H_{R(\lambda)}$ be the RKHS (following Theorem 3.1.1 sum of reproducing kernel is a reproducing kernel) with kernel $R_\lambda(t, t')$ and norm (again by Theorem 3.1.1)

$$\|g()\|_{R(\lambda)}^2 = \min_{g \,|\, g_0+\xi,\, g_0\in H_R ,\xi\in H_P} (\|g_0\|_R^2 + \|\xi\|_{\lambda P}^2).$$

https://doi.org/10.26219/heal.aueb.6279

It is worth noting that $H_{R(\lambda)}$ is a Hilbert space of functions of the form $g = g_0 + \xi$ (Aronszajn p.352) where $g_0 \in H_P$ and $\xi \in H_Q$.Following Aronszajn this decomposition is not unique unless $H_P \cap H_Q = \{0\}$.

Theorem 4.6.3. Nashed et al. (1974)*Suppose $D(A^*)$ is dense in Y, $H_Q \subset D(A)$ and A and H_Q defined as in Theorem 4.5.1.Suppose H_Q, H_R and $H_P \subset Y$ all have continuous kernels. Then for $g \in H_R(\lambda)$, the unique minimizing element $f_\lambda \in H_Q$ of the functional $\phi_g(f)$ is given by*

$$f_\lambda(s) \;=< AQ_s, g >_{R(\lambda)} = \;(QA^*(AQA^* + \lambda P)^{\dagger}_{Y,Y}g)(s)'$$

The linear mapping which assigns to each $g \in H_{R(\lambda)}$ the unique minimizing element f_λ is called the regularization operator of the equation $Af = g$.

Theorem 4.6.4. *If $H_P \cap H_R = \{0\}$, then the minimizing element f_λ of (3.3) is the solution to the problem: Find $f \in \Omega$ to minimize $\|f\|_Q$, where*

$$\Omega = \{f \in H_Q \;:\; \|Af - g\|_{R(\lambda)} \;=\; \inf_{h \in H_Q} \|Ah - g\|_{R(\lambda)}\}$$

In the setting of this section we have $A(H_Q) \;=\; H_R \subset H_{R(\lambda)} \subset Y$. Replacing $H_{\tilde{R}}$ by $H_{R(\lambda)}$ in (3.2), we get

$$A^{\dagger}_{(Q,R(\lambda))}y \;=\; Q^{1/2}[(R + \lambda P)^{-1/2}AQ^{1/2}]^{\dagger}_{(X,Y)}(R + \lambda P)^{-1/2}y$$

for $y \in D(A^{\dagger}_{(Q,R(\lambda))})$.

Let us denote that the topology on H_R is not, in general the restriction of the topology on $H_R + \lambda P$ except the case that $H_R \cap H_P \;=\; \{0\}$ see section 3.2

4.7 A numeric approach for regularized equations

In this section we will present a numerical approach for regularized operator equations.

Let $T_n \;=\; \{t_1, ..., t_n\}$, where $t_i \in T$, $t_1 < t_2 < ... < t_n$. For a generic function h on T, let $h_n \;=\; (h(t_1), ..., h(t_n))$. Let P_n denote the $n \times n$ matrix whose ij-th element is $P(t_i, t_j)$, and define $\|h_n\|_{P_n} \;=\; \min\{\|e\| \;:\; e \in R^n \;,\; P_n^{1/2} \;=\; h_n\}$, if $h_n \in R(P_n)$ otherwise it takes the value $+\infty$

https://doi.org/10.26219/heal.aueb.6279

Theorem 4.7.1. Nashed et al. (1974) *Let the operator A be as in section 4.6. Let $f_{\lambda,n}$ be the minimizing element in H_Q of the functional $J_n = \|(Af)_n - g_n\|^2_{P_n} + \lambda\|f\|^2_Q$ for $\lambda > 0$. Let $P_{T_n}(\lambda)$ be the orthogonal projector of $H_{R(\lambda)}$ onto the subspace spanned by $\{R_t(\lambda) \ : \ t \in T_n\}$. Then*

$$\begin{aligned} |f_\lambda(s) - f_{(\lambda,n}(s)| &= | < P_{T_n}(\lambda)n^*_s, P_{T_n}(\lambda)g >_{R(\lambda)} | \\ &\leq \|n^*_s - P_{T_n}(\lambda)n^*_s\|_{R(\lambda)}\|g - P_{T_n}(\lambda)g\|_{R(\lambda)}. \end{aligned}$$

Furthermore let $\Delta = max|t_{i+1} - t_i|$, $|f_\lambda(s) - f_{(\lambda,n)}(s)| = O(\Delta^m)$ or $O(\Delta^{2m})$ depending on the smoothness properties of the kernel $R_\lambda(t,t')$ and the functions g and n^*_s. In the particular case when $H_P \cap H_R = 0$, $\{f_{\lambda,n}\}$ converges to $A^\dagger g$.

This simultaneous approach of ill-posed linear operator equations applies to a large class of operator equations such as Fredholm integral equation of the first kind or boundary-value problems. Furthermore, it is worth noting that the Sobolev Spaces W^m_2, a class of functions which will be further discussed in the next chapter are included in the class of spaces considered.

https://doi.org/10.26219/heal.aueb.6279

Chapter 5

Functional Regression in RKHS

5.1 Introduction

This presentation follows closely the recent paper of Yuan et al. (2010). Let Y be a response variable that is related to a square integrable function $X(.)$ by the following model

$$Y = a_0 + \int_T X(t)\beta_0(t)dt + \epsilon \tag{5.1}$$

where a_0 is the intercept, T is the domain of $X(.)$ which is assumed to be compact, β_0 is an unknown slope function which lies in an RKHS $H \subset \mathscr{L}(T)^2$ and ϵ is a centered noise random variable (for example $\epsilon \sim N(0, \sigma^2)$). Our goal is to estimate a_0 and β_0 as well as to retrieve

$$n_0(X) = a_0 + \int_T X(t)\beta_0(t)dt \tag{5.2}$$

through a set $(x_1, y_1), ..., (x_n, y_n)$ of n independent copies of (X, Y).
For the estimation of both n_0 and (a_0, β_0) we will use the regularization method described in Chapter 4. Let l_n be a data fit functional that measures the goodness of fit of the data and J be the penalty functional that assesses the plausibility of n. The estimated $\hat{n}_{n\lambda}$ of n_0 are taken by

$$\hat{n}_{n\lambda} = argmin_n[_n(n|data) + \lambda J(n)] \tag{5.3}$$

where the minimization is taken over

$$\left\{ n \ : \ \mathscr{L}^2(T) \to \mathbb{R} \mid n(X) = a_0 + \int_T X\beta, \ a \in \mathbb{R}, \ \beta \in H \right\} \tag{5.4}$$

and λ be a non negative parameter that intuitively secures the fidelity and the plausibility of the data. Another option is that instead of estimating n_0, we can minimize

 https://doi.org/10.26219/heal.aueb.6279

both (a, β) to obtain estimates for the intercept denoted by $\hat{a}_{n\lambda}$ and the slope function denoted by $\hat{\beta}_{n\lambda}$. In the context of our subject a convenient choice for the data fit functional is the squared error

$$l_n(n) = \frac{1}{n}\sum_{i=1}^{n}[y_i - n(x_i)]^2 \tag{5.5}$$

or any functional which is convex in n and $E(l_n)$ is uniquely minimized by n_0. Also, the penalty functional could be defined through $\beta(t)$ as a squared norm or a semi-norm in H. A good example of H is the Sobolev spaces. For this section we assume that $T = [0, 1]$, the Sobolev space of order m is defined as

$$W_2^m = \{\beta : \ [0,1] \rightarrow \mathbb{R} : \ \beta, \beta^{(1)}, ..., \beta^{(m-1)} \text{are absolutely continuous and } \beta^{(m)} \in \mathscr{L}^2\}$$

with norm

$$\|\beta\|^2_{W_2^m} \ = \sum_{q=0}^{m-1}(\int \beta^{(q)})^2 + (\int \beta^{(m)})^2 \tag{5.6}$$

