**Espacios de Hilbert con Núcleo Reproductor: el Teorema del Representante. Interpretación del representante como aproximante en Aprendizaje**

***(Reproducing Kernel Hilbert Spaces: the Representer Theorem. An interpretation of the representer as an approximant in Learning)***


**Víctor Martínez Crespo**


**Trabajo de Fin de Grado**

para acceder al

**Grado en Matemáticas**

Facultad de Ciencias

Universidad de Cantabria

Director: Luis Miguel Pardo Vasallo

Octubre - 2020

*Quiero comenzar dando las gracias a mi familia, en especial, a mi padre, a mi madre, a mi hermano y a mi tía Belén. Por vuestro cariño y vuestra paciencia. Por haber respetado y haberme acompañado en toda decisión que he tomado. Por las comidas en el Jovi y las barbacoas en Oreña. Gracias por haberme ayudado a ser quien quiero ser.*

*Gracias a los profesores que han hecho posible que haya llegado hasta aquí. A Cecilia Valero por dudar inicialmente de mí y conseguir así que aprobase su asignatura. Por tener siempre la puerta del despacho abierta. A Luis Miguel Pardo por creer en mí en todo momento. Por su labor como tutor, profesor, y finalmente, como director de este trabajo. Gracias a ambos por ser inicio y final de esta aventura.*

*Gracias a Antonio por su disposición a ayudarme durante estos años, incluso en este texto. Gracias a Irene por su contagiosa alegría, sea cual sea la situación. Gracias a ambos por su humor* Follow *y por darme los mejores recuerdos y anécdotas de estos años.*

*Gracias a Pablo por las tardes jugando al bádminton. Por su papel como confidente y por su atención. Gracias a Marina por aquella llamada que lo inicio todo. Por ser un apoyo y mi subdelegada en las sombras. Gracias a Iñigo por las cenas en el McDonald's. Por mostrar siempre su mundo interior.*

*Gracias a Eloy y a Daniel por las cenas de pizza de los jueves. Por abrirme las puertas de su piso y hacerme aprender de lo que puede uno encontrar al ver ms allá y cuestionarse las cosas.*

*Gracias a Lucía y a Verónica por recorrer conmigo todas las bibliotecas de Santander. Por mostrarme sus perspectivas y como seguir sus pasos para tomarme la vida con más humor.*

*Gracias a Patricia y a Mónica, mis físicas favoritas. Por ser fuente inagotable de cotilleos y de alegría. Gracias a quien decidió juntar matemáticos y físicos en la misma aula. Algo bueno tuvo.*

*Gracias a Marta, a Pablo y a Alicia, mis eternos* primerines*. Gracias por la confianza depositada al acudir a mí en busca de consejo. Por ser un soplo de aire fresco.*

*Gracias a Miguel por cuidarnos siempre detrás de la barra y crear así un entorno donde poder desconectar y descansar.*

*Se que dejo muchos nombres sin mencionar, no puede ser de otra manera. Gracias a todos los que vinieron, los que se fueron y los que se quedaron. Por todas las partidas al mus, la brisca, el vidas o el juego que tocase ese cuatrimestre. Por haber conseguido, a pesar de todo, que la facultad se sienta hogar.*

*No puedo acabar estos agradecimientos sin mencionar a la persona más importante de estos años. Muchísimas gracias (si, las doy) a Adrián. Por siempre apostar por mí. Por creer en mí y hacerme ver de lo que soy capaz. Por ser el mejor amigo que uno pudiese pedir. Porque la carga pesa menos si se lleva entre dos. Espero que estés orgulloso, porque tenías toda la razón.*

*Lo conseguí.*

*"Simply put, we are not in this alone"*
Chidi Anagonye. *The Good Place.*
*Season 2, Episode 12.*

Abstract. In this manuscript we present a complete and self-contained proof of the main mathematical statement that supports what is known as the “kernel trick” in computational learning: the Representer Theorem. We have chosen the more complete statement of this result due to B. Scholkopf, R. Herbrich and A. J. Smola, published on 2001, involving regularization functions. This statement requires an introduction to reproducing kernel Hilbert spaces (RKHS for short), introduced by J. Mercer in 1909 and strongly developed by N. Aronszajn in 1950. The main theorem of Aronszajn theory is the characterization of RKHS in a statement known as the Aronszajn-Moore Theorem, which we also prove in detail. With these tools we may finally prove the Representer Theorem in reproducing kernel Hilbert spaces of R. Herbrich, B. Scholkopf and A. J. Smola. Additionally, we prove a result by A. Argyriou, C. A. Micchelli and M. Pontil of 2009 which characterizes the Tikhonov regularization functions admissible in a Representer Theorem.
On the other hand, although the Representer Theorem is claimed to be a used tool in Learning, we have not found a precise form to link this statement with the foundations of Learning established by F. Cucker and S. Smale in 2003. Thus, we just filled the gap with our own statements by giving an interpretation of the representer as an approximant of regression functions within the context of Cucker-Smale foundations of learning.
In order to achieve these goals, we first introduce well-known concepts and results on Hilbert spaces theory (Chapter 1). Then, we establish the foundations of reproducing kernel Hilbert spaces, we prove Aronszajn-Moore Theorem and, finally, the Representer Theorem (Chapter 2). Finally, we link this theorem and its application in learning by giving an original result which proves that the Representer Theorem's “output” is a good approximation of our desired regression function (Chapter 3).

Key words: Hilbert spaces, reproducing kernel, Representer theorem, computational learning, approximation error.

Resumen. En este manuscrito presentamos una prueba completa y autocontenida de uno de los principales enunciados matemáticos en lo que se conoce como el “kernel trick” en aprendizaje computacional: el Teorema del Representante. Hemos escogido la versión más completa del enunciado de este resultado por R. Herbrich, B. Scholkopf y A. J. Smola, publicado en 2001, incluyendo funciones de regularización. Este enunciado requiere la introducción de espacios de Hilbert con núcleo reproductor (siglas RKHS), introducido por J. Mercer en 1909 y desarrollado por N. Aronszajn en 1950. El teorema principal de la teoría de Aronszajn es la caracterización de RKHS en un enunciado conocido como el Teorema de Aronszajn-Moore, el cual también probamos. Con estas herramientas probamos finalmente el Teorema del Representante en espacios de Hilbert con núcleo reproductor de R. Herbrich, B. Scholkopf y A. J. Smola. Además, probamos un resultado de A. Argyriou, C. A. Micchelli y M. Pontil de 2009 que caracteriza las funciones de regularización de Tikhonov admisibles en el Teorema del Representante.
Por otro lado, aunque se dice que el Teorema del Representante es usado como herramienta en Aprendizaje, no hemos encontrado una forma precisa de enlazar este enunciado con los fundamentos del Aprendizaje establecidos por F. Cucker y S. Smale en 2003. Por ello, hemos rellenado ese hueco con nuestros propios enunciados dando una interpretación del representante como un aproximante de funciones de regresión en el contesto de los fundamentos del aprendizaje de Cucker y Smale.
Para conseguir estos objetivos, primero introducimos conceptos y resultados conocidos sobre la teoría de espacios de Hilbert (Capítulo 1). Luego, establecemos los fundamentos de los espacios de Hilbert con núcleo reproductor, probamos el Teorema de Aronszajn-Moore y, finalmente, el Teorema del Representante (Capítulo 2). Finalmente, enlazamos este teorema y su aplicación en aprendizaje aportando un resultado original que prueba que el “output” del Teorema del Representante es una buena aproximación de la función de regresión deseada (Capítulo 3).

Palabras clave: Espacios de Hilbert, núcleo reproductor, Teorema del Representante, aprendizaje computacional, error de aproximación.

# Índice

CAPíTULO 0

# Introducción y Resumen de Contenidos de la Memoria

## Índice



### 0.1. Introducción

El propósito de esta memoria es la presentación, con pruebas, del prinicipal resultado matemático que subyace a las técnicas de Aprendizaje Computacional conocido como *"kernel tricks"*: se trata de el *Teorema del Representante*, traducción libre del comúnmente conocido como *Representer Theorem.* Adicionalmente, hemos incluido una interpretación como aproximante de ciertas funciones de regresión del resultado afirmado por el Teorema del Representante.
A continuación, y tras una introducción del contenido del trabajo, se incluyen los resúmenes de las principales contribuciones para ayudar a una más cercana lectura del texto.
El problema inicial del que se parte es el siguiente: Se considera un espacio de Hilbert $H$ de funciones a valores reales o complejos sobre un espacio topológico $X$, es decir, $H \subseteq \mathbb{K}^X$, donde $\mathbb{K} = \mathbb{R} \vee \mathbb{C}$. Se considera una función de riesgo definida sobre $H$ y dependiente para cada $f \in H$ de un número finito de datos que representan puntos de $X$ y potenciales valores de $f$ en esos puntos. Formalmente, para cada $n \in \mathbb{N}$, $n \geq 1$ se considera una *función de riesgo*

$$E_n : H \times (X \times \mathbb{K})^n \longrightarrow \mathbb{R}_+ \cup \{\infty\}.$$

A cada lista $((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{K})^n$ se le denomina datos observables. Supondremos, como hipótesis fundamental del problema, que la función de riesgo $E_n$ *depende solamente de los datos observados y del comportamiento de $f$ sobre estos datos.* Formalmente, esto se puede interpretar como la existencia de una función de riesgo

$$\mathcal{E}_n : (X \times \mathbb{K}^2)^n \longrightarrow \mathbb{R} \cup \{\infty\}. \tag{0.1.1}$$

de tal modo que para cada $f \in H$, y para cada lista de datos observables $((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{K})^n$ se tiene:

$$E_n(f, ((x_1, y_1), ..., (x_n, y_n)) = \mathcal{E}_n((x_1, y_1, f(x_1)), ..., (x_n, y_n, f(x_n)).$$

El objetivo del problema consiste en minimizar el riesgo $E_n$ para una lista de datos observados dada "a priori". Es decir, dada una observación $\omega = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{K})^n$, hallar $f \in H$ tal que

$$E_n(f, \omega) = \min\{E_n(g, \omega) : g \in H\}.$$

En general, la función de riesgo $E_n$ se regulariza añadiendo una *función de regularización* $\Omega : H \longrightarrow \mathbb{R}$. Típicamente, se presupone que existe una función estrictamente creciente $g : [0, \infty) \longrightarrow \mathbb{R}$ tal que

$$\Omega(f) = g(||f||_H), \qquad \forall f \in H,$$

donde $|| \cdot ||_H$ es la norma de $f$ en $H$. Así, el problema de referencia es el siguiente:

Problema. *Sea $H$ un espacio de Hilbert, sea $g : [0, \infty) \longrightarrow \mathbb{R}$ una función estrictamente creciente y sea $E_n : H \times (X \times \mathbb{K})^n \longrightarrow \mathbb{R}$ una función de riesgo tal que existe $\mathcal{E}_n$ como en 0.1.1 de tal modo que $\forall f \in H$, $\forall \omega = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{K})^n$,*

$$E_n(f, \omega) = \mathcal{E}_n((x_1, y_1, f(x_1)), ..., (x_n, y_n, f(x_n))).$$

*Dado $z \in (X \times \mathbb{K})^n$, se pide hallar una función $f_z \in H$ tal que*

$$E_n(f_z, z) + g(||f_z||) = \min\{h \in H : E_n(h, z) + g(||h||)\}.$$

Nótese que se pretende hallar para cada $z \in (X \times \mathbb{K})^n$ un elemento $f_z$ que minimize la suma de $E_n$ con $g(|| \cdot ||)$.
Un ejemplo clásico es el riesgo dado por el error cuadrático medio

$$E_n(f, (x_1, y_1), ..., (x_n, y_n)) = \frac{1}{n}\sum_{i=1}^{n} |f(x_i) - y_i|^2$$

y una función de regularización donde $g(t) = t^2$. En ese caso, pretendemos minimizar sobre $H$ la expresión

$$E_n(f, z) = \frac{1}{n}\sum_{i=1}^{n} |f(x_i) - y_i|^2 + ||f||_H^2.$$

La respuesta al Problema precedente se dividiría en dos partes:

i) *Problema de Consistencia o Satisfactibilidad*: Bajo ciertas hipótesis sobre $H$, $E_n$, $g$ y $z$, ¿existe $f_z \in H$ que minimiza $E_n(\cdot, z) + \Omega$ ?.

ii) *Problema de Resolución*: Sabiendo que el problema de minimización es consistente, hallar un $f_z$ que minimize $E_n(\cdot, z) + \Omega$.