Or any norm, through which, W_2^m becomes a Sobolev space. In this scenario, a possible choice for the data fit functional is

$$J(\beta) = \int_0^1 [\beta^{(m)}(t)]^2 dt \tag{5.7}$$

Analogously, we can generalize the above form of the penalty functional into 2 dimensions. Consequently, under the assumption that $T = [0, 1]$, a natural choice of J is

$$J(\beta) = \int_0^1 \int_0^1 (\frac{\partial^2 \beta}{\partial x_1^2})^2 + (\frac{\partial^2 \beta}{\partial x_1 \partial x_2})^2 + (\frac{\partial^2 \beta}{\partial x_2^2})^2 dx_1 dx_2 \tag{5.8}$$

which is actually the thin plate spline where (x_1, x_2) are the arguments of a bivariate function β. The readers are referred to Wahba (1990) for more information of this subject.
The main advantage of the aforementioned regularized estimators is that in contrast of other related methods such as the functional principal components analysis for X(.), no assumptions on the spacing of the eigenvalues of the covariance function of X(.) or the Fourier coefficients of β_0 are required. Because of that, we are in position to obtain a better and stronger results on the convergence rates as well as a more efficient computation of the estimators with the help of the representer theorem which will be presented in our next section.
In section 3 we study the relationship between the eigen structures and the covariance operator of X(.) as in Example 3.5.3. We prove that by simultaneously diagonalizing the Covariance operator of X(.) and the RKHS H we obtain a useful tool in order to investigate the minimax rates of convergence.
In section 4 we examine the rates of convergence of the smoothness regularized estimators under a class norms that provides us a unified treatment for the prediction of

https://doi.org/10.26219/heal.aueb.6279

the n_0 and β. The results show that the optimal rates of convergence of the prediction and the estimation could be achieved under weaker conditions than the relative methods used in FPCA.

5.2 Minimizers through the Representer Theorem

In this section we will prove through Theorem 5.2.1 that although our minimization problem of (5.3) could lie over an infinite dimensional space, the solution can be found over a finite dimensional subspace.

Theorem 5.2.1. *Assume that l_n depends on n only through$n(x_1), ..., n(x_n)$. Then there exist $\mathbf{d} = (d_1, ...d_n) \in \mathbb{R}^N$ and $\mathbf{c} = (c_1, ..., c_n)' \in \mathbb{R}^n$ such that*

$$\hat{\beta}_{n\lambda} = \sum_{k=1}^{N} d_k \xi_k(t) + \sum_{i=1}^{n} c_i (K x_i)(t) \tag{5.9}$$

Proof. Let J be the penalty functional, conveniently defined through a squared seminorm such that the null space

$$H_0 := \ \{\beta \in H : \ J(\beta) = 0)\} \tag{5.10}$$

is finite dimensional $dim(H_0) = N$ with orthonormal basis $(\xi_1, .., \xi_N)$.Then, since H_0 is finite dimensional we can define its orthogonal complement as:

$$H_1 \ : \ \{f_1 \in H := < f_0, f_1 >_H = 0, \ \forall f_0 \in H_0\}.$$

But by Theorem 4.2.4 we have that $H = H_0 \oplus H_1$ and any $f \in H$ can uniquely be written as $f = f_0 + f_1$ where $f_0 \in H_0$ and $f_1 \in H_1$. Let K be the Reproducing Kernel of H. Then similarly we can write K as:

$$K = K_0 + K_1 \text{ where } K_0 \in H_0 \text{ and } K_1 \in H_1.$$

Note that by Theorem 3.7.1, H_1 forms a RKHS with $K_1(.,.)$ be it's corresponding reproducing kernel.

Then for any $f \in H_1$ $J(f_1) = \|f_1\|^2_{K_1} = \|f_1\|^2_H$.

As we already mentioned in chapter 1 and 3 any positive continuous function defines an non negative definite operator in $\mathscr{L}^2$ as below

$$(Kf)(.) = \int_T K(., s) f(s) ds \tag{5.11}$$

 https://doi.org/10.26219/heal.aueb.6279

Also for any $f \in \mathscr{L}^2$ and for any $f_0 \in H_0$ we have:

$$\begin{aligned} < Kf, f_0 > &= \int_T \left(\int_T f(s)K_1(t,s)ds \right) f_0(t)dt \\ &= \left(\int_T f_0(t)K_1(s,t)dt \right) f(s)ds = 0 \text{ (since } K_1 \perp f_0) \end{aligned}$$

where the second equality comes directly from Fubini's theorem. Furthermore from the reproducing property of K_1 we have that

$$\int_T f(t)b(t)dt =< Kf, b >_H \tag{5.12}$$

Our goal is to find $\beta_{n\lambda} \in H$ in order to minimize

$$\mathcal{F}(l_n, \beta) = l_n(x_i, \beta) + \lambda J(\beta) \tag{5.13}$$

The basic idea of this proof is trying to take advantage of the fact that l_n depends only through $n(x_i)$. Let $H_{10} = span\{Kx_i\}$ Then similarly to H and following Theorem 4.2.4 we can write

$$H_1 = H_{10} + H_{11}$$

Then any $\beta \in H$ and so can be written as

$$\beta_{n\lambda} = \beta^0_{n\lambda} + \beta^{10}_{n\lambda} + \beta^{11}_{n\lambda} = \sum_{k=1}^{N} d_k \xi_k(t) + \sum_{i=1}^{n} c_i (Kx_i)(t) + \beta_{n\lambda}$$

with $\beta_{n\lambda}$ perpendicular to $\{\xi\}_k$ and $\{K(x_i)\}_i$

Then by (5.12) and the fact that $Kx_i \in H_1$ it comes out that

$$\begin{aligned} \int_T X_i(t)\beta(t)dt &= \int_T (X_i(t)(\beta_0(t) + \beta_{10}(t) + \beta_{11}(t))dt \\ &= \int_T X_i(t)\,(\beta_0(t) + \beta_{10}(t))dt = \\ &=< X_i, \beta_0 > + < Kx_i, \beta_{10} > \end{aligned} \tag{5.14}$$

Also by the properties of J and (5.14) we have that

$$\begin{aligned} \mathcal{F}(l_n, \beta_{n\lambda}) &= l_n(y_i, \beta_{n\lambda}) + \lambda J(\beta_{n\lambda}) = \\ &= l_n(y_i, \beta^0_{n\lambda} + \beta^{10}_{n\lambda} + \beta^{11}_{n\lambda}) + \lambda J(\beta^0_{n\lambda} + \beta^{10}_{n\lambda} + \beta^{11}_{n\lambda}) = \\ &= l_n(y_i, \beta^0_{n\lambda} + \beta^{10}_{n\lambda}) + \lambda J(\beta^0_{n\lambda}) + \lambda J(\beta^{10}_{n\lambda}) + \lambda J(\beta^{11}_{n\lambda}) = \\ &= l_n(y_i, \beta^0_{n\lambda} + \beta^{10}_{n\lambda}) + \lambda J(\beta^{10}_{n\lambda}) + \lambda J(\beta^{11}_{n\lambda}) \geq l_n(y_i, \beta^0_{n\lambda} + \beta^{10}_{n\lambda}) + \lambda J(\beta^{10}_{n\lambda}) \\ &= \mathcal{F}(l_n, \hat{\beta}_{n\lambda}) \end{aligned}$$

 https://doi.org/10.26219/heal.aueb.6279

The result follows. □

For example, take as $n(x_i)$ the squared error loss. Then the regularized estimator is given by

(5.15)

$$(\hat{a}_{n\lambda}, \hat{\beta}_{n\lambda}) = argmin_{a\in\mathbb{R},\, \beta\in H}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \left(a + \int_T x_i(t)\beta(t)dt\right)\right]^2 + \lambda J(\beta)\right\}$$

Taking derivatives with respect to a we get

$$\begin{aligned}\frac{\partial\mathcal{F}}{\partial a} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial a}\left[y_i - \left(a + \int_T x_i(t)\beta(t)dt\right)\right]^2 = \\ &= -\frac{2}{n}\sum_{i=1}^{n}\left[y_i - \left(a + \int_T x_i(t)\beta(t)dt\right)\right] \qquad (5.16)\end{aligned}$$

The minimizer $\hat{a}_{n\lambda}$ arises from

$$\begin{aligned}\frac{\partial\mathcal{F}}{\partial a} = 0 &=> na = \sum_{i=1}^{n} y_i - \int\sum_{i=1}^{n} x_i(t)\beta(t)dt \\ &=> \hat{a}_{n\lambda} = \bar{y} - \int_T \bar{x}(t)\hat{\beta}_{n\lambda}d \qquad (5.17)\end{aligned}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$.

Consequently, (5.15) yields

(5.18)

$$\hat{\beta}_{n\lambda} = argmin_{\beta\in H}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[(y_i - \bar{y}) - \left(a + \int_T (x_i(t) - \bar{x}(t))\beta(t)dt\right)\right]^2 + \lambda J(\beta)\right\}$$

A good example for applying the above results is the space W_2^2 with $J(\beta) = \int_T (\beta'')^2 = 0$.

Then $H_0 := \left\{\beta \in H : \int_T (\beta'')^2)\right\}$, that is the functions $\beta(t)$ of the form $\beta = a_0 + a_1 t$. Thus H_0 is the space spanned by $\xi_1(t) = 1$ and $\xi_2(t) = t$. A popular Reproducing Kernel associated with H_1 is

$$K(s,t) = \frac{1}{2!}B_2(s)B_2(t) - \frac{1}{4!}B_4(|s-t|), \qquad (5.19)$$

where $B_m(.)$ is the m-th Bernoulli polynomial. (The readers are referred to Wahba(1990) for further information of this subject)

Now following 5.2.1, $\beta(t)$ is of the following form.

$$\beta(t) = d_1 + d_2 t + \sum_{i=1}^{n} c_i \int_T [x_i(s) - \bar{x}(s)]K(t,s)ds \qquad (5.20)$$

for some $d \in \mathbb{R}^2$ and $c \in \mathbb{R}^n$. Correspondingly,

$$\begin{aligned}
&\int_T (X(t) - \bar{x}(t))\beta(t)dt = \\
&= d_1 \int_T (X(t) - \bar{x}(t))dt + d_2 \int_T (X(t) - \bar{x}(t))tdt + \\
&+ \sum_{i=1}^{n} c_i \int_T \int_T [x_i(s) - \bar{x}(s)]K(t,s)[X(t) - \bar{x}(t)]dsdt
\end{aligned}$$

Moreover for β given in (5.20)

$$J(\beta) = c'\Sigma c \tag{5.21}$$

where $\Sigma = \Sigma_{ij}$ is a $n \times n$ matrix whose (i,j) entry is

$$\Sigma_{ij} = \int_T \int_T [x_i(s) - \bar{x}(s)]K(t,s)[x_j(t) - \bar{x}(t)]dsdt \tag{5.22}$$

Using Fubini's theorem ,it is easy to see that Σ_{ij} is symmetric. Denote by $T = T_{ij}$ an $n \times 2$ matrix whose (i,j) entry is

$$T_{ij} = \int [x_i(t) - \bar{x}(t)]t^{j-1}dt \tag{5.23}$$

for $j = 1, 2$. Set $\mathbf{y} = (y_1, ..., y_n)'$. Then

$$l_n + \lambda J(\beta) = \frac{1}{n}\|\mathbf{y} - (T\mathbf{d} + \Sigma\mathbf{c})\|_{l_2}^2 + \lambda\mathbf{c}'\Sigma\mathbf{c} \tag{5.24}$$

which is equivalent to

$$\begin{aligned}
&\frac{1}{n}(y - Td - \Sigma c)'(y - Td - \Sigma c) + \lambda c'\Sigma c \\
&= \frac{1}{n}(y' - d'T' - c'\Sigma')(y - Td - \Sigma c) + \lambda c'\Sigma c \\
&= \frac{1}{n}(y'y - y'Td - y'\Sigma c - d'T'y + d'T'y + dT'Td + d'T'\Sigma c - \\
&- c'\Sigma'y + c'\Sigma'y + c'\Sigma'Td + c'\Sigma'\Sigma c) + \lambda c'\Sigma c
\end{aligned}$$

Our goal is to minimize the above equation with respect to c and d. Since all the sets of solutions lead to the same estimate (Gu 2002), we will assume that T is a full rank matrix and Σ is non singular.
Taking derivatives with respect to c and d it comes out that :

$$\frac{\partial \mathcal{F}}{\partial d} = \frac{2}{n}(-T'y + T'Td + T'\Sigma c)$$

and

$$\frac{\partial \mathcal{F}}{\partial c} = \frac{2}{n}(-\Sigma'y + \Sigma'Td + \Sigma'\Sigma c) + 2\lambda\Sigma c$$

https://doi.org/10.26219/heal.aueb.6279

Taking derivatives equal to zero yields :

$$\frac{\partial \mathcal{F}}{\partial d} = 0 => T'y = T'Td + T'\Sigma c \tag{5.25}$$

and

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial c} = 0 => \Sigma' y = \Sigma' Td + \Sigma' Wc => y = Td + Wc \\ => \hat{c} = W^{-1}(y - Td) \end{aligned} \tag{5.26}$$

Where $W = \Sigma + n\lambda I$ substituting (5.26) to (5.25) yields,

$$\begin{aligned} & T'y = T'Td + T'\Sigma W(y - Td) => \\ & T'(I - \Sigma W^{-1})y = T'(I - \Sigma W^{-1})Td => \\ & d = \left[T'(I - \Sigma W^{-1})T\right]^{-1} T' \left[T'(I - \Sigma W^{-1})y\right] \end{aligned} \tag{5.27}$$

But

$$\begin{aligned} I - \Sigma W^{-1} &= WW^{-1} - \Sigma W^{-1} = \\ &= (W - \Sigma)W^{-1} = \\ &= n\lambda W^{-1} \end{aligned} \tag{5.28}$$

Substituting (5.28) to (5.27) it comes out that

$$\mathbf{d} = (T'W^{-1}T)^{-1}T'W^{-1}y \tag{5.29}$$

Finally, c arises by substituting (5.29) into (5.26), that is

$$\mathbf{c} = W^{-1}[I - T(T'W^{-1}T)^{-1}T'W^{-1}]y \tag{5.30}$$

If we write $W = \Sigma + n\lambda I$, then the minimizer of (5.24) is given by

$$\mathbf{d} = (T'W^{-1}T)^{-1}T'W^{-1}\mathbf{y},$$

$$\mathbf{c} = W^{-1}[I - T(T'W^{-1}T)^{-1}T'W^{-1}]\mathbf{y}$$

5.3 Simultaneous Diagonization

In this section we will examine the eigen structures of the Covariance operator $X(.)$ and the Reproducing Kernel of the functional space H.
Let K be the Reproducing Kernel of H_1. As we mentioned in section (1.5), it follows from Theorem 2.5.6 K can be written with the above spectral decomposition.