El Teorema del Representante que tratamos aquí no habla del problema de consistencia sino del problema de resolución: básicamente, se pretende mostrar cómo dado un problema de minimización consistente podemos expresar el mínimo de manera simple, de tal modo que, "a posteriori", existieran algoritmos capaces de calcular una función que minimice el riesgo $E_n(\cdot, z) + \Omega$.
La solución más natural pasa por transformar el problema de minimización en $H$ en un problema de minimización (tipicamente no lineal) en un espacio de Hilbert de dimensión finita a través de una proyección ortogonal. Por tanto, el Problema de Resolución pasa a ser el siguiente:

PROBLEMA (de Resolución). *Sean $E_n$, $H$, $\Omega$ como en el enunciado del Problema 0.1. Sea $z \in (X \times \mathbb{K})^n$ y supongamos que el problema de minimizar $E_n(\cdot, z) + \Omega$ es consistente en $H$. Hallar un subespacio de dimensión finita $V_z \subseteq H$ tal que $\exists f_z \in V_z$ verificando*

$$E_n(f_z, z) + \Omega(f_z) = \min\{E_n(h, z) + \Omega(h) : h \in H\}.$$

*Más áun, si fuera posible, dar un sistema generador de $V_z$ en el que expresar $f_z$.*

La respuesta al Problema de Resolución pasa por un resultado inicialmente introducido en 1970 por [**KW, 70**], posteriormente estudiado por [**Gi, 18**] y generalizado en [**HSS, 01**]: el *Teorema del Representante.*

TEOREMA 0.1.1 (del Representante). *Con las hipótesis precedentes, si $H = H_K$ es un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{K}$, el Problema de Resolución admite respuesta positiva siendo $V_z$ el $\mathbb{K}$-subespacio vectorial de $H_K$ generado por*

$$\{K_{x_i} : 1 \leq i \leq n\},$$

*donde*

$$\begin{aligned} K_{x_i} : X &\longrightarrow \mathbb{K} \\ y &\longmapsto K(x_i, y). \end{aligned}$$

Este resultado así enunciado hace referencia a los *espacios de Hilbert con núcleo reproductor* (en acrónimo RKHS). Los espacios de Hilbert con núcleo reproductor fueron introducidos por J. Mercer en 1909 (cf. [**Me, 09**]) y estudiados en profundidad por N. Aronszajn en 1950 (cf. [**Ar, 50**]). A partir del Teorema del Representante, estos espacios de Hilbert con núcleo reproductor se convierten en un tema central del Aprendizaje Computacional.
En este TFG, consideramos la teoría de Espacios de Hilbert con núcleo reproductor y desarrollamos con detalles algunos de sus teoremas esenciales, concluyendo con una prueba del Teorema del Representante, asi como un recíproco (caracterización de las funciones de regularización admisibles) siguiendo [**AMP, 09**]. Este desarrollo se realiza en el Capítulo 2 y que resumiremos en la Sección 0.2 de esta Introducción.
Usualmente, la intervención del Teorema del Representante en el Aprendizaje Computacional

no es siempre explícita en la literatura sobre el tema. Más aún si se considera dentro del marco abstracto del Aprendizaje Computacional fijado por F. Cucker y S. Smale en su artículo fundamental [CS, 02]. Estos autores fijan como objetivo del aprendizaje el calculo (o, al menos la aproximación) de una función de regresión

$$f_\rho : X \longrightarrow \mathbb{K},$$

asociada a una distribución de probabilidad $\rho$ sobre $X \times \mathbb{K}$, que minimiza el error cuadrático medio

$$\mathcal{E}(f_\rho) = \int_{X\times\mathbb{K}} |f(x) - y|^2 d\rho.$$

En la tarea de aproximar $f_\rho$ el problema, bajo las técnicas propuestas, se descompone en dos problemas:

i) Hallar un espacio de Hilbert con núcleo reproductor $H_K$ tal que existe $f_H \in H_K$ "próximo" a $f_\rho$.
ii) Una vez fijado $H_K$, aplicar el Teorema de Representación para hallar un $f_z \in V_z$ que aproxime a $f_\rho$.

De nuevo, este Trabajo de Fin de Grado omite discutir el aspecto $i)$ en detalle y asume que ya disponemos de un espacio de Hilbert con núcleo reproductor $H_K$ y que $f_\rho \in L^2_\nu(X)$, donde $\nu = f_X$ es la distribución marginal inducida por $\rho$ sobre $X$, de tal modo que existe $\delta > 0$ verificando:

$$d(f_\rho, H_K)_{L^2_\nu(X)} \leq \delta.$$

Entonces, probaremos que dada una observación $z \in (X \times \mathbb{K})^n$, bajo ciertas hipótesis sobre $K$ y $z$, tendremos que el *output* $f_z$ del Teorema del Representante será un buen aproximante de $f_\rho$, es decir, obtendremos una estimación de la forma

$$||f_\rho - f_z||_{L^2_\nu(X)} < \delta + 2\varepsilon.$$

Este resultado de aproximación se exhibe en el Capítulo 3 y que resumimos en la Sección 0.3 siguiente.

## 0.2. Resumen del Capítulo 2

Tras un capítulo de resultados preliminares y básicos, comunmente conocidos, en este capítulo desarrollamos la teoría básica de espacios de Hilbert con núcleo reproductor y el Teorema del Representante.

El primer resultado fundamental del que nos ocupamos es el Teorema de Aronszajn-Moore sobre la base del Teorema de Riesz-Fréchet (que probamos en la Sección 2.2), damos un teorema que caracteriza los RKHS mediante la propiedad de que la evaluación en un punto es una función lineal continua y acotada. Resumimos formalmente este resultado del modo siguiente:

Sea $\mathbb{K} = \mathbb{R} \vee \mathbb{C}$, sea $X$ un espacio topológico y $H \subseteq \mathbb{K}^X$ un espacio de Hilbert cuyos elementos son funciones de $X$ en $\mathbb{K}$. Para cada $x \in X$, consideremos la función evaluación en $x$

$$\begin{aligned} L_x : H &\longrightarrow \mathbb{K} \\ f &\longmapsto f(x). \end{aligned}$$

Un espacio de Hilbert $(H, \langle\cdot,\cdot\rangle)$ formado por funciones de $X$ en $\mathbb{K}$ (ie, $H \subseteq \mathbb{K}^X$) se denomina *espacio de Hilbert con núcleo reproductor* si para cada $x \in X$, la función $L_x$ es lineal, continua y de norma finita.

Esta definición no hace referencia a la presencia de núcleos sino a la función evaluación en conjunto. Un *núcleo* sobre $X$ es una función $K : X \times X \longrightarrow \mathbb{K}$ que satisface:

i) $K$ es simétrica si $\mathbb{K} = \mathbb{R}$ o sesquilineal si $\mathbb{K} = \mathbb{C}$.
ii) $K$ es definida positiva, es decir, dada cualquier familia finita $\{x_1, ..., x_n\} \subseteq X$ la matriz de Gram de $K$ en esta familia

$$Gram[K](x_1, ..., x_n) \coloneqq (K(x_i, x_j))_{1\leq i,j\leq n}$$

es definida positiva.

El Teorema de Aronszajn-Moore relaciona los espacios de Hilbert con núcleo reproductor con los núcleos de la forma siguiente:

TEOREMA 0.2.1. *Sea $X$ un espacio topológico, $K : X \times X \longrightarrow \mathbb{K}$ un núcleo sobre $X$. Entonces, existe un único (salvo isometría) espacio de Hilbert con núcleo reproductor $H_K \subseteq \mathbb{K}^X$ de tal modo que:*

*i) Para cada $x \in X$, la función*

$$\begin{aligned} K_x : X &\longrightarrow \mathbb{K} \\ y &\longmapsto K_x(y) := K(x, y) \end{aligned}$$

*verifica que $K_x \in H_K$.*

*ii) Para cada $f \in H_K$ y cada $x \in X$, si $L_x : H_K \longrightarrow \mathbb{K}$ es la función de evaluación,*

$$L_x(f) = f(x) = \langle K_x, f \rangle_K,$$

*donde $\langle \cdot, \cdot \rangle_K$ es el producto interno de $H_K$.*

Más aún, probaremos que $H_K$ es el completado del espacio preHilbert $H_0$ generado por la familia

$$\{K_x : x \in X\}.$$

Hemos incluido la condición de espacio topológico sobre $X$, aunque no es necesaria para nuestra demostración.

Un caso particular de núcleos son los núcleos de Mercer: son aquellos núcleos donde $X$ es un espacio topológico y $K : X \times X \longrightarrow \mathbb{K}$ es una función continua. Los núcleos de Mercer, aunque muy citados cotidianamente, son algo más restrictivos. Así, por ejemplo, en el Teorema 2.3.2 probaremos que si $(X, d)$ es un espacio métrico compacto y $K : X \times X \longrightarrow \mathbb{K}$ es un núcleo de Mercer, entonces el espacio de Hilbert $H_K$ es separable y, por tanto, isométricamente identificable con $\ell^2$.

En todo caso, dedicamos la Sección 2.4 del Capítulo a resumir algunos resultados básicos sobre núcleos de Mercer. Hemos elegido la forma resumida para no extender en exceso este trabajo. Los resultados han sido tomados de [**CZ, 07**] y es a él a quien referiremos para las demostraciones. Debe hacerse notar que [**CZ, 07**] establece sus resultados para el caso $(X, d)$ espacio métrico compacto. El resultado más destacable de esta Sección es el Teorema de Mercer (cf. Teorema 2.4.3) en el que se muestra que los espacios de Hilbert con núcleo de Mercer sobre espacios métricos compactos son separables, mediante una descipción explícita de $H_K$ partiendo de una base ortonormal numerable descrita a partir de los valores y vectores propios de un operador integral $L_K$.

Tras este breve resumen sobre núcleos de Mercer, dedicamos una sección a estudiar un resultado sobre el Error de Aproximación entre espacios de Banach, que reusaremos en el capítulo siguiente.

El capítulo concluye con dos secciones:

- La Sección 2.6 que enuncia y demuestra la versión general del Teorema del Representante conforme [**HSS, 01**]
- La Sección 2.7 que discute un pseudorecíproco del Teorema del Representante, conforme a [**AMP, 09**].

Como ya hemos hablado del Teorema del Representante, centrémonos en lo que se discute en la Sección 2.7. El objetivo aquí es tratar de discutir que propiedades deben verificar las funciones de regularización $\Omega : H_K \longrightarrow \mathbb{R}$ para el caso en que el Teorema del Representante se satisface. Para ello se considera el *problema de interpolación sobre espacios de Hilbert con núcleo reproductor.*

PROBLEMA (de interpolación en espacios de Hilbert con núcleo reproductor). *Sea $X$ un conjunto, $H \subseteq \mathbb{C}^X$ un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{C}$. Llamaremos problema de interpolación sobre $H$ con base en $X$ al siguiente: Dada una instancia $z = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{C})^n$ y dada una función de interpolación $\Omega : H \longrightarrow \mathbb{R}$, llamaremos problema de interpolación en $H$ con instancia $z$ y función de regularización $\Omega : H \longrightarrow \mathbb{R}$ al problema de minimizar el siguiente conjunto:*

$$\{\Omega(\omega) : \omega \in H, \omega(x_i) = y_i, 1 \leq i \leq m\}.$$

Se consideran funciones de regularización $\Omega : H \longrightarrow \mathbb{R}$ *admisibles* como aquellas para las que el problema de interpolación es satisfactible (o consistente, ie. admite solución) en un subespacio

generado por $\{K_{x_1}, ..., K_{x_n}\}$. Nótese que el problea de interpolación puede ser visto como una instancia particular del problema de minimización inicial: basta con suponer que la función de riesgo es $E_n = 0$.
En la Sección 2.7 probamos la siguiente caracterización de las funciones de regularización:

TEOREMA 0.2.2 (Caracterización de las regularizaciones de Tikhonov). *Sea $H$ un espacio de Hilbert y $\Omega : H \longrightarrow \mathbb{R}$ una regularización de Tikhonov continua y diferenciable (en el sentido de la Definición 25). Supongamos que $dim(H) \geq 2$. Entonces, $\Omega$ es admisible para el problema de interpolación sobre $H$ si y solamente si existe una función monótona creciente $h : \mathbb{R}_+ \longrightarrow \mathbb{R}$ tal que*

$$\Omega(\omega) = h(||\omega||^2), \forall \omega \in H.$$

## 0.3. Resumen del Capítulo 3

Como se indica en casi todas las referencias al Teorema del Representante, una de sus principales aplicaciones consiste en su utilización en Aprendizaje. Desafortunadamente, no hemos sido capaces de localizar un resultado conclusivo, válido para un contenido abstracto como el de [**CS, 02**], que permita mostrar la manera de usar este resultado para aproximar funciones de regresión. Es notable observar que ni la referencia clásica, ni [**CS, 02**] sean capaces de ligar el Teorema del Representante con los fundamentos matemáticos del Aprendizaje que ellos mismos pretender establecer. Así, la Sección 6 del Capítulo III de [**CS, 02**] introduce el Teorema del Representante aunque no explica en qué medida este resultado está ligado a las páginas que le preceden.
Ante esta situación hemos optado por introducir un resultado original que liga el Teorema del Representante con los principios de Teoría del Aprendizaje tal y como está establecido en [**CS, 02**]. El resultado no es ni óptimo ni satisfactorio, dado que sólo establece condiciones suficientes de aplicabilidad del Teorema del Representante al Aprendizaje sin mostrar casos concretos en los que esa aplicabilidad es además factible. Es sólo una primera aproximación a las palabras usadas regularmente por los autores del ámbito.
Este capítulo se dedica así a enunciar y dar una demostración bastante completa de este teorema original que no hemos podido encontrar en la literatura. Para poder enunciarlo, necesitamos introducir algunas nociones preliminares.
En lo que sigue, $(X, d)$ será un espacio métrico, $I \subseteq \mathbb{R}$ un intervalo, $Z = X \times I \subseteq X \times \mathbb{R}$ y $\mathscr{B}(Z)$ la clase de los conjuntos de Borel del espacio producto $X \times I$. Supondremos $\rho : \mathscr{B} \longrightarrow [0, 1]$ una distribución de probabilidad definida sobre $\mathscr{B}(Z)$. Para cualquier función medible $f : X \longrightarrow \mathbb{R}$, consideramos el error cuadrático promedio de $f$ con respecto a la probabilidad $\rho$ dado mediante

$$\mathcal{E}(f) = \int_Z |y - f(x)|^2 d\rho.$$

El propósito de la teoría del Aprendizaje, conforme a los principios de [**CS, 02**] consiste en minimizar este error cuadrático promedio. Consideremos $\mathscr{B}(X)$ la clase de los conjuntos de Borel sobre $X$ y sea $\rho_X : \mathscr{B}(X) \longrightarrow [0, 1]$ la *distribución marginal de $\rho$ sobre $X$* a través de la proyección canónica $\pi : Z \longrightarrow X$. Es decir, para $E \in \mathscr{B}(X)$

$$\rho_X(E) := \rho(\pi^{-1}(E)) = \rho(E \times I).$$

Con respecto a $\rho_X$ podemos definir el espacio $L^2(X)$ de funciones cuadrado integrales. Es decir, las funciones Borel medibles $f : X \longrightarrow \mathbb{R}$ tales que

$$\int_X |f(x)|^2 d\rho_X < \infty.$$

Bajo la hipótesis de que el espacio métrico $(X, d)$ es separable (i.e, posee un subconjunto denso contable), el teorema de Desintegración de la Medida nos garantiza que existe una función $f_\rho : X \longrightarrow \mathbb{R}$ que minimiza $\mathcal{E}$:

$$\mathcal{E}(f_\rho) \leq \mathcal{E}(f), \qquad \forall f \in L^2(X).$$

La función $f_\rho$ se denomina *función de regresión de $f$ sobre $X$*. Para que todo sea más conforme, la Teoría del Aprendizaje de [**CS, 02**] añade una hipótesis sobre la distribución de probabilidad $\rho$:

DEFINICIÓN 1. *Con las notaciones precedentes una distribución de probabilidad $\rho$ sobre $Z$ es* admisible *para su tratamiento en Teoría del Aprendizaje si la función de regresión $f_\rho$ es acotada sobre $X$. Es decir, $\rho$ admisible si*

$$||f_\rho||_\infty := sup\{|f_\rho(x)| : x \in X\} < \infty.$$

Bajo la hipótesis de la Teoría del Aprendizaje, si $\rho$ es una distribución de probabilidad admisible se tiene $f_\rho \in L^2(X)$ :

$$\int_X |f_\rho(x)|^2 d\rho_X \leq \int_X ||f_\rho||_\infty^2 d\rho_X = ||f_\rho||_\infty^2 \int_X d\rho_X = ||f_\rho||_\infty^2 < \infty.$$

Más aún, si $|| \cdot ||_{L^2(X)}$ es la norma en el espacio de Hilbert $L^2(X)$, tendremos que

$$||f_\rho||_{L^2(X)} \leq ||f_\rho||_\infty.$$

Por tanto, bajo la hipótesis de la Teoría del Aprendizaje, $f_\rho \in L^2(X)$ es la función que minimiza $\mathcal{E}$ sobre $L^2(X)$. De hecho, probamos, en la Proposición 3.2.3, que para cada $f \in L^2(X)$ se tiene

$$\mathcal{E}(f) - \mathcal{E}(f_\rho) = |\mathcal{E}(f) - \mathcal{E}(f_\rho)| = ||f - f_\rho||_{L^2(X)}^2.$$

Como, normalmente, $f_\rho$ no es fácilmente computable, se hace una aproximación de $f_\rho$ mediante un subespacio de Hilbert $H \subseteq L^2(X)$ con buenas propiedades de aproximación de $f_\rho$ que, al mismo tiempo, sean manipulables.

DEFINICIÓN 2. *Con las notaciones precedentes, sea $\delta > 0$ un número real positivo y $H \subseteq L^2(X)$ un subespacio de Hilbert. Diremos que $H$ posee una* $\delta$-aproximación de la función de regresión *$f_\rho : X \longrightarrow \mathbb{R}$ anterior si existe $f_H \in H$ tal que*

$$||f_\rho - f_H||_{L^2(X)}^2 < \delta.$$

En la Proposición 3.2.5 probaremos que dado un espacio de Hilbert $H$ que contiene una $\delta$-aproximación de $f_\rho$ y siendo $f_H \in H$ un elemento que minimiza $\mathcal{E}$ sobre $H$ se tiene:

$$||f_\rho - f_H||_{L^2(X)}^2 < \delta,$$
$$|\mathcal{E}(f_H) - \mathcal{E}(f_\rho)| < \delta.$$

En este sentido, $f_H \in H$ sería una buena aproximación de $f_\rho$ y, por tanto, una buena codificación de la función de regresión a través de elementos de $H$.

No es el propósito de este Trabajo de Fin de Grado hacer una disquisición sobre la búsqueda de buenos espacios de Hilbert para aproximar la función de regresión. Nuestro objetivo es más modesto: *pretendemos aproximar $f_H$ a partir del Teorema del Representante.* Sin embargo, para que nuestra exposición no quedase corta en este aspecto, hemos descrito en la Subsección 3.2.4 una breve discusión sobre los ratios de Aproximación en el caso de espacios de Hilbert definidos por núcleos de Mercer.

A partir de aquí nos ocupamos solamente del caso en que $H$ sea un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$ y que $H$ contenga una $\delta$-aproximación de $f_\rho$. Nuestro objetivo será calcular $f_H$ o, en su defecto, aproximar $f_H$ y aquí es donde interviene el Teorema del Representante. Para ello, sea $z = ((x_1, y_1), ..., (x_n, y_n)) \in Z^n$ una observación empírica de un evento donde $n \in \mathbb{N}, n > 1$. Para cada $f \in H$, llamaremos error empírico de $f$ en $z$ a la cantidad

$$\mathcal{E}_z(f) = \frac{1}{n} \sum_{i=1}^{n} |y_i - f(x_i)|^2.$$

No queda claro, a priori, que el error empírico tenga relación con $\mathcal{E}(f)$ para clases amplias de funciones. Por ello, dado $\epsilon > 0$ y dado $\mathcal{F} \subseteq H$ definimos:

$$H_\rho^{(n)}(\mathcal{F}, \varepsilon) := \{z \in Z^n : \sup_{f \in \mathcal{F}} |\mathcal{E}_z(f) - \mathcal{E}(f)| < \epsilon\}.$$

Las grandes contribuciones de V. N. Vapnik y A. Ya. Chervonenkis (cf. [**CV, 71**]), para $\mathcal{F}$ formado por funciones características de conjuntos de Borel de $X$, o de autores como D. Pollard (cf. [**Po, 84**]), con el uso de la esperanza de números de recubrimiento, consisten en mostrar clases $\mathcal{F}$ para las cuales $H_g^{(n)}(\mathcal{F}, \mathcal{E}) \neq \emptyset$. Más aún, los teoremas de estos autores son de la forma

$$Prob\left(H_\rho^{(n)}(\mathcal{F}, \varepsilon)\right) \geq 1 - N_\rho(\mathcal{F}, \varepsilon, n) e^{-\frac{\epsilon^2 m}{D}},$$

para una cierta constante $D$ y una cierta función $N_\rho(\mathcal{F}, \mathcal{E}, n)$ usualmente denominada *número de recubrimiento*. En la Subsección 3.2.6 mostramos un ejemplo de una cota inferior de este tipo basada en el método combinatorio de Pollard y ciertas disgresiones estudiadas en [**PS, 20**].
Para concluir nuestras notaciones, si $x = (x_1, ..., x_n) \in X^n$, denotaremos por $G_x$ a la matriz de Gram asociada al núcleo $K$ y a la lista de puntos de $x$. Es decir,

$$G_x = \begin{pmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{pmatrix} = (K(x_i, x_j))_{1 \le i,j \le n},$$

Denotamos por $(G_x{}^{\frac{1}{2}})$ a la matriz raíz cuadrada de $G_x$, por $\mu(G_x{}^{\frac{1}{2}})$ su número de condicionamiento y por $||G_x^{\frac{1}{2}}||_F$ su norma de Frobenius.
Finalmente, dado $z \in Z^n$, denotamos por $f_z \in H$ a un elemento que minimiza el error empírico $\mathcal{E}_z$. Nuestro primer resultado es el siguiente, que acota la norma en $H$ del $f_z$ dado por el Teorema del Representante.

Teorema 0.3.1. *Con las notaciones precedentes, sea $(X, d)$ un espacio métrico, $H$ un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$. Sea $z = ((x_1, y_1), ..., (x_n, y_n)) \in Z^n$ y sea*

$$W_z := Span(\{K_{x_1}, ..., K_{x_n}\}),$$

*el subespacio vectorial de $H$ generado por $\{K_{x_1}, ..., K_{x_n}\}$. Entonces, existe $f_z \in W_z$ tal que*

*i)*

$$\mathcal{E}_z(f_z) = \min\{\mathcal{E}_z(f) : f \in H\}$$

*ii)*

$$||f_z||_H \le \frac{\mu(G_x^{\frac{1}{2}})}{||G_x^{\frac{1}{2}}||_F} ||y||_2.$$

Este resultado (Teorema 3.1.1) se prueba en la Sección 3.1. Utilizando todas estas herramientas, probamos finalmente el siguiente enunciado en la Sección 3.2.

Teorema 0.3.2. *Sea $(X, d)$ un espacio métrico separable, $I \subseteq \mathbb{R}$ un intervalo, $Z = X \times I$ y $\rho$ una distribución de probabilidad sobre los subconjuntos de Borel de $Z$. Supongamos que $\rho$ es admisible para el aprendizaje y que la norma $||f_\rho||_\infty$ de la función de regresión es finita. Sean $R, \varepsilon, \delta \in \mathbb{R}$ números reales con $0 < \varepsilon, 0 < \delta < 1$. Sea $H$ un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$ que contiene una $\delta$-aproximación $f_H \in H$ de $f_\rho$. Supongamos que $R$ satisface las siguientes propiedades:*

*i)* $H_g^{(n)}(B_H(0, R), \varepsilon) \neq \emptyset$, *donde $B_H(0, R)$ es la bola en $H$ de centro $0$ y radio $R$.*
*ii)* $||f_\rho||_\infty \le R - 1$.