$$K(s,t) = \sum_{k=1}^{\infty} \rho_k \psi_k(s) \psi_k(t) \tag{5.31}$$

https://doi.org/10.26219/heal.aueb.6279

Where $\rho_1 \geq \rho_2 \geq ....$ are the eigenvalues of K, and $\{\psi_1, \psi_2, ...\}$ are the corresponding eigenfunctions as in Definition 2.5.1 such that

$$< \psi_i, \psi_j >_{\mathscr{L}^2} = \delta_{ij} \text{ and } < \psi_i, \psi_j >_K = \delta_{ij}/\rho_j \tag{5.32}$$

For example consider the Sobolev space $H = W^2_m([0,1])$ with norm (5.6) and penalty (5.7). As in our previous section we define H_0 as

$$\begin{aligned} H_0 = \{f_0 \in H :\ J(f) = \int_T \left[f_0^{(m)}\right]^2 = 0\} = \\ = \{f_0 \in H :\ \left[f_0^{(m)}\right]^2 = 0\} = \\ = \{f_0 \in H :\ f_0^{(m)} = 0\} \end{aligned} \tag{5.33}$$

Then

$$\begin{aligned} H_1 &= \left\{ f_1 \in H : \sum_{q=1}^{m-1} \int_T f_0^{(k)} \int_T f_1^{(k)} + \int f_0^{(m)} f_1^{(m)} = 0,\ \forall f_0 \in H_0 \right\} = \\ &= \left\{ f_1 \in H : \sum_{q=0}^{m-1} \int_T f_0^{(k)} \int_T f_1^{(k)} = 0,\ \forall f_0 \in H_0 \right\} \\ &= \left\{ f_1 \in H : \int_T f_1^{(k)} = 0,\ \forall k = 0, 1, ..., m-1 \right\} \end{aligned} \tag{5.34}$$

Then the reproducing Kernel of H_1 is Wahba (1990)

$$K(s,t) = \frac{1}{(m!)^2} B_m(s) B_m(t) + \frac{(-1)^{m-1}}{(2m)!} B_2 m(|s-t|) \tag{5.35}$$

with $\rho_k \asymp k^{-2m}$ (Micchelli and Wahba(1981)), where $a_k \asymp b_k$ means that a_k/b_k is bounded away from 0 and ∞ as $k \to \infty$.
Let C be the Covariance operator i.e.

$$E\{[(X(s) - E(X(s))][(X(t) - E(X(t)))]\} \tag{5.36}$$

It is also Known that there is a duality between RKHS and Covariance operators (Stein(1999)). Furthermore, under same assumptions for K, we can write C as

$$C(s,t) = \sum_{k=1}^{\infty} m_k \phi_k(s) \phi_k(t) \tag{5.37}$$

where $\{\phi_k, \mu_k\},\ k \geq 1$ are the eigenfunctions and the corresponding eigenvalues such that

$$C\phi_k = \int_T C(.,t)\phi_k(t)dt = \mu_k \phi_k \tag{5.38}$$

The decay rates of μ_k can be determined through some smoothness conditions about C which are called "Sacks-Ylvisacker conditions".

https://doi.org/10.26219/heal.aueb.6279

Definition 5.3.1. *Denote by*

$$\Omega_+ = \{(s,t) \in (0,1)^2 \ : \ s > t\}$$

$$\Omega_- = \{(s,t) \in (0,1)^2 \ : \ s < t\}$$

Let $cl(A)$ be the closure of the set A. Suppose that L is a continuous function on $\Omega_+ \cup \Omega_-$ such that $L|_{\Omega_j}$ is continuously expendable to $cl(\Omega_j)$ for $j \in \{+,-\}$. By L_j we denote the extension of L to $[0,1]^2$ which is continuous on $cl(\Omega_j)$, and on $[0,1]^2 cl(\Omega_j)$. Furthermore write $M^{(k,l)}(s,t) = (\frac{\partial^{k+l}}{\partial^s \partial^t} M(s,t)$. We say that a covariance function M on $[0,1]^2$ satisfies the Sacks-Ylvisacker conditions of order r if the following conditions hold :

(A)$L = M^{(r,r)}$ is continuous on $[0,1]^2$, and its partial derivatives up to order 2 are continuous on $\Omega_+ \cup \Omega_-$, and they are continuously extendable to $cl(\Omega_+)$, and $cl(\Omega_+)$

(B)

$$\min_{0 \le s \le 1} L_-^{(1,0)}(s,s) - L_+^{(1,0)}(s,s) > 0 \tag{5.39}$$

(C)$L_+^{(2,0)}(s,.)$ belongs to the Reproducing Kernel Hilbert Space spanned by L, furthermore

$$\sup_{0 \le \le 1} \|L_+^{(2,0)}(s,.)\|_L < \infty \tag{5.40}$$

More precisely C satisfies the Sacks-Ylvisacker conditions of order s, $s \in \mathbb{N}$, then $\mu_k \asymp k^{-2(s+1)}$[Sacks and Sacks-Ylvisacker(1966,1968,1970)]. In general, a covariance operator C is said to satisfy the Sacks-Ylvisacker conditions of order r, $r \in \mathbb{N}$ if

$$\frac{\partial^{2r} C(s,t)}{\partial s^r \partial t^r}$$

satisfies the Sacks-Ylvisacker conditions of order 0. Moreover, we highlight that there is a connection between Sobolev spaces and covariance functions which will turn to be very useful when simultaneously diagonalizing K and C. It is worth noting that although $\{\psi_k\}$ and $\{\phi_k\}$ may be different, the corresponding Kernels K and C can be simultaneously diagonalized. In order to ensure that El_n is uniquely minimized we shall assume that $Cf \neq 0$ for any $f \neq 0$.

We can define a norm $\|.\|_R$ in H by

$$\|f\|_R =< Cf, f >_{\mathscr{L}^2} + J(f) = \int_{T \times T} f(s) C(s,t) f(t) ds dt \tag{5.41}$$

https://doi.org/10.26219/heal.aueb.6279

which is in quadratic form and zero if $f = 0$ and therefore, it is a norm.

Proposition 5.3.2. *If $Cf \neq 0$ for any $f \in H_0$ and $f \neq 0$, then $\|.\|_R$ and $\|.\|_R$ are equivalent (i.e. there exists positive constants c_1, c_2 such that $c_1\|f\|_R \leq \|f\|_H \leq c_2\|f\|_R$)*

Equivalence between the $\|.\|_H$ and $\|.\|_R$ guaranties us that any convergence of a sequence with respect to $\|.\|_R$ implies its convergence with respect to $\|.\|_H$. In other words all the topological properties of the $\|.\|_R$ are inherited to the $\|.\|_H$

Let R the Reproducing Kernel associated with $\|.\|_R$. As in section (1.5) we can define a linear map $R : \mathscr{L}^2 \to \mathscr{L}^2$ such that

$$R\psi'_k = \int_T R(., t)\psi'_k(t)dt = \rho'_k\psi'_k \tag{5.42}$$

Where $\{\psi'_k, \rho'_k\}, k \leq 1$ are the eigenfunctions and the corresponding eigenvalues of R. As R is positive definite and continuous we can define the square root $R^(1/2)$ of R as the linear map from $\mathscr{L}^2$ to $\mathscr{L}^2$ such that

$$R^{1/2}\psi'_k = (p'_k)^{1/2}\psi'_k \tag{5.43}$$

Let $\{\zeta_k, \nu_k\}$ the orthogonal eigenfunctions in $\mathscr{L}^2$ and the corresponding eigenvalues of the bounded linear operator $R^{1/2}CR^{1/2}$.If we write $w_k = \nu_k^{-1/2}R^{1/2}\zeta_k,\ k = 1, 2, ...,$ then it is not hard to see that

(5.44)

$$< w_j, w_k >_R = \nu_j^{-1/2}\nu_k^{-1/2} < R^{1/2}\zeta_j, R^{1/2}\zeta_k >_R = \nu_k^{-1} < \zeta_j, \zeta_k >_{\mathscr{L}^2} = \nu_k^{-1}\delta_{jk},$$

Where the second equality comes from Lemma 2.5.9. Moreover,

$$\begin{aligned} < C^{1/2}w_j, C^{1/2}w_k >_{\mathscr{L}^2} &= \nu_j^{-1/2}\nu_k^{-1/2} < C^{1/2}R^{1/2}\zeta_j, C^1/2R^{1/2}\zeta_k >_{\mathscr{L}^2} \\ &= \nu_j^{-1/2}\nu_k^{-1/2} < R^{1/2}C^{1/2}R^{1/2}\zeta_j, \zeta_k >_{\mathscr{L}^2} \\ &= \nu_j^{1/2}\nu_k^{-1/2} < \zeta_j, \zeta_k >_{\mathscr{L}^2} = \delta_{jk} \end{aligned}$$

Thus we extract

$$< C\omega_j, \omega_k >= \delta_{jk} \tag{5.45}$$

where $< ., . >_R$ is the inner product associated with $\|.\|_R$, that is, for any $f, g \in H$,

$$< f, g >_R = \frac{1}{4}(\|f + g\|_R^2 - \|f - g\|_R^2) \tag{5.46}$$

The following theorem validates our previous claim that quadratic forms $\|f\|_R^2$ and $< Cf, f >_{\mathscr{L}^2}$ can be simultaneously diagonalized through the basis $\{w_k, k \geq 1\}$.

 https://doi.org/10.26219/heal.aueb.6279

Theorem 5.3.3. *For any $f \in H$,*

$$f = \sum_{k=1}^{\infty} f_k w_k, \tag{5.47}$$

in the absolute sense where $f_k = \nu_k < f, w_k >_R$. Furthermore, if $\gamma_k = (\nu_k^{-1} - 1)^{-1}$,

$$< f, f >_R = \sum_{k=1}^{\infty} (1 + \gamma_k^{-1}) f_k^2 \text{ and } < Cf, f >_{\mathscr{L}^2} = \sum_{k=1}^{\infty} f_k^2 \tag{5.48}$$

Consequently,

$$J(f) = < f, f >_R - < Cf, f >_{\mathscr{L}^2} = \sum_{k=1}^{\infty} \gamma^{-1} f_k^2 \tag{5.49}$$

Proof. Recall that ζ_k is forms an orthonormal base in $\mathscr{L}^2$. Therefore one can write

$$\begin{aligned} R^{-1/2} f &= \sum_{i=1}^{\infty} < R^{-1/2} f, \zeta_k >_{\mathscr{L}^2} = \sum_{i=1}^{\infty} < R^{-1/2} f, \nu_k R^{-1/2} \omega_k >_{\mathscr{L}^2} \nu_k^{1/2} R^{-1/2} \\ &= R^{-1/2} \left(\sum_{i=1}^{\infty} < R^{-1/2} f, R^{-1/2} \omega_k >_{\mathscr{L}^2} \right) \omega_k = R^{-1/2} \left(\sum_{i=1}^{\infty} \nu_k < f, \omega_k >_R \omega_k \right) \end{aligned} \tag{5.50}$$

Applying $R^{1/2}$ to both sides yields to

$$f = \sum_{i=1}^{\infty} \nu_k < f, \omega_k >_R \omega_k \tag{5.51}$$

Moreover

$$\begin{aligned} \|f\|_R &= \left\langle \sum_{k=1}^{\infty} \nu_j \nu_k < f, \omega_k >_R \omega_k, \sum_{j=1}^{\infty} \nu_j < f, \omega_j >_R \omega_j \right\rangle = \\ &= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \nu_j \nu_k < f, \omega_k >_R < f, \omega_j >_R <, \omega_j, \omega_k >_R = \\ &= \sum_{k=1}^{\infty} \nu_k < f, \omega_k >_R^2 = \sum_{k=1}^{\infty} \nu_k^{-1} f_k^2 \end{aligned} \tag{5.52}$$

where the last inequality comes directly from (5.44). Thus by setting $\gamma_k = (\nu_k^{-1} - 1)^{-1}$, we obtain that

$$\nu_k^{-1} = 1 + \gamma_k^{-1} \tag{5.53}$$

and therefore $\|f\|_R$ becomes

$$\sum_{k=1}^{\infty}(1+\gamma_k^{-1})f_k^2.$$

Similarly with the help of (5.45) one can prove that

$$< Cf, f >_{\mathscr{L}^2}= \sum_{k=1}^{\infty} \nu_k^2 < f, \omega_k >_R^2 \tag{5.54}$$

Indeed,

$$\begin{aligned}
< Cf, f >_{\mathscr{L}^2} &= \left\langle C\left(\sum_{k=1}^{\infty} \nu_k < f, \omega_k >_R \omega_k\right), \sum_{j=1}^{\infty} \nu_j < f, \omega_j >_R \right\rangle_{\mathscr{L}^2 \omega_j} = \\
&= \left\langle \sum_{k=1}^{\infty} \nu_k < f, \omega_k >_R C\omega_k, \sum_{j=1}^{\infty} \nu_j < f, \omega_j >_R \omega_j \right\rangle_{\mathscr{L}^2} = \\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \nu_j \nu_k < f, \omega_j >_R < f, \omega_k >_R < C\omega_k, \omega_j >_R \\
&= \sum_{k=1}^{\infty} \nu_k^2 < f, \omega_k >_R^2 .
\end{aligned}$$

□

In addition if we set $f_k = \nu_k < f, w_k >_R$ and $g_k = \nu_k < f, w_k >_R$, by (5.46)

$$\begin{aligned}
< f, g >_R &= \frac{1}{4}(\|f+g\|_R^2 - \|f-g\|_R^2) = \\
&= \frac{1}{4}\sum_{k=1}^{\infty} \nu_k(< f+g, \omega_k >_R^2 - < f-g, \omega_k >_R^2) = \\
&= \sum_{k=1}^{\infty} \nu_k^{-1} f_k g_k = \sum_{k=1}^{\infty}(1+\gamma_k^{-1}) f_k g_k \qquad (5.55)
\end{aligned}$$

Proposition 5.3.4 will provide us useful relationship for the determination of $\{\gamma_k, w_k \; k \geq 1\}$ through $\{\gamma_k, w_k \; k \geq 1\}$ and $\{\rho_k, \psi_k \; : k \geq 1\}$ in the special case $\psi_k = \phi_k$ (i.e. C and K are commutable)

Proposition 5.3.4. *Assume that* $\psi_k = \phi_k, \;\; k = 1, 2, ...$ *then* $\gamma_k = \rho_k \mu_k$ *and* $w_k = \mu_k^{-1/2} \psi_k$.