*Supongamos que existe $z = ((x_1, y_1), ..., (x_n, y_n)) \in H_g^{(n)}(B_H(0, R), \varepsilon)$ tal que*

$$\mu(G_x{}^{\frac{1}{2}})||y||_2 \le R||G_x{}^{\frac{1}{2}}||_F.$$

*Entonces, existe $f_z \in W_z = Span(\{K_{x_1}, .., K_{x_n})$ tal que*

$$||f_z - f_\rho||^2_{L^2(X)} \le 2\varepsilon + \delta.$$

$$|\mathcal{E}(f_z) - \mathcal{E}(f_\rho)| \le 2\varepsilon + \delta.$$

## 0.4. Sobre el estilo y la ortografía usados en este TFG

En algún caso precedente se ha discutido el estilo y la ortograf ía de las memorias presentadas como Trabajo de Fin de Grado en Matemáticas. En evitación de intervenciones innecesarias, queremos clarificar algunos aspectos relativos al estilo elegido en este texto.
Se ha elegido el formato de libro (book) de la *American Mathematical Society (AMS)*. Aunque el idioma utilizado es el español, hemos tratado de seguir lo más fielmente posible las recomendaciones del Libro de Estilo de esta asociación[1], juntamente con las reglas de estilo recomendadas

[1]M. Letourneau, J. Wright Sharp, *AMS Style Guide, Journals, October 2017*, AMS, Providence, 2017

por D. E. Knuth y co-autores para la *Mathematical Asociation of America (MAA)*[2]. Específicamente, hemos tratado de seguir atentamente las siguientes dos reglas:

- *"Numbered theorems, lemmas, etc. are proper nouns and, thus, are capitalized: Theorem 2.3, Lemma 3.1, Figure 4.5"* (p. 79 del *AMS Style Guide*).
- *"Rule 19. Capitalize names like Theorem 1, Lemma 2, Algorithm 3, Method 4"'* (en D. E. Knuth *et al.*).

## 0.5. Apéndices Finales

El trabajo presentado se finaliza con 2 Apéndices para simplificar su lectura. El Apéndice A.1 contiene parte de las pruebas de algunos resultados de la Sección 1.1 del Capítulo 1 que hemos decidido incluir aquí por completitud del trabajo. Del mismo modo, el Apéndice A.2 está dedicado a las demostraciones de algunos resultados que podemos encontrar en la Sección 1.2 del Capítulo 1.

[2]D. E. Knuth, T. Larrabee, P. M. Roberts, *Mathematical Writing*, MAA, 1989

CAPíTULO 1

# Resultados Básicos de los Espacios de Hilbert

## Índice



A lo largo de este capítulo, más concretamente en la Sección 1.1 y en la Sección 1.2, se exponen resultados básicos y algunos ejemplos de la teoría de espacios de Hilbert. Para su estructura se ha seguido el trabajo de [**HN, 01**]. Omitiremos las pruebas de estos resultados básicos que pueden seguirse en el Apéndice A.

### 1.1. Definiciones y Resultados Básicos sobre Espacios de Hilbert

DEFINICIÓN 3. *Sea* $(\mathbb{K}, \sigma)$ *un cuerpo con un automorfismo* $\sigma : \mathbb{K} \longrightarrow \mathbb{K}$. *Sean* $V$ *y* $W$ *dos espacios vectoriales sobre* $\mathbb{K}$. *Un morfismo de grupos* $L : (V, \sigma) \longrightarrow (W, \sigma)$ *se denomina* función semilineal *con respecto al automorfismo* $\sigma$ *si satisface la siguiente propiedad:*

$$\forall \lambda \in \mathbb{K}, \forall v \in V \quad L(\lambda v) = \sigma(\lambda) L(v).$$

En el caso del cuerpo $\mathbb{C}$ de los números complejos con el automorfismo dado por la conjugación $\bar{\cdot} : \mathbb{C} \longrightarrow \mathbb{C}$, dada una aplicación entre los $\mathbb{C}$-espacios vectoriales $L : V \longrightarrow W$, diremos simplemente que $L$ es semilineal cuando $L$ sea semilineal con respecto a la conjugación compleja.

DEFINICIÓN 4. *Sea* $\mathbb{K} = \mathbb{R} \vee \mathbb{C}$, *es decir, o bien el cuerpo de los números reales o el cuerpo de los números complejos. Sea* $V$ *un espacio vectorial sobre* $\mathbb{K}$. *Dada una aplicación*

$$\phi : V \times V \longrightarrow \mathbb{K},$$

*podemos considerar las aplicaciones obtenidas al fijar una de las coordenadas:*

- *Para* $x \in V$, *consideramos*

$$\begin{aligned} \phi(x, \cdot) : V &\longrightarrow \mathbb{K} \\ y &\longmapsto \phi(x, y) \end{aligned}$$

- *Para* $y \in V$, *consideramos*

$$\begin{aligned} \phi(\cdot, y) : V &\longrightarrow \mathbb{K} \\ x &\longmapsto \phi(x, y) \end{aligned}$$

*i) Diremos que* $\phi$ *es una* forma bilineal *si para todo* $x, y \in V$ *las aplicaciones* $\phi(x, \cdot)$ *y* $\phi(\cdot, y)$ *son aplicaciones lineales.*
*ii) En el caso* $\mathbb{K} = \mathbb{C}$ *diremos que* $\phi$ *es una* forma sesquilineal *si para todo* $x, y \in V$, *la aplicación* $\phi(\cdot, y)$ *es lineal y la aplicación* $\phi(x, \cdot)$ *es semilineal, con respecto a la conjugación compleja.*

DEFINICIÓN 5. *Sea* $\mathbb{K}$ *un cuerpo y sea* $H$ *un espacio vectorial sobre* $\mathbb{K}$. *Se denomina* producto interno *(*hermítico *en el caso* $\mathbb{K} = \mathbb{C}$*) a toda aplicación:*

$$\langle \cdot, \cdot \rangle : H \times H \longrightarrow \mathbb{K},$$

*que satisface:*

*i) Si* $\mathbb{K} = \mathbb{R}$, $\langle \cdot, \cdot \rangle$ *es una forma bilineal y si* $\mathbb{K} = \mathbb{C}$, $\langle \cdot, \cdot \rangle$ *es una forma sesquilineal.*
*ii) El producto interno* $\langle \cdot, \cdot \rangle$ *es simétrico, es decir,*

$$\forall \ x, y \in H, \ \langle x, y \rangle = \overline{\langle y, x \rangle}.$$

*Así, en el caso* $\mathbb{K} = \mathbb{R}$, $\langle \cdot, \cdot \rangle$ *será una forma bilineal simétrica.*

*iii)* *El producto interno* $\langle \cdot, \cdot \rangle$ *es semidefinido positivo, es decir,*
$$\forall\ x \in H,\ \langle x, x\rangle \in \mathbb{R} \text{ y } \langle x, x\rangle \geq 0.$$
*iv)* *El producto interno* $\langle \cdot, \cdot \rangle$ *es definido positivo, es decir, además de la propiedad iii) se cumple*
$$\forall\ x \in H \setminus \{0\},\ \langle x, x\rangle > 0.$$

DEFINICIÓN 6. *Un* espacio preHilbert *es un espacio vectorial dotado de un producto interno.*

Una propiedad clásica y esencial de los espacios preHilbert es la *desigualdad de Cauchy-Schwarz.* Publicado originalmente por A. L. Cauchy para sumatorios y por H. A. Schwarz para integrales, el resultado dice lo siguiente:

TEOREMA 1.1.1 (de Cauchy-Schwarz). *Sea* $(H, \langle \cdot, \cdot \rangle)$ *un espacio preHilbert sobre* $\mathbb{K}$*. Entonces, para cada* $x, y \in H$ *se cumple*
$$|\langle x, y\rangle| \leq \sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}.$$

Los productos internos definen una estructura natural de espacio normado sobre los espacios preHilbert, como vemos en el siguiente teorema:

TEOREMA 1.1.2. *Sea* $(H, \langle \cdot, \cdot \rangle)$ *un espacio preHilbert. Definimos la siguiente función*
$$|| \cdot || : H \longrightarrow \mathbb{R}_+$$
$$x \mapsto \sqrt{\langle x, x\rangle}.$$
*Entonces,* $(H, || \cdot ||)$ *es un* $\mathbb{K}$*-espacio vectorial normado. Es decir,* $|| \cdot ||$ *satisface las siguientes propiedades*

*i)* $||x|| \geq 0,\ \forall\ x \in H.$
*ii)* $||x|| = 0 \iff x = 0.$
*iii)* $||\lambda x|| = |\lambda| ||x||,\ \forall\ x \in H,\ \lambda \in \mathbb{K}.$
*iv)* $||x + y|| \leq ||x|| + ||y||,\ \forall\ x, y \in H.$

EJEMPLO 1.1.3. *Consideramos* $(\mathbb{C}^n, \langle \cdot, \cdot \rangle)$ *con el producto hermítico canónico dado por*
$$\langle u, v\rangle = \sum_{i=1}^{n} u_i \overline{v_i},$$
*con* $u = (u_1, ..., u_n), v = (v_1, ..., v_n)$*. Dada una matriz* $A$ *diremos que es hermítica si se cumple que si* $A = (a_{ij})$*, entonces*
$$A = (a_{ij}) = (\overline{a_{ji}}) = A^*.$$
*Podemos así definir una nueva operación dada por*
$$\langle u, v\rangle_A = uA\overline{v}^T.$$
*No es díficil ver que se cumple que* $A$ *es hermítica si y solo si* $\langle \cdot, \cdot \rangle_A$ *cumple las propiedades i) y ii) de la definición 5. Además, es producto interno si y solo si añadimos la condición de que* $A$ *sea definida positiva. El producto interno canónico viene dado por la matriz identidad.*

EJEMPLO 1.1.4. *Sea* $d \in \mathbb{N}$ *un número entero positivo y* $\{x_0, ..., x_n\}$ *un conjunto de variables algebraicamente independientes sobre el cuerpo* $\mathbb{C}$ *de los complejos. Para* $\mathbb{K} = \mathbb{R} \vee \mathbb{C}$*, consideremos el espacio vectorial de los polinomios homogéneos de grado* $d$ *sobre* $\mathbb{K}$*:*
$$H_d^{\mathbb{K}} := H_d^{\mathbb{K}}(x_0, ..., x_n) = \{f \in \mathbb{K}[x_0, ..., x_n] : f = 0 \text{ ó } f \text{ es homogéneo de grado } d\}.$$
*Es un espacio vectorial cuya base natural viene dada por los monomios de grado* $d$*. Así, dado* $\mu = (\mu_0, ..., \mu_n) \in \mathbb{N}^{n+1}$*, definiremos el grado del exponente* $\mu$ *mediante*
$$|\mu| := \mu_0 + ... + \mu_n.$$
*Consideremos el conjunto de todos los monomios de grado* $d$ *en las variables* $\{x_0, ..., x_n\}$*:*
$$\mathscr{B}_d^{\mathbb{K}} := \{{x_0}^{\mu_0}, ..., {x_n}^{\mu_n} \in H_d^{\mathbb{K}} : \mu_0 + ... + \mu_n = d\}.$$
*Es fácil verificar que* $\mathscr{B}_d^{\mathbb{K}}$ *es una base de* $H_d^{\mathbb{K}}$ *como* $\mathbb{K}$*-espacio vectorial y es sencillo comprobar la siguiente igualdad:*
$$\dim_{\mathbb{K}}(H_d^{\mathbb{K}}) = \#(\mathscr{B}_d^{\mathbb{K}}) = \binom{n+d}{n}.$$

*En ocasiones, conviene considerar la clase de polinomios con grado acotado en las variables* $\{x_1, ..., x_n\}$*, que denominaremos polinomios afines. Se define para cada* $d \in \mathbb{N}$ *mediante:*

$$P_d^{\mathbb{K}} := \{f \in \mathbb{K}[x_1, ..., x_n] : deg(f) \leq d\}.$$

*Un proceso clásico permite identificar los elementos de* $P_d^{\mathbb{K}}$ *y los elementos de* $H_d^{\mathbb{K}}$*: son los procesos de homogeneización y afinización de polinomios.*

- *Dado* $f \in H_d^{\mathbb{K}}$*, llamaremos* afinización con respecto a la variable $x_0$ *(o, simplemente, afinización) al polinomio*

$$f^a := f(1, x_1, ..., x_n) \in \mathbb{K}[x_1, ..., x_n].$$

  *Observamos que si* $f \in H_d^{\mathbb{K}}$*, entonces* $f^a \in P_d^{\mathbb{K}}$*.*
- *Dado* $f \in P_d^{\mathbb{K}}$*, llamaremos* homogeneización con respecto a la variable $x_0$ *(o, simplemente, homogeneización) al polinomio*

$$f^h := x_0^{def(f)} f\left(\frac{x_1}{x_0}, ..., \frac{x_n}{x_0}\right) \in \mathbb{K}[x_0, ..., x_n].$$

  *En términos monomiales, podemos definir la homogenización de* $f$ *de la siguiente forma: supongamos* $\delta := deg(f) \leq d$ *y supongamos que* $f$ *tiene la forma*

$$f := \sum_{|\mu| \leq \delta} a_\mu x_1^{\mu_1} ... x_n^{\mu_n}.$$

  *Entonces,* $f^h$ *tiene la forma siguiente:*

$$f^h := \sum_{|\mu| \leq \delta} a_\mu x_0^{\delta - (\mu_1 + ... + \mu_n)} x_1^{\mu_1} ... x_n^{\mu_n}.$$