Proof. For any $f \in H_R$,

$$< f, g >_R = \frac{1}{4}(\|f+g\|_R^2 - \|f-g\|_R^2) =$$
$$= \frac{1}{4}(2 < Cf, g >_{\mathscr{L}^2} +2 < Cg, f >_{\mathscr{L}^2}) + \frac{1}{4}(J(f+g) - J(f-g)) =$$
$$=< Cf, g >_{\mathscr{L}^2} + < f, g >_{K_1} =$$
$$= \int_{T\times T} f(s)C(s,t)g(t)dsdt \tag{5.56}$$

Let $\Phi = span\{\phi_k, k \geq 1\}$. Then, since $Cf \neq 0 <=> f \neq 0$ for any $f \in H_0$ it comes out that $H_0 \cap \Phi^\perp = \{0\}$. Moreover from the fact that $H_0 \cap H_1 = \{0\}$ we conclude that $H = H_1 = \Phi$.

Also by (5.56)

$$< \psi_j, \psi_k >_R= (\mu_k + \rho_k^{-1})^{-1}\delta_{jk}, \tag{5.57}$$

which implies that $\{(\mu_k + \rho_k^{-1})^{-1}, \psi_k : k \geq 1\}$ is also an eigen-system of R. Thus by Theorem 2.5.6 we have:

$$R(s,t) = \sum_{i=1}^{\infty}(\mu_k + \rho_k^{-1})^{-1}\psi_k(s)\psi_k(t) \tag{5.58}$$

Then

$$R\psi_k = \int_T R(.,t)\psi_k dt = (\mu_k + \rho_k^{-1})^{-1}\psi_k(s) \tag{5.59}$$

Therefore,

$$R^{1/2}CR^{1/2}\psi_k = R^{1/2}C(\mu_k + \rho_k^{-1})^{-1/2}\psi_k =$$
$$= R^{1/2}C\mu_k(\mu_k + \rho_k^{-1/2})\psi_k =$$
$$= (1 + \rho_k^{-1})^{-1}\psi_k \tag{5.60}$$

Therefore $\zeta_k = \psi_k = \phi_k$, $\nu_k = (1 + \rho_k^{-1}\mu_k^{-1})^{-1}$ and $\gamma_k = \rho_k\mu_k$. Consequently,

$$\omega_k = \nu_k^{-1/2}R^{1/2}\psi_k = \nu_k^{-1/2}(\mu_k + \rho_k^{-1})^{-1/2}\psi_k = \mu_k^{-1/2}\psi_k \tag{5.61}$$

□

Theorem 5.3.5 will reveal us the asymptotic behavior of γ_k under specific conditions .

 https://doi.org/10.26219/heal.aueb.6279

Theorem 5.3.5. *Consider the one dimensional case when $T = [0,1]$. If H is the Sobolev space $W_2^m([0,1])$ endowed with norm (5.6), and C satisfies the Sacks-Ylvisacker conditions, then $\gamma_k \asymp \mu_k \rho_k$*

5.4 Convergence Rates

In this section we will study the asymptotic properties of the regularized estimators defined in section 5.2. We will focus on the squared error loss defined in (5.5) which leads us to the above form of the estimators

(5.62)

$$(\hat{a}_{n\lambda}, \hat{\beta}_{n\lambda}) = argmin_{a\in\mathbb{R},\, \beta\in H} \left\{ \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \left(a + \int_T x_i(t)\beta(t)dt\right)\right]^2 + \lambda J(\beta) \right\}.$$

In that case after after computing $\hat{\beta}_{n\lambda}$ with the help Theorem 5.2.1, one can easily compute $\hat{a}_{n\lambda}$ which is of the form

$$\hat{a}_{n\lambda} = \bar{y} - \int_T \bar{x}(t)\hat{\beta}_{n\lambda} \tag{5.63}$$

So our main problem to the minimization of (5.62) is the computation of $\hat{\beta}_{n\lambda}$. Derived by that, we will pay more attention into studying the asymptotic properties of $\hat{\beta}_{n\lambda}$. In addition, we will assume that the eigenvalues of K satisfies $\rho_k \asymp k^{-2r}$ for some $r > 1/2$.

Let $\mathcal{F}(s, M, K)$ be the collection of distributions F of the process X with covariance operator C that satisfy the following conditions

a) the eigenvalues μ_k of C satisfy $\mu_k \asymp k^{-2s}$ for some $s > 1/2$.

b) for any $f \in \mathscr{L}^2(T)$,

(5.64)

$$E\left(\int_T f(t)\left[X(t) - (EX)(t)\right]dt\right)^4 \le M\left[E\left(\int_T f(t)[X(t) - (EX)(t)]\right)^2\right]^2$$

c) When simultaneously diagonalizing K and C, $\gamma_k \asymp \rho_k\mu_k$, where $\nu_k = (1+\gamma_k^{-1})^{-1}$ is the k-th largest eigenvalue of $R^{1/2}CR^{1/2}$ with R be the Reproducing Kernel associated with (5.41).

Condition (a) is connected with the smoothness of the sample path of $X(.)$. The second condition is actually the forth moment of a compact linear functional $(\mathcal{G}X)(t) = \int_T X(t)f(t)$. If we assume that $X(t)$ is a Gaussian process, then this condition is

https://doi.org/10.26219/heal.aueb.6279

satisfied for $M = 3$ since the $\int f(t)X(t)dt$ is normally distributed. It is clear that condition (c) is satisfied if K and C have the same the eigenfunctions (i.e. they commute). An interesting result arises if we assume that $H = W_m^2$. Then since $\gamma_k \asymp \rho_k \mu_k$ and by Theorem 5.3.5, condition (c) is satisfied by any covariance function C that satisfies the Sacks-Ylvisaker conditions defined in section 4.3.

For $0 \leq a \leq 1$ we define a norm $\|.\|_a$

$$\|f\|_a = \sum_{k=1}^{\infty}(1+\gamma_k^{-a})f_k \tag{5.65}$$

where $f_k = \nu_k < f, \omega_k >_R$ as in Theorem 5.3.3. In addition, Similarly to Equation 5.55, one can prove that

$$< f, g >_a = \sum_{k=1}^{\infty}(1+\gamma_k^{-a})f_k g_k \tag{5.66}$$

Observe that

$$\|f\|_0^2 = 2\sum_{k=1}^{\infty} f_k^2 = 2 < Cf, f >_{\mathscr{L}^2} \tag{5.67}$$

and

$$\begin{aligned}\|f\|_1^2 &= \sum_{k=1}^{\infty}(1+\gamma_k^{-1})f_k^2 = \\ &= \sum_{k=1}^{\infty} f_k^2 + \sum_{k=1}^{\infty}\gamma_k^{-1}f_k^2 = < Cf, f >_{\mathscr{L}^2} + J(f) \\ &= \|f\|_R^2 \end{aligned} \tag{5.68}$$

where the last equality comes from Theorem 5.4.1. Intuitively, it will be convenient to consider $\|.\|_a$ as an interpolated norm between $< Cf, f >$ and $J(f)$. As one could notice from the above results, the main advantage of $\|.\|_a$ is that it has been constructed in such manner that it will give us a more unified treatment for $< Cf, f >$ and $\|.\|_K$ and therefore for both the prediction and estimation error. Moreover we denote that the optimal rates of convergence given in Theorem 5.4.1 are valid for $0 \leq a \leq 1$.