*Con la simple construcción basada en la homogeneización y la afinización podemos concluir*

$$(f^h)^a = f, \forall f \in P_d^{\mathbb{K}}.$$

*Sin embargo, no siempre se obtendría* $(f^a)^h = f$ *para* $f \in H_d^{\mathbb{K}}$*. La razón es la presencia de posibles factores múltiples de* $x_0$ *entre los factores irreducibles de* $f$*. Por ejemplo, consideramos*

$$f = {x_0}^2 x_1 \in H_3^{\mathbb{K}}.$$

*Entonces,* $f^a = x_1$ *que es directamente homogéneo y, por tanto,* $(f^a)^h = x_1$ *y* $x_1 \neq f$*. Para poder acompasar mejor la relación entre afinización y homogeniezación introducimos el proceso de* homogeneización con grado fijado*:*

$$(f^h)^{(d)} := x_0^{d - deg(f)} f^h.$$

*Entonces podemos concluir que el siguiente es un isomorfismo de espacios vectoriales:*

$$\begin{aligned} \cdot^a : H_d^{\mathbb{K}} &\longrightarrow P_d^{\mathbb{K}} \\ f &\longmapsto f^a \end{aligned}$$

*y que la inversa es la homogeneización con grado d fijado:*

$$\begin{aligned} P_d^{\mathbb{K}} &\longrightarrow H_d^{\mathbb{K}} \\ f &\longmapsto (f^h)^{(d)} \end{aligned}$$

*Concluimos así que* $P_d^{\mathbb{K}}$ *es también un* $\mathbb{K}$*-espacio vectorial de la misma dimensión de* $H_d^{\mathbb{K}}$*, cuya base es dada por la familia*

$${}^a\mathscr{B}_d^{\mathbb{K}} := \{f^a : f \in \mathscr{B}_d^{\mathbb{K}}\} = \{x_1^{\mu_1} ... x_n^{\mu_n} : \mu_1 + ... + \mu_n \leq d\}.$$

*Vamos a introducir una métrica en* $H_d^{\mathbb{K}}$ *(y, por isomorfismo, en* $P_d^{\mathbb{K}}$*) conocida como* la métrica de Bombieri-Weyl. *En realidad, es una estructura de espacio de Hilbert definida del modo siguiente: supongamos dados dos polinomios homogéneos* $f, g \in H_d^{\mathbb{K}}$ *por sus expansiones monomiales:*

$$f = \sum_{|\mu| = d} a_\mu x_0^{\mu_0} ... x_n^{\mu_n}, \quad g = \sum_{|\mu| = d} b_\mu x_0^{\mu_0} ... x_n^{\mu_n}.$$

*Definiremos el producto de Bombieri-Weyl mediante:*

$$\langle f, g \rangle_d := \sum_{|\mu| = d} \binom{d}{\mu}^{-1} a_\mu \overline{b_\mu},$$

*donde el número combinatorio se define mediante:*

$$\binom{d}{\mu} = \frac{d!}{\mu_0!...\mu_n!}.$$

*Es claramente un producto escalar en el caso* $\mathbb{K} = \mathbb{R}$ *y un producto hermítico en el caso* $\mathbb{K} = \mathbb{C}$. *Dejaremos los detalles al lector. En particular,* $(H_d^{\mathbb{K}}, \langle \cdot, \cdot \rangle_d)$ *es un espacio de Hilbert. Podemos profundizar un poco más esta estructura de espacios de Hilbert. Para ello, comencemos considerando la esfera unidad* $S^{2n+1} \subseteq \mathbb{C}^{n+1}$, *es decir,*

$$S^{2n+1} := \{z = (z_0, ..., z_n) \in \mathbb{C}^{n+1} : ||z||_2^2 = |z_0|^2 + ... + |z_n|^2 = 1\}.$$

*La esfera unidad compleja posee una estructura natural de variedad de Riemann, con una forma de volumen (que denotaremos mediante* $d\nu_S$*), que confiere un volumen de* $S^{2n+1}$ *conocido:*

$$vol[S^{2n+1}] = \nu_S[S^{2n+1}] = \frac{2\pi^{n+1}}{\Gamma(n+1)}.$$

*Por otro lado, dado un polinomio homogéneo* $f \in H_d^{\mathbb{K}}$, *podemos considerar la función que define sobre* $S^{2n+1}$

$$\begin{aligned} f : S^{2n+1} &\longrightarrow \mathbb{C} \\ (z_0, ..., z_n) &\longmapsto f(z_0, ..., z_n). \end{aligned}$$

*Nótese que, al ser homogéneo de grado d, para cualquier otro punto* $u = (u_0, ..., u_n) \in \mathbb{C}^{n+1} \setminus \{0\}$, *tenemos que el valor* $f(u)$ *está determinado por* $||u||$ *y el valor de* $f$ *sobre* $S^{2n+1}$, *es decir,*

$$f(u) = ||u||^d f\left(\frac{u}{||u||}\right).$$

*Más aún, podemos considerar la estructura de espacio de Hilbert definida mediante* $L^2(S^{2n+1})$ *y considerar el producto hermítico definido para cada* $g_1, g_2 \in L^2(S^{2n+1})$ *a valores complejos, normalizado con* $vol[S^{2n+1}]$.

$$\langle g_1, g_2 \rangle_{L^2} = \frac{1}{\nu_S[S^{2n+1}]} \int_{S^{2n+1}} g_1(z)\overline{g_2(z)} d\nu_S z.$$

*Pues bien, como observa* [**De, 06**], *la norma de Bombieri-Weyl es una normalización adicional de la norma de* $L^2(S^{2n+1})$, *es decir,*

TEOREMA 1.1.5. *Con las notaciones precedentes, dados* $f, g \in H_d^{\mathbb{C}}$, *se tiene*

$$\langle f, g \rangle_d := \binom{d+n}{n} \langle f, g \rangle_{L^2} = \binom{d+n}{n} \frac{1}{\nu_S[S^{2n+1}]} \int_{S^{2n+1}} f(z)\overline{g(z)} d\nu_S(z).$$

*Omitiremos la demostración porque se sale de los objetivos de este trabajo y dejamos la referencia para el lector (cf.* [**De, 06**]*). Una de las consecuencias más interesantes de está igualdad es la invarianza de la métrica de Bombieri-Weyl por la acción del grupo unitario (o del grupo ortogonal en el caso* $\mathbb{K} = \mathbb{R}$*). Discutamos el caso* $\mathbb{K} = \mathbb{C}$. *Recordemos que el grupo unitario* $U(n+1)$ *es el grupo de matrices dado por la siguiente igualdad:*

$$U(n+1) := \{U \in \mathscr{M}_{n+1}(\mathbb{C}) : U^* = U^{-1}\},$$

*donde* $U^*$ *es la traspuesta conjugada de* $U$. *A través del grupo unitario podemos definir una acción de grupo sobre* $H_d^{\mathbb{C}}$ *del modo siguiente:*

$$\begin{aligned} U(n+1) \times H_d^{\mathbb{C}} &\longrightarrow H_d^{\mathbb{C}} \\ (U, f) &\longmapsto f \circ U^* \end{aligned} \tag{1.1.1}$$

*Por otro lado, el grupo unitario es un grupo de isometrías de la esfera* $S^{2n+1}$ *con el producto inducido por el producto hermítico canónico de* $\mathbb{C}^{n+1}$ :

$$\begin{aligned} U(n+1) \times S^{2n+1} &\longrightarrow S^{2n+1} \\ (U, u) &\longmapsto Uu. \end{aligned}$$

*Claramente, si* $\langle \cdot, \cdot \rangle_2$ *es el producto hermítico canónico en* $\mathbb{C}^{n+1}$ *se tiene que* $\forall x, y \in \mathbb{C}^{n+1}$.

$$\langle Ux, Uy \rangle_2 = \langle x, y \rangle_2.$$

*La idea de Bombieri-Weyl y de la acción de grupo descrita en 1.1.1 consiste en transferir la métrica hermítica canónica de $S^{2n+1}$ (ó $\mathbb{C}^{n+1}$) a $H_d^{\mathbb{C}}$ dado que se verifica:*

$$\forall u \in S^{2n+1}, \qquad (f \circ U^*)(Uu) = f(u).$$

*El resultado de Dedieu antes citado nos muestra que la métrica de Bombieri-Weyl es invariante por la acción del grupo unitario.*

Corolario 1.1.6. *El producto hermítico de Bombieri-Weyl es invariante por la acción del grupo unitario en $H_d^{\mathbb{C}}$, es decir, se verifica:*

$$\langle f, g\rangle_d = \langle f \circ U^*, g \circ U^*\rangle_d, \qquad \forall f, g \in H_d^{\mathbb{C}}.$$

*Para probarlo basta con observar que $U$ define una isometría sobre $S^{2n+1}$. Por tanto, preserva la integral y se tiene:*

$$\int_{S^{2n+1}} (f \circ U^*)(Uz)\overline{(g \circ U^*)(Uz)}d\nu_S(z) = \int_{S^{2n+1}} (f \circ U^*)(z')\overline{(g \circ U^*)(z')}d\nu_S(z'),$$

*por tanto, usando las normalizaciones oportunas, se concluye*

$$\langle f, g\rangle_d = \binom{d+n}{n} \frac{1}{\nu_S[S^{2n+1}]} \int_{S^{2n+1}} (f \circ U^*)(z')\overline{(g \circ U^*)(z')}d\nu_S(z').$$

*Esta propiedad de la invarianza tendrá su utilidad cuando se trate la condición de $H_d^{\mathbb{K}}$ como espacio de Hilbert con núcleo reproductor. Una forma alternativa, puramente calculatoria, de demostrar esta invarianza unitaria de la métrica de Bombieri-Weyl puede verse también en* [**BCSS, 98**]. *No sólo los polinomios individuales tienen estructura de espacio de Hilbert, sino también las familias de sistemas de ecuaciones, que interesan especialmente en Geometría Algebraica. Sea $m \in \mathbb{N}$ un entero positivo y sea $(d) = (d_1, ..., d_m)$ una lista de grados. Denotaremos por $\mathscr{H}_{(d)}^{\mathbb{K}}$ a las listas de polinomios homógeneos de grados respectivos determinados por $(d_1, ..., d_m) = (d)$. Es decir,*

$$\mathscr{H}_{(d)}^{\mathbb{K}} := \prod_{i=1}^{m} H_{d_i}^{\mathbb{K}} = \{(f_1, ..., f_m) \in \mathbb{K}[x_0, ..., x_n]^m : f_i \in H_{d_i}^{\mathbb{K}}, 1 \leq i \leq m\}.$$

*De nuevo, $\mathscr{H}_{(d)}^{\mathbb{K}}$ posee una estructura de espacio de Hilbert con la métrica de Bombieri-Weyl inducida en el espacio producto cartesiano del modo natural:*

$$\langle f, g\rangle_{(d)} = \langle f_1, g_1\rangle_{d_1} + ... + \langle f_m, g_m\rangle_{d_m},$$

*donde $f = (f_1, ..., f_m) \in \mathscr{H}_{(d)}^{\mathbb{K}}$ y $g = (g_1, ..., g_m) \in \mathscr{H}_{(d)}^{\mathbb{K}}$. De nuevo, será una forma sesquilineal invariante por la acción del grupo unítario $U(n+1)$.*

$$(U, f) \longmapsto f \circ U^* := (f_1 \circ U^*, ..., f_m \circ U^*).$$

*De modo análogo, se construye $\mathscr{P}_{(d)}^{\mathbb{K}} := \prod_{i=1}^{m} P_{d_i}^{\mathbb{K}}$ y se le dota la estructura de espacio de Hilbert con la métrica de Bombieri-Weyl.*

Se puede demostrar un resultado recíproco al resultado precedente. Los espacios preHilbert se pueden caracterizar como aquellos espacios normados cuyas normas satisfacen una cierta propiedad que generaliza el paso entre una forma cuadrática y su forma polar: *la Ley del Paralelogramo.*

Teorema 1.1.7 (Ley del Paralelogramo). *Sea $\mathbb{K} = \mathbb{R} \vee \mathbb{C}$ y sea $(V, ||\cdot||)$ un $\mathbb{K}$-espacio vectorial normado. Entonces, la norma $||\cdot||$ sobre $V$ es la inducida por un producto interno $\langle\cdot,\cdot\rangle$ sobre $V$ y una estructura de espacio preHilbert $(V, \langle\cdot,\cdot\rangle)$ si y solamente si la norma $||\cdot||$ satisface la siguiente propiedad (conocida como* Ley del Paralelogramo*):*

$$\forall x, y \in V \qquad ||x+y||^2 + ||x-y||^2 = 2||x||^2 + 2||y||^2.$$

Ejemplo 1.1.8. *Definimos el espacio $\ell^2$ como el espacio de sucesiones en $\mathbb{C}$ cuya serie de cuadrados converge, es decir*

$$\ell^2 = \{(z_n)_n : \sum_{n=1}^{\infty} |z_n|^2 < \infty, z_n \in \mathbb{C}\}.$$

*No es díficil comprobar que es un espacio vectorial con las operaciones*

- $\lambda(z_n)_n = (\lambda z_n)_n$,
- $(z_n)_n + (w_n)_m = (z_n + w_n)_n$.

*Están bien definidas: está claro que*

$$\sum_{n=1}^{\infty} |\lambda z_n|^2 = |\lambda|^2 \sum_{n=1}^{\infty} |z_n|^2 < \infty.$$

*Para la suma, si vemos $\mathbb{C}$ como un espacio preHilbert tenemos por la ley del paralelogramo que*

$$|z_n + w_n|^2 + |z_n - w_n|^2 = 2|z_n|^2 + 2|w_n|^2,$$

*con lo que deducimos que*

$$|z_n + w_n|^2 \leq 2|z_n|^2 + 2|w_n|^2,$$

*y por tanto*

$$\sum_{n=1}^{\infty} |z_n + w_n|^2 \leq 2\sum_{n=1}^{\infty} |z_n|^2 + 2\sum_{n=1}^{\infty} |w_n|^2 < \infty.$$

*Podemos definir la aplicación*

$$\langle (z_n)_n, (w_n)_n \rangle := \sum_{n=1}^{\infty} z_n \overline{w_n}.$$

*Está claro que cumple las propiedades del producto interno, por lo que simplemente vamos a ver que la aplicación está bien definida: Notemos antes que dados dos números reales $r$ y $s$ se tiene que*

$$0 \leq (r-s)^2 = r^2 + s^2 - 2rs,$$

*es decir*

$$rs \leq \frac{1}{2}(r^2 + s^2).$$

*Por tanto, para cada $n \in \mathbb{N}$*

$$|z_n \overline{w_n}| = |z_n||w_n| \leq \frac{1}{2}(|z_n|^2 + |w_n|^2)$$

*por lo que concluimos que*

$$\sum_{n=1}^{\infty} |z_n \overline{w_n}| \leq \sum_{n=1}^{\infty} \frac{1}{2}(|z_n|^2 + |w_n|^2) \leq \sum_{n=1}^{\infty} |z_n|^2 + \sum_{n=1}^{\infty} |w_n|^2 < \infty.$$

*Es decir, la serie es absolutamente convergente y por tanto convergente. Por ello, $\ell^2$ dotado con el producto interno que acabamos de definir es un espacio preHilbert.*

Con las notaciones precedentes, si $(H, \langle \cdot, \cdot \rangle)$ es un espacio preHilbert, tenemos una métrica definida por la norma $||\cdot||$. Así, dados $x, y \in H$, la distancia viene dada por $d(x,y) := ||x - y||$. Por tanto, tenemos una topología inducida en $H$ por esa métrica. Se tiene entonces

Teorema 1.1.9. *Si $(H, \langle \cdot, \cdot \rangle)$ es un espacio preHilbert, el producto interno es una función continua con respecto a la topología producto en $H \times H$ inducida por la topología del espacio normado $(H, ||\cdot||)$.*

Introducimos ahora un concepto básico tanto en las matemáticas como en la fśicia y esencial para este trabajo. Nombrado tras el matemático David Hilbert, los espacios de Hilbert generalizan la noción de espacio euclídeo, permitiendo extender conceptos del análisis y del álgebra a espacios de dimensión finita o infinita. Podemos definirlos formalmente de la siguiente forma:

Definición 7. *Un espacio preHilbert $(H, \langle \cdot, \cdot \rangle)$ se dice que es un* espacio de Hilbert *si es completo respecto a la métrica sobre $H$ inducida por la norma $||\cdot||$. Es decir, si toda sucesión de Cauchy con respecto a $||\cdot||$ es convergente en $H$.*

Definición 8. *Decimos que un espacio vectorial es un* espacio de Banach *si es un espacio normado y completo con respecto a la métrica definida por su norma.*

Corolario 1.1.10. *Un espacio de Banach $(V, ||\cdot||)$ es un espacio de Hilbert con norma $||\cdot||$ dado por un producto interno $\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{K}$ si y solamente si $||\cdot||$ satisface la Ley del Paralelogramo (Teorema 1.1.7).*

PROPOSICIÓN 1.1.11. *El espacio preHilbet $(\ell^2, \langle\cdot,\cdot\rangle)$ introducido en el ejemplo 1.1.8 es completo y, por tanto, un espacio de Hilbert.*

EJEMPLO 1.1.12. *Sea $(X,\nu)$ un espacio de probabilidad y consideramos el conjunto $L^2(X)$ de las funciones con valores complejos con cuadrado integrable, es decir*

$$L^2(X) := \{f : X \longrightarrow \mathbb{C} : \int_X |f(z)|^2 d\nu(z) < \infty\}.$$

*Este espacio es un espacio de Hilbert con el producto hermítico*

$$\langle f, g\rangle = \int_X f(z)\overline{g(z)} d\nu(z),$$

*siempre y cuando consideremos la relación de equivalencia $f \equiv g$ si y solo si el conjunto de puntos $x \in X$ donde $f(x) \neq g(x)$ tiene medida 0. Esta última propiedad es la que nos garantiza que $\langle f, f\rangle = 0 \iff f \equiv 0$. El resto de propiedades son sencillas, por lo que no las escribiremos. Así, sabemos que $L^2(X)$ es un espacio preHilbert. Vamos a ver que es completo. Sea $f^{(k)}$ una sucesión de Cauchy en $L^2(X)$. Entonces para todo $\varepsilon > 0$ existe un $\delta$ tal que para todo $k, l > \delta$*

$$||f^{(k)} - f^{(l)}|| < \varepsilon.$$

*Por ello, existe un $k_1$ tal que para todo $l > k_1$*

$$||f^{(k_1)} - f^{(l)}|| < \frac{1}{2}.$$

*Del mismo modo, existe $k_2 > k_1$ tal que para todo $l > k_2$*

$$||f^{(k_2)} - f^{(l)}|| < \frac{1}{2^2}.$$

*Podemos así encontrar una sucesión $k_1, k_2, k_3, ...$ tal que para todo $j = 1, 2, ...$*

$$||f^{(k_j)} - f^{(k_{j+1})}|| < \frac{1}{2^j}.$$

*Si consideramos la sucesión $f^{(k_j)}$ podemos escribir*

$$f^{(k_j)} = f^{(k_1)} + [f^{(k_2)} - f^{(k_1)}] + [f^{(k_3)} - f^{(k_2)}] + ... + [f^{(k_j)} - f^{(k_{j-1})}].$$

*Si fijamos*

$$F^{(j)} = |f^{(k_1)}| + |f^{(k_2)} - f^{(k_1)}| + |f^{(k_3)} - f^{(k_2)}| + ... + |f^{(k_j)} - f^{(k_{j-1})}|,$$

*tenemos que*

$$|f^{(k_j)}| \leq F^{(j)}.$$

*La sucesión $F^{(j)}$ es una sucesión monótona creciente y por ello converge a una función $F$. Esto implica que la sucesión $f^{(k_j)}$ converge a una función $f$ ya que las sumas parciales son absolutamente convergentes. Es más,*

$$||F^{(j)}|| \leq ||f^{(k_1)}|| + ||f^{(k_2)} - f^{(k_1)}|| + ||f^{(k_3)} - f^{(k_2)}|| + ... + ||f^{(k_j)} - f^{(k_{j-1})}||,$$

*que está acotada por*

$$||f^{(k_1)}|| + \frac{1}{2} + \frac{1}{2^2} + ... + \frac{1}{2^j} < ||f^{(k_1)}|| + 1.$$

*Por el Teorema de Convergencia Monótona existe un $F$ integrable tal que*

$$|f^{(k_j)} - f(x)| \leq |f^{(k_j)}| + |f| \leq 2F.$$

*Como $f^{(k_j)}$ converge a $f$, por el Teorema de la Convergencia Dominada tenemos que*

$$\lim_{j\to\infty} \int_X |f^{(k_j)}(x) - f(x)|^2 d\nu = 0.$$

*Es decir, tenemos que para la subsucesión $k_j$ se cumple*

$$\lim_{j\to\infty} ||f^{k_j} - f|| = 0.$$

*Si fijamos $\varepsilon > 0$ y consideramos un $\delta$ tal que si $k_j > \delta$, $||f^{k_j} - f|| < \frac{\varepsilon}{2}$ y para $l > \delta$, $||f^{(k_j)} - f^{(l)}|| < \frac{\varepsilon}{2}$. Entonces, para todo $l > \delta$*

$$||f^{(l)} - f|| \leq ||f^{(k_j)} - f^{(l)}|| + ||f^{(k_j)} - f|| < \varepsilon.$$

*Concluimos así que $L^2(X)$ es un espacio de Hilbert. Si consideramos $X = \mathbb{Z}_+$, es decir, el conjunto de los números enteros positivos, tenemos que $L^2(X) = \ell^2$, siendo $\ell^2$ el conjunto visto en el ejemplo 1.1.8.*

El siguiente resultado clásico permite observar que el paso al completado de un espacio pre-Hilbert $H_0$ permite construir un único (salvo isometría) espacio de Hilbert $H$ en el que $H_0$ es denso. Recordamos que como el resto de demostraciones, esta se encuentra en la Sección A.1 del Apéndice A,

TEOREMA 1.1.13. *Sea $(H, \langle\cdot,\cdot\rangle)$ un espacio preHilbert sobre $\mathbb{K}$. Entonces, existe un espacio de Hilbert $(H_0, \langle\cdot,\cdot\rangle_{H_0})$ tal que existe una aplicación lineal continua $I : H \longrightarrow H_0$ verificando:*

*i) $I(H)$ es denso en $H_0$ para la topología inducida por la norma $||\cdot||_{H_0}$.*

*ii) $I$ es una isometría, es decir,*

$$\langle I(x), I(y)\rangle_{H_0} = \langle x, y\rangle_H,\ \forall\ x, y \in H.$$

*Más aún, $H_0$ es único con estas propiedades salvo isomorfismo lineal isométrico. Es decir, dados $(H', \langle\cdot,\cdot\rangle_{H'})$ un espacio de Hilbert y una aplicación lineal $L : H \longrightarrow H'$ que satisface las propiedades i) y ii), entonces existe una isometría lineal $\varphi : H_0 \longrightarrow H'$, que además cumple que $\varphi\,|_{I(H)}: I(H) \longrightarrow J(H)$ es isomorfismo.*

DEFINICIÓN 9. *Bajo las notaciones del teorema anterior, llamamos* completación *del espacio preHilbert $H$ al espacio de Hilbert $H_0$.*

DEFINICIÓN 10. *Sea $(V, ||\cdot||)$ un espacio normado sobre $\mathbb{K}$ y sea $L \in Hom_{\mathbb{K}}(V, \mathbb{K})$ una aplicación lineal $L : V \longrightarrow \mathbb{K}$.*

*i) Decimos que $L$ es una* aplicación lineal de norma acotada *si existe $M \in \mathbb{R}$, $M \geq 0$ tal que*

$$|L(x)| \leq M||x|| \qquad \forall x \in V.$$

*ii) Decimos que $L$ es una* aplicación lineal continua *si es una función continua considerando en $V$ la topología inducida por $||\cdot||$ y en $\mathbb{K}$ la topología usual inducida por su valor absoluto $|\cdot|$.*

El siguiente resultado es bien conocido:

TEOREMA 1.1.14. *Sea $(V, ||\cdot||)$ un espacio normado sobre $\mathbb{K}$ y $L \in Hom_{\mathbb{K}}(V, \mathbb{K})$ una función lineal. Entonces, $L$ es continua si y solamente si es de norma acotada.*

Denotaremos por $V^* := \{L \in Hom_{\mathbb{K}}(V, \mathbb{K}) : L \text{ es continua}\}$ al $\mathbb{K}$-espacio vectorial formado por todas las aplicaciones lineales continuas definidas en $V$. Más aún, para cada $L \in V^*$, denotaremos su norma mediante

$$||L|| := \sup_{||x||=1} \{||L(x)||.$$

## 1.2. Ortogonalidad. Bases Ortonormales de un Espacio de Hilbert

DEFINICIÓN 11. *Sean $x, y$ dos vectores de un espacio de Hilbert $H$. Decimos que $x$ e $y$ son* ortogonales, *representado por $x \perp y$, si $\langle x, y\rangle = 0$. Decimos que dos subconjuntos $A, B$ son ortogonales si $x \perp y$ para todo $x \in A$ e $y \in B$. El* complementario ortogonal *$A^\perp$ de un subconjunto $A$ es el conjunto de vectores ortogonales a $A$,*

$$A^\perp := \{x \in H : x \perp y\ \forall y \in A\}.$$

A continuación enunciaremos algunas propiedades elementales relativas a la ortogonalidad en espacios de Hilbert. Las pruebas se han incluido en la sección A.2 del apéndice con el fin de hacer este trabajo lo más autocontenido posible.