Theorem 5.4.1. *Assume that $E(\epsilon_i) = 0$ and $Var(\epsilon_i) \leq M_2$. Suppose the eigenvalues ρ_k of the reproducing kernel K_1 of the RKHS H satisfy $\rho_k \asymp k^{-2r}$ for some $r > 1/2$. Then the regularized estimator $\hat{\beta}_{n\lambda}$ with*

$$\lambda \asymp n^{-2(r+s)/2(r+s)+1} \tag{5.69}$$

satisfies

$$\lim_{D\to\infty} \overline{\lim_{n\to\infty}} \sup_{f\in\mathcal{F}(s,M,K),\beta_0\in H} P(\|\hat{\beta}_{n\lambda} - \beta_0\|_a^2) > Dn^{-2(r+s)(1-a)/2(r+s)+1} = 0 \tag{5.70}$$

https://doi.org/10.26219/heal.aueb.6279

It is worth noting that the optimal choice of λ does not depend on a. Theorem 5.4.2 will ensure us that the rate of convergence of $\hat{\beta}_{n\lambda}$ (i.e. $n^{-2(r+s)(1-a)/2(r+s)+1}$) is indeed optimal.

Theorem 5.4.2. *Under the assumptions of Theorem 5.4.1, there exists a constant $d > 0$ such that*

$$\underline{\lim}_{n\to\infty} \inf_{\tilde{\beta}\in\mathcal{B}} \sup_{f\in\mathcal{F}(s,M,K),\beta_0\in H} P(\|\tilde{\beta}_{n\lambda} - \beta_0\|_a^2) > dn^{-2(r+s)(1-a)/2(r+s)+1} > 0 \tag{5.71}$$

Consequently, the regularized estimator $\hat{\beta}_{n\lambda}$ with $\lambda \asymp n^{-2(r+s)/2(r+s)+1}$ is rate optimal.

Proof. The proof follows a similar argument as of that of Hall and Horowitz (Hall et al. (2007)). Consider a setting where $\psi_k = \phi_k, k = 1, 2, ...$ Clearly in this case we also have $\omega_k = \mu_k \phi_k$. It suffices to show that the rate is optimal in this case. Recall that $\phi_{k_{k\geq 1}}$ forms an orthonormal base. Let $\beta_0 = \sum \alpha_k \phi_k$ where

$$a_k = \begin{cases} L_n^{-1/2} k^{-r}\theta_k, & L_n + 1 \leq k \leq 2L_n \\ 0 & otherwise \end{cases} \tag{5.72}$$

where L_n is the integer part of $n^{1/2(r+s)+1}$, and θ_k is either 0 or 1. It is clear that

$$\|\beta_0\|_{K_1}^2 \leq \sum_{L_n+1}^{2L_n} L_n^{-1} = 1 \tag{5.73}$$

Therefore $\beta_0 \in H$. Now let X admit the following expansion: $\sum \xi_k k^{-s}\phi_k$ (i.e. Karhunen-Loeve expansion) where $\{\xi_k\}_{k\geq 1}$ are independent random variables drawn from a uniform distribution on $[-\sqrt{3}, \sqrt{3}]$. Simple algebraic manipulation shows that the distribution X belongs to $\mathcal{F}(s, 3)$. The observed data are

$$y_i = \sum_{k=L_n+1}^{2L_n} L_n^{-1/2} k^{(-r+s)} \xi_{ik}\theta_k + \epsilon_i, \quad i = 1, 2, .., n \tag{5.74}$$

where the noise is assumed to be independently sampled from $N(0, M_2)$. As shown in Hall et al. (2007)

$$\lim_{n\to\infty} \inf_{L_n\leq j\leq 2L_n} \inf_{\tilde{\theta}_j} \overset{\star}{\sup} E(\tilde{\theta}_j - \theta_j)^2 \geq 0 \tag{5.75}$$

 https://doi.org/10.26219/heal.aueb.6279

where $\sup^\star$ denotes the over all 2^{L_n} choices of $(\theta_{L_{n+1}}, ..., \theta_{2L_n})$, and $\inf_{tilde\theta}$ is taken over all measurable functions $\tilde{\theta}_j$ of data. Therefore, for any estimate $\tilde{\beta}$,

$$\begin{aligned}\sup^\star \|\beta \tilde{-} \beta_0\|_a^2 &= \sup^\star \sum_{k=L_n+1}^{2L_n} L_n^{-1} k^{-2(1-a)(r+s)} E(\theta_j - \theta_j)^2 \\ &\geq Mn^{-2(1-a)(r+s)/2(r+s)+1} \qquad (5.76)\end{aligned}$$

for some constant $M > 0$. Denote

$$\tilde{\tilde{\theta}}_k = \begin{cases} 1, & \tilde{\theta}_k \geq 1 \\ \tilde{\theta}_k, & 0 \leq \tilde{\theta}_k \leq 1 \\ 0, & \theta_k \leq 0 \end{cases} \qquad (5.77)$$

It is easy to see that

$$\sum_{k=L_n+1}^{2L_n} L_n^{-1} k^{-2(1-a)(r+s)} (\tilde{\theta}_j - \theta_j)^2 \geq \sum_{k=L_n+1}^{2L_n} L_n^{-1} k^{-2(1-a)(r+s)} (\tilde{\tilde{\theta}}_j - \theta_j)^2 \qquad (5.78)$$

Hence, we can assume that $0 \leq \theta_j \leq 1$ without loss of generality in establishing the lower bound. Subsequently,

$$\begin{aligned}\sum_{k=L_n+1}^{2L_n} L_n^{-1} k^{-2(1-a)(r+s)} (\tilde{\theta}_j - \theta_j)^2 &\leq \sum_{k=L_n+1}^{2L_n} L_n^{-1} k^{-2(1-a)(r+s)} \\ &\leq L_n^{-2(1-a)(r+s)} \qquad (5.79)\end{aligned}$$

Together with (5.76), this implies that

$$\lim_{n\to\infty} \inf_{\tilde{\beta}} \sup^\star P(\|\tilde{\beta} - \beta\|_a^2 > dn^{-2(1-a)(r+s)/2(r+s)+1}) > 0 \qquad (5.80)$$

for some constant $d > 0$ □

Denote by $\mathcal{B}$ the collection of all measurable functions of the observations $(X_1, Y_1), ..., (X_n, Y_n)$.If $X^\star$ is an independent copy of X. By (5.67) we have,

$$\|\tilde{\beta} - \beta_0\|_0^2 =< C\left(\tilde{\beta} - \beta_0\right), \tilde{\beta} - \beta_0 >_{\mathscr{L}^2}$$

which by (3.12) takes the form

$$\begin{aligned}E_{X^\star}(< X, \tilde{\beta} - \beta_0 >_{\mathscr{L}^2} < X, \tilde{\beta} - \beta_0 >_{\mathscr{L}^2}) &= E_{X^\star}\left(\int \left[\tilde{\beta}(t) - \beta_0(t)\right] X^\star(t)dt\right)^2 \\ &= E_{X^\star}\left(\int \tilde{\beta}(t)X^\star(t)dt - \int \beta_0(t)X^\star(t)dt\right)^2 \qquad (5.81)\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6279

Observe that the right hand side of (5.81) is actually the mean squared prediction error of $\mathcal{G} = \int X^\star \tilde{\beta}(t)$ in regression method. (5.4.3) will give us an interesting result about this quantity. To be accurate (5.4.3) will reveal us that there is an analogue between the convergence rate of the eigenvalues of the covariance operator C and the the mean squared prediction error.

Corollary 5.4.3. *Under the assumptions of Theorem 5.4.1, the mean squared optimal prediction error of a slope function estimator over $\mathcal{H} \in \mathcal{F}(s, M, K)$ and $\beta_0 \in H$ is of order $n^{-\frac{2(r+s)}{2(r+s)+1}}$ and it can be achieved bey the regularized estimator $\hat{\beta}_{n\lambda}$ with λ satisfying (5.69)*

Observe that the fastest the convergence rate of the eigenvalues, the smaller the prediction error.