TEOREMA 1.2.1. *El complementario ortogonal de un subconjunto de un espacio de Hilbert es un subespacio lineal cerrado.*

El siguiente resultado expresa una propiedad geométrica fundamental de los espacios de Hilbert.

TEOREMA 1.2.2. *Sea $\mathcal{M}$ un subespacio lineal cerrado de un espacio de Hilbert $H$.*

*a) Para cada $x \in H$ existe un único punto $y \in \mathcal{M}$ tal que*

$$||x - y|| = \min_{z\in\mathcal{M}} ||x - z||.$$

*b) El punto $y \in \mathcal{M}$ que cumple a) es el único elemento de $\mathcal{M}$ con la propiedad de que $(x - y) \perp \mathcal{M}$.*

Definición 12. *Sean $\mathcal{M}$ y $\mathcal{N}$ dos subespacios lineales, cerrados y ortogonales de un espacio de Hilbert. Definimos la* suma directa *de ambos, representado $\mathcal{M} \oplus \mathcal{N}$ como el conjunto*

$$\mathcal{M} \oplus \mathcal{N} := \{x + y : x \in \mathcal{M}, y \in \mathcal{N}\}.$$

El Teorema 1.2.2 nos dice que si $\mathcal{M}$ es un subespacio cerrado, cualquier $x \in H$, siendo $H$ un espacio de Hilbert, puede ser escrito como $x = y + z$, con $y \in \mathcal{M}$ es el punto más próximo a $x$ y $z \perp \mathcal{M}$. Por ello tenemos el siguiente corolario:

Corolario 1.2.3. *Si $\mathcal{M}$ es un subespacio cerrado de un espacio de Hilbert $H$, entonces*

$$H = \mathcal{M} \oplus \mathcal{M}^{\perp}.$$

Recordemos que un subconjunto no nulo $U$ de vectores de un espacio de Hilbert $H$ es *ortogonal* si dos elementos cualesquiera distintos son ortogonales. Un conjunto de vectores se dice *ortonormal* si es ortogonal y $||u|| = 1$ para todo $u \in U$.

Sea $(V, || \cdot ||)$ un espacio de Banach, $I$ un conjunto cualquiera y $X := \{x_i : i \in I\} \subseteq V$ un subconjunto cualquiera de elementos de $V$ indexados por $I$. Para cada $J \subseteq I$ finito denotaremos las sumas parciales indexadas por $I$ mediante

$$S_J(\{x\}) = \sum_{j \in J} x_j.$$

Definición 13. *Sea $(V, || \cdot ||)$ un espacio de Banach, $X = \{x_i : i \in I\} \subseteq V$ un subconjunto de elementos de $V$ indexado por $I$. Diremos que la suma $x$ de los elementos de $X$* converge incondicionalmente *si para cada $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, existe un subconjunto finito $J^{\varepsilon} \subseteq I$ tal que para todo otro subconjunto finito $J \subseteq I$ tal que $J^{\varepsilon} \subseteq J$ se tiene*

$$||S_J(X) - x|| < \varepsilon.$$

Obsérvese que si un conjunto $X \subseteq V$ de un espacio de Banach $(V, || \cdot ||)$ converge incondicionalmente a un elemento $x \in V$, este elemento es único. En caso contrario, si $X$ convergiese a $x \in V$ y a $x' \in V$, para cada $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, tendríamos dos conjuntos finitos $J_1^{\varepsilon} \subseteq I$ y $J_2^{\varepsilon} \subseteq I$ tales que para todo $J \subseteq I$ finito se tiene:

- si $J_1^{\varepsilon} \subseteq J$, entonces $||S_J(X) - x|| < \varepsilon$.
- si $J_2^{\varepsilon} \subseteq J$, entonces $||S_J(X) - x|| < \varepsilon$.

Tomando $J^{\varepsilon} := J_1^{\varepsilon} \cup J_2^{\varepsilon} \subseteq I$, que es finito, tendríamos que para todo $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$

$$||x - x'|| \leq ||x - S_J(X)|| + ||x' - S_J(X)|| < 2\varepsilon,$$

con lo que $x = x'$.

Definición 14. *Dado un espacio de Hilbert $H$ diremos que un subconjunto ortonormal $U$ de $H$ es* maximal *si para cualquier otro subconjunto ortonormal $V$, con $U \subseteq V$, se cumple que $U = V$. Por otro lado, diremos que $U$ es una* base ortonormal *si el conjunto generado por los elementos de $U$ es denso en $H$, i.e, todo elemento de $H$ está en $U$ o existe una sucesión en $U$ que converja a él.*

El siguiente resultado da varias formulaciones equivalentes de la noción conjunto ortonormal maximal. La demostración puede encontrarse en la Sección A.2 del Apéndice A.

Teorema 1.2.4. *Si $U = \{u_i : i \in I\}$ es un subconjunto ortonormal de un espacio de Hilbert $H$, entonces las siguientes condiciones son equivalentes:*

*i) $\langle x, u_i \rangle = 0$ para todo $i \in I$ implica $x = 0$;*
*ii) $x = \sum_{i \in I} \langle x, u_i \rangle u_i$ para todo $x \in H$;*
*iii) $||x||^2 = \sum_{i \in I} |\langle x, u_i \rangle|^2$ para todo $x \in H$;*
*iv) $[U] = \{\sum_{u \in U} c_u u : c_u \in \mathbb{C}$ y $\sum_{u \in U} c_u u$ converge incondicionalmente$\} = H$;*
*v) $U$ es un conjunto ortonormal maximal.*

Está claro que si un conjunto ortonormal $U$ cumple una (y por tanto todas) de las propiedades del teorema precedente, entonces $U$ es una base ortonormal. Asumiendo el lema de Zorn, podemos afirmar que todo espacio de Hilbert posee una base ortonormal, esto es, se tiene:

Teorema 1.2.5. *Asumiendo el* lema de Zorn*, todo espacio de Hilbert $H$ posee una base ortonormal.*

## 1.3. Separabilidad en Espacios de Hilbert

Por su interés en nuestro estudio de espacios de Hilbert con núcleo reproductor, formado por funciones continuas sobre un espacio métrico compacto, nos interesan especialmente los espacios de Hilbert separables.

DEFINICIÓN 15. *Un espacio de Hilbert* $(H, \langle\cdot,\cdot\rangle)$ *se dice* separable *si posee un subconjunto denso contable, es decir, finito o numerable.*

EJEMPLO 1.3.1. *El espacio de Hilbert* $(\ell^2, \langle\cdot,\cdot\rangle)$ *del Ejemplo 1.1.8 es un espacio de Hilbert separable. Consideramos* $x = (x_n)_n \in \ell^2$*. Por tanto, tenemos que* $\sum_{n=1}^{\infty} |x_n|^2 < \infty$*. Luego:*

$$\sum_{n=1}^{\infty} |x_n|^2 < \infty \Rightarrow \left(\sum_{n=1}^{\infty} |x_n|^2\right)^{\frac{1}{2}} < \infty.$$

*Por ello, se da la siguiente equivalencia*

$$\left(\sum_{n=1}^{\infty} |x_n|^2\right)^{\frac{1}{2}} < \varepsilon \iff \sum_{n=1}^{\infty} |x_n|^2 < \varepsilon^2,$$

*para algún* $\varepsilon > 0$*. Entonces, va a existir un* $k \in \mathbb{N}$ *tal que*

$$\sum_{n=k+1}^{\infty} |x_n|^2 < \left(\frac{\varepsilon}{2}\right)^2.$$

*Consideramos* $y = (y_n)_n$ *la sucesión dada por*

$$y_n = \begin{cases} x_n, & n \leq k \\ 0, & n > k \end{cases}$$

*entonces*

$$||x - y||_2 = \left(\sum_{n=k+1}^{\infty} |x_n|^2\right)^{\frac{1}{2}} < \frac{\varepsilon}{2}.$$

*Definimos*

$$D_n = \{\sum_{j=1}^{n} q_j e_j : q_j \in \mathbb{Q}[i]\},$$

*con* $e_j$ *el correspondiente elemento de la base canónica* $\{e_n : n \in \mathbb{N}\}$ *de* $\ell^2$ *dada por*

$$(e_n)_m = \begin{cases} 1, & n = m \\ 0, & \text{en otro caso,} \end{cases}$$

*y* $\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}$*. Como* $(\mathbb{Q}[i])^n$ *es numerable,* $D_n$ *también y por ello*

$$D = \cup_{n\in\mathbb{N}} D_n$$

*también es numerable, al ser unión numerable de conjuntos numerables. Si consideramos la clausura* $\overline{D}$ *del conjunto* $D$*, está claro que* $y \in \overline{D}$*, por lo que si consideramos un elemento* $z$ *de* $D$ *tenemos que*

$$||x - z||_2 \leq ||x - y||_2 + ||y - z||_2 < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

*Concluimos que* $D$ *es numerable y denso en* $\ell^2$ *y por tanto este espacio es separable.*

TEOREMA 1.3.2. *Un espacio de Hilbert* $H$ *es separable si y solo si posee una base ortonormal numerable.*

DEMOSTRACIÓN. Recordemos que un espacio topológico es separable si contiene un subconjunto denso numerable. Por tanto, tenemos que probar que un espacio de Hilbert contiene un subconjunto denso numerable si y solo si tiene una base ortonormal numerable. Probemos ambas implicaciones

$\Longrightarrow$) Supongamos que $H$ tiene un subconjunto denso numerable $\{a_j\}$ con $j \in \mathbb{N}$. Sea $\{e_i\}_{i\in I}$ una base ortonormal de $H$, que sabemos de su existencia por el teorema anterior. Supondremos que la base no es numerable. Tenemos que para cualesquiera $e_n \neq e_m$, $n, m \in I$, debido a la ortonormalidad

$$||e_n - e_m||^2 = \langle e_n - e_m, e_n - e_m\rangle = \langle e_n, e_n\rangle + \langle e_n, e_m\rangle + \langle e_m, e_n\rangle + \langle e_m, e_m\rangle = ||e_n|| + ||e_m|| = 2.$$

Por tanto, dos elementos cualesquiera de la base están a distancia $\sqrt{2}$. Para todo $i \in I$ consideramos las bolas

$$B(e_i, \frac{1}{2}),$$

que tienen cada una diámetro menor que 1. Por tanto, cada una de ellas solo puede contener un elemento de la base, el propio centro. Además, las bolas tienen que ser disjuntas, pues de no ser así la desigualdad triangular nos permitiría probar que la distancia entre los centros de las bolas es menor que 1, llegando a contradicción. Como $\{a_j\}$ es un subconjunto denso y cada bola es no vacía, debe haber al menos un elemento del subconjunto en cada bola. Como ya hemos dicho, las bolas son disjuntas, por lo que podemos definir una función sobreyectiva de un subconjunto de $\{a_j\}$ a las bolas. Pero hemos definido las bolas a partir de un conjunto no numerable, por lo que debería haber una cantidad no numerable de $a_j$, llegando a una contradicción. Por tanto, la base ortonormal tiene que ser numerable.