For purpose of illustration consider as H the Sobolev space $W_m^2([0,1])$ and the Wiener stochastic process $X(.)$. Moreover let $C(s,t) = \min\{s,t\}$ be it's corresponding covariance operator. It is not hard to see that $C(s,t)$ satisfies the Sacks-Ylvisaker conditions of order 0 and therefore $\mu_k \asymp k^{-2}$. But by Corollary 5.4.3 , the minimax rate of the prediction error of estimating β_0 is $n^{-2(m+1)/2(2m+3)}$.

A subject of particular interest is when C and K have the same set of eigen functions. As we mentioned Theorem 5.3.5 when $\psi_k = \phi_k$. Then $\gamma_k = \rho_k \mu_k$ and by the assumptions made in the previous section $\rho_k \asymp k^{-2r}, r > 1/2$ and $\mu_k \asymp k^{-2s}, s > 1/2$. Thus $\gamma_k \asymp k^{-2(r+s)}, k \geq 1$. Consider estimating estimating $\int x^\star \beta_0$ where $x^\star$ satisfies $| < x^\star, \phi_k > | \asymp k^{-s+q}$ for some $0 < q < s - 1/2$ (required to ensure that $x^\star \in \mathscr{L}^2$). Then the squared prediction error

$$\left(\int \tilde{\beta}(t) x^\star(t) dt - \int \beta(t) x^\star(t) dt \right)^2 \tag{5.82}$$

is equivalent to $\|\tilde{\beta} - \beta_0\|^2_{(s-q)/(r+s)}$. It is also clear that $\|.\|_{s/r+s}$ is equivalent to $\|.\|_{\mathscr{L}^2}$ Derived by that and Theorems 5.4.1 and 5.4.2 we have the following Corollaries.

Corollary 5.4.4. *Suppose $x^\star$ is a function satisfying $| < x^\star, \phi_k >_{\mathscr{L}^2} |$ for some $0 < q < s - 1/2$. Then under the assumptions of Theorem 5.4.1*

(5.83)

$$\lim_{n\to\infty} \inf_{\tilde{\beta}\to\infty} \sup_{f\in\mathcal{F}(s,M,K)} P\left\{ \left(\int \tilde{\beta}(t) X^\star(t) dt - \int \beta_0(t) X^\star(t) dt \right)^2 > dn^{-2(r+q)/2(r+s)+1} \right\} > 0$$

for some constant $d > 0$, and the regularized estimator $\hat{\beta}_{n\lambda}$ with λ satisfying (5.69) achieves the optimal rate of convergence under the prediction error (5.82).

 https://doi.org/10.26219/heal.aueb.6279

Corollary 5.4.5. *if $\phi_k = \psi_k$ for all $k \geq 1$, then under the assumptions of Theorem 5.4.1*

$$\lim_{n\to\infty} \inf_{\tilde{\beta}\to\infty} \sup_{f\in\mathcal{F}(s,M,K)} P\left(\|\hat{\beta}_{n\lambda} - \beta_0\|^2_{\mathscr{L}^2} > dn^{-2r/2(r+s)+1}\right) > 0 \tag{5.84}$$

for some constant $d > 0$ and the regularized estimate $\beta_{n\lambda}$ with λ satisfying (5.69) achieves the optimal rate.

In contrast with the mean squared prediction error, the optimal convergence rates of the estimation error are larger as the eigenvalues of the covariance operator decay faster. If $\beta_0 \in H$ we have

$$\sum_{i=1}^{\infty} \rho_k^{-1} < \beta_0, \psi_k >^2_{\mathscr{L}^2} = \sum_{i=1}^{\infty} \rho_k^{-1} < \beta_0, \phi_k >^2_{\mathscr{L}^2} < \infty \tag{5.85}$$

and if $\rho_k \asymp k^{-2r}$ then the above condition (5.85) is estimated as

$$| < \beta_0, \phi_k >_{\mathscr{L}^2} | \leq M_0 k^{-r-1/2} \tag{5.86}$$

for some constant $M_0 > 0$.

Theorems 5.4.1 and 5.4.2 turn to be very useful for estimating derivatives. A natural way that one can think to estimate the q-th derivative of β_0 is the q-th derivative of $\hat{\beta}_\lambda$. In Corollary 5.4.6, we will show that this holds true, under some specific conditions.
Assume that $\psi_k = \phi_k$ and $\|\psi_k^{(q)}/\psi_k\|_\infty \asymp k^q$ for all $k \geq 1$. This clearly holds when $H = W_2^m$. In this case

$$\|\tilde{\beta}^{(q)} - \beta_0^{(q)}\|_{\mathscr{L}^2} \leq C_0 \|\tilde{\beta} - \beta_0\|_{(s+q)/(r+s)} \tag{5.87}$$

Thus, with the help of Theorem 5.4.1 and Theorem 5.4.2 we obtain the following Corollary

Corollary 5.4.6. *Assume that $\psi_k = \phi_k$ and $\|\psi_k^{(q)}/\psi_k\|_\infty \asymp k^q$ for all $k \geq 1$. Then under the assumptions of Theorem 5.4.1 for some constant $d > 0$*

$$\lim_{n\to\infty} \inf_{\tilde{\beta}\to\infty} \sup_{f\in\mathcal{F}(s,M,K)} P\left(\|\hat{\beta}^{(q)}_{n\lambda} - \beta_0^{(q)}\|^2_{\mathscr{L}^2} > dn^{-2(r-q)/2(r+s)+1}\right) > 0 \tag{5.88}$$

and the regularized estimate $\hat{\beta}_{n\lambda}$ with λ satisfying (5.69) achieves the optimal rate.

https://doi.org/10.26219/heal.aueb.6279

https://doi.org/10.26219/heal.aueb.6279

Bibliography

Adams, R. A. and J. J. Fournier (2003). *Sobolev spaces*, Volume 140. Academic press.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society*, 337–404.

Berlinet, A. and C. Thomas-Agnan (2011). *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media.

Engl, H. W., M. Hanke, and A. Neubauer (1996). *Regularization of inverse problems*, Volume 375. Springer Science & Business Media.

Ferreira, J. and V. Menegatto (2013). Positive definiteness, reproducing kernel hilbert spaces and beyond. *Ann. Funct. Anal 4*(1), 64–88.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics Springer, Berlin.

Hall, P., J. L. Horowitz, et al. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Heckman, N. et al. (2012). The theory and application of penalized methods or reproducing kernel hilbert spaces made easy. *Statistics Surveys 6*, 113–141.

Horn, R. A. and C. R. Johnson (2012). *Matrix analysis*. Cambridge university press.

Kennedy, R. A. and P. Sadeghi (2013). *Hilbert Space Methods in Signal Processing*. Cambridge University Press.

Nair, M. T. (2009). *Linear operator equations: approximation and regularization*. World Scientific.

Nashed, M., G. Wahba, et al. (1974). Regularization and approximation of linear operator equations in reproducing kernel spaces. *Bulletin of the American Mathematical Society 80*(6), 1213–1218.

Paulsen, V. I. (2009). An introduction to the theory of reproducing kernel hilbert spaces. *Lecture Notes*.

Poggio, T. and C. Shelton (2002). On the mathematical foundations of learning. *American Mathematical Society 39*(1), 1–49.

https://doi.org/10.26219/heal.aueb.6279

Small, C. G. and D. L. McLeish (2011). *Hilbert space methods in probability and statistical inference*, Volume 920. John Wiley & Sons.

Wahba, G. (1990). *Spline models for observational data*, Volume 59. Siam.

Yuan, M., T. T. Cai, et al. (2010). A reproducing kernel hilbert space approach to functional linear regression. *The Annals of Statistics 38*(6), 3412–3444.