$\Longleftarrow$) Sea $\{e_i\}_{i\in\mathbb{N}}$ una base ortonormal numerable de $H$. Consideramos

$$\mathbb{Q}[i] = \{a + bi \in \mathbb{C} : a, b \in \mathbb{Q}\},$$

que es un subconjunto numerable denso de $\mathbb{C}$. Definimos además, para todo $n \in \mathbb{N}$ los siguientes subconjuntos de $H$:

$$A_n = \{\sum_{j=1}^{n} q_j e_j : q_j \in \mathbb{Q}[i]\},$$

que también son numerables, pues los podemos identificar con $\mathbb{Q}[i]^n$. Consideramos

$$A = \bigcup_{n\in\mathbb{N}} A_n,$$

que es numerable por ser unión numerable de conjuntos numerables. $A$ va a ser el subconjunto que buscamos a faltar de probar que es denso en $H$. La clausura de cada $A_n$ la podemos ver como el $\mathbb{C}$-subespacio generado por $e_1, e_2, ..., e_n$, por lo que la clausura de $A$ contiene cualquier combinación finita de la forma $\sum_{i=1}^{n} c_i e_i$ con $c_i \in \mathbb{C}$. Por el teorema 1.2.4 y la definición de base ortonormal tenemos que para todo $x \in H$, $x = \sum_{i\in\mathbb{N}}\langle x, e_i\rangle e_i$, con $\langle x, e_i\rangle \in \mathbb{C}$, luego podemos expresar $x$ como el lmite de una sucesión de elementos de $A$, por lo que la clausura de $A$ es $H$ y por ello $A$ es denso. Para el caso real, la prueba sería idéntica considerando los $q_j \in \mathbb{Q}$.

□

Observación 1.3.3. Ya hemos visto que $\ell^2$ es separable. Por tanto, con el resultado que acabamos de probar, sabemos de la existencia de una base ortonormal numerable en este espacio. Pero en este caso ya la conocemos, pues se trata de la misma base $\{e_n : n \in \mathbb{N}\}$ que hemos usado en el ejemplo anterior. Con el producto interno definido en el Ejemplo 1.1.8 no es díficil ver que efectivamente la base es ortonormal.

Fijándonos en el ejemplo y en la prueba, no nos sorprende el siguiente resultado. Se trata del papel que juega $\ell^2$ en la caracterización de los espacios de Hilbert separables.

Corolario 1.3.4. *Un espacio de Hilbert $(H, \langle\cdot,\cdot\rangle_H)$ de dimensión numerable no finita es separable si y solamente si es isomorfo isométricamente a $\ell^2$. Es decir, $(H, \langle\cdot,\cdot\rangle_H)$ es separable si y solamente si existe un isomorfismo de espacios vectoriales normados*

$$\Phi : H \longrightarrow \ell^2,$$

*que es isometría, es decir, tal que $\forall f, g \in H$*

$$\langle f, g\rangle_H = \langle\Phi(f), \Phi(g)\rangle_{\ell^2}.$$

Demostración. Una de las implicaciones es obvia dado que si $(H, \langle \cdot, \cdot \rangle_H)$ es isomorfo isométricamente a $(\ell^2, \langle \cdot, \cdot \rangle_{\ell^2})$, entonces posee una base ortonormal numerable, lo que implica que $(H, \langle \cdot, \cdot \rangle_H)$ es separable. Recíprocamente, supongamos que $(H, \langle \cdot, \cdot \rangle_H)$ es un espacio de Hilbert separable. Por el Teorema precedente, posee una base ortonormal numerable. Sea $\mathcal{B} = \{e_n : n \in \mathbb{N}\}$ esa base. Sea $\{l_n : n \in \mathbb{N}\}$ la base ortonormal separable de $(\ell^2, \langle \cdot, \cdot \rangle_{\ell^2})$ descrita en el Ejemplo anterior. Definimos entonces el siguiente isomorfismo continuo $\Phi : H \longrightarrow \ell^2$ tal que $\Phi(e_n) = l_n$, $\forall n \in \mathbb{N}$. Es decir, $\Phi$ viene dado por:

$$H \longrightarrow \ell^2$$
$$\sum_{n \in \mathbb{N}} a_n e_n \longmapsto \sum_{n \in \mathbb{N}} a_n l_n.$$

Es un isomorfismo isométrico dado que si $f = \sum_{n \in \mathbb{N}} a_n e_n, g = \sum_{n \in \mathbb{N}} b_n e_n \in H$

$$\langle f, g \rangle_H = \langle \sum_{n \in \mathbb{N}} a_n e_n, \sum_{n \in \mathbb{N}} b_n e_n \rangle_H = \sum_{n \in \mathbb{N}} a_n \overline{b_n}.$$

$$\langle \Phi(f), \Phi(g) \rangle_{\ell^2} = \langle \sum_{n \in \mathbb{N}} a_n l_n, \sum_{n \in \mathbb{N}} b_n l_n \rangle_{\ell^2} = \sum_{n \in \mathbb{N}} a_n \overline{b_n}.$$

□

CAPíTULO 2

# Espacios de Hilbert con Núcleo Reproductor. Teorema del Representante

## Índice



Tras un capítulo de resultados preliminares y básicos, comunmente conocidos, en este capítulo desarrollamos la teoría básica de espacios de Hilbert con núcleo reproductor y el Teorema del Representante. La idea de núcleo reproductor se remonta a un trabajo de 1907 de S. Zaremba. Simultáneamente, J. Mercer estudió ideas similares en el contexto de estudio de las ecuaciones integrales. Las ideas quedan parcialmente estancadas durante 20 años, aunque se produjeron algunos avances en trabajos de G. Szeg, S. Bergman o S. Bochner. El tema fue desarrollado ampliamente por N. Aronszajn en los años 50 (cf. [**Ar, 50**]). El primer resultado fundamental del que nos ocupamos es el Teorema de Aronszajn-Moore sobre la base del Teorema de Riesz-Fréchet (que probamos en la Sección 2.2), damos un teorema que caracteriza los RKHS mediante la propiedad de que la evaluación en un punto es una función lineal continua y acotada. Resumimos formalmente este resultado del modo siguiente: Sea $\mathbb{K} = \mathbb{R} \vee \mathbb{C}$, sea $X$ un espacio topológico y $H \subseteq \mathbb{K}^X$ un espacio de Hilbert cuyos elementos son funciones de $X$ en $\mathbb{K}$. Para cada $x \in X$, consideremos la función evaluación en $x$

$$\begin{aligned} L_x : H &\longrightarrow \mathbb{K} \\ f &\longmapsto f(x). \end{aligned}$$

Un espacio de Hilbert $(H, \langle\cdot,\cdot\rangle)$ formado por funciones de $X$ en $\mathbb{K}$ (ie, $H \subseteq \mathbb{K}^X$) se denomina *espacio de Hilbert con núcleo reproductor* si para cada $x \in X$, la función $L_x$ es lineal, continua y de norma finita. Esta definición no hace referencia a la presencia de núcleos sino a la función evaluación en conjunto. Un *núcleo* sobre $X$ es una función $K : X \times X \longrightarrow \mathbb{K}$ que satisface:

i) $K$ es simétrica si $\mathbb{K} = \mathbb{R}$ o sesquilineal si $\mathbb{K} = \mathbb{C}$.
ii) $K$ es definida positiva, es decir, dada cualquier familia finita $\{x_1, ..., x_n\} \subseteq X$ la matriz de Gram de $K$ en esta familia

$$Gram[K](x_1, ..., x_n) \coloneqq (K(x_i, x_j))_{1 \le i,j \le n}$$

es definida positiva.

El Teorema de Aronszajn-Moore relaciona los espacios de Hilbert con núcleo reproductor con los núcleos de la forma siguiente:

Teorema 2.0.1. *Sea $X$ un espacio topológico, $K : X \times X \longrightarrow \mathbb{K}$ un núcleo sobre $X$. Entonces, existe un único (salvo isometría) espacio de Hilbert con núcleo reproductor $H_K \subseteq \mathbb{K}^X$ de tal modo que:*

*i) Para cada $x \in X$, la función*

$$\begin{aligned} K_x : X &\longrightarrow \mathbb{K} \\ y &\longmapsto K_x(y) \coloneqq K(x, y) \end{aligned}$$

*verifica que* $K_x \in H_K$.

*ii) Para cada* $f \in H_K$ *y cada* $x \in X$*, si* $L_x : H_K \longrightarrow \mathbb{K}$ *es la función de evaluación,*

$$L_x(f) = f(x) = \langle K_x, f \rangle_K,$$

*donde* $\langle \cdot, \cdot \rangle_K$ *es el producto interno de* $H_K$.

Más aún, probaremos que $H_K$ es el completado del espacio preHilbert $H_0$ generado por la familia

$$\{K_x : x \in X\}.$$

Hemos incluido la condición de espacio topológico sobre $X$, aunque no es necesaria para nuestra demostración. Un caso particular de núcleos son los núcleos de Mercer: son aquellos núcleos donde $X$ es un espacio topológico y $K : X \times X \longrightarrow \mathbb{K}$ es una función continua. Los núcleos de Mercer, aunque muy citados cotidianamente, son algo más restrictivos. Así, por ejemplo, en el Teorema 2.3.2 probaremos que si $(X, d)$ es un espacio métrico compacto y $K : X \times X \longrightarrow \mathbb{K}$ es un núcleo de Mercer, entonces el espacio de Hilbert $H_K$ es separable y, por tanto, isométricamente identificable con $\ell^2$. En todo caso, dedicamos la Sección 2.4 del Capítulo a resumir algunos resultados básicos sobre núcleos de Mercer. Hemos elegido la forma resumida para no extender en exceso este trabajo. Los resultados han sido tomados de [**CZ, 07**] y es a él a quien referiremos para las demostraciones. Debe hacerse notar que [**CZ, 07**] establece sus resultados para el caso $(X, d)$ espacio métrico compacto. El resultado más destacable de esta Sección es el Teorema de Mercer (cf. Teorema 2.4.3) en el que se muestra que los espacios de Hilbert con núcleo de Mercer sobre espacios métricos compactos son separables, mediante una descipción explícita de $H_K$ partiendo de una base ortonormal numerable descrita a partir de los valores y vectores propios de un operador integral $L_K$. Tras este breve resumen sobre núcleos de Mercer, dedicamos una sección a estudiar un resultado sobre el Error de Aproximación entre espacios de Banach, que reusaremos en el capítulo siguiente. El capítulo concluye con dos secciones:

- La Sección 2.6 que enuncia y demuestra la versión general del Teorema del Representante conforme [**HSS, 01**]
- La Sección 2.7 que discute un pseudorecíproco del Teorema del Representante, conforme a [**AMP, 09**].

Como ya hemos hablado del Teorema del Representante, centrémonos en lo que se discute en la Sección 2.7. El objetivo aquí es tratar de discutir que propiedades deben verificar las funciones de regularización $\Omega : H_K \longrightarrow \mathbb{R}$ para el caso en que el Teorema del Representante se satisface. Para ello se considera el *problema de interpolación sobre espacios de Hilbert con núcleo reproductor.*

PROBLEMA 1 (de interpolación en espacios de Hilbert con núcleo reproductor). *Sea* $X$ *un conjunto,* $H \subseteq \mathbb{C}^X$ *un espacio de Hilbert con núcleo reproductor* $K : X \times X \longrightarrow \mathbb{C}$*. Llamaremos problema de interpolación sobre* $H$ *con base en* $X$ *al siguiente: Dada una instancia* $z = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{C})^n$ *y dada una función de interpolación* $\Omega : H \longrightarrow \mathbb{R}$*, llamaremos problema de interpolación en* $H$ *con instancia* $z$ *y función de regularización* $\Omega : H \longrightarrow \mathbb{R}$ *al problema de minimizar el siguiente conjunto:*

$$\{\Omega(\omega) : \omega \in H, \omega(x_i) = y_i, 1 \leq i \leq m\}.$$

Se consideran funciones de regularización $\Omega : H \longrightarrow \mathbb{R}$ *admisibles* como aquellas para las que el problema de interpolación es satisfactible (o consistente, ie. admite solución) en un subespacio generado por $\{K_{x_1}, ..., K_{x_n}\}$. Nótese que el problea de interpolación puede ser visto como una instancia particular del problema de minimización inicial: basta con suponer que la función de riesgo $E_n = 0$. En la Sección 2.7 probamos la siguiente caracterización de las funciones de regularización:

TEOREMA 2.0.2 (Caracterización de las regularizaciones de Tikhonov). *Sea* $H$ *un espacio de Hilbert y* $\Omega : H \longrightarrow \mathbb{R}$ *una regularización de Tikhonov continua y diferenciable (en el sentido de la Definición 25). Supongamos que* $dim(H) \geq 2$*. Entonces,* $\Omega$ *es admisible para el problema de interpolación sobre* $H$ *si y solamente si existe una función monótona creciente* $h : \mathbb{R}_+ \longrightarrow \mathbb{R}$ *tal que*

$$\Omega(\omega) = h(||\omega||^2), \forall \omega \in H.$$

## 2.1. Espacios de Hilbert con Núcleo Reproductor

Haremos la discusión sobre espacios de Hilbert complejos, pero es análogo para el caso real. Comencemos con un poco de notación. Sea $X$ un conjunto y $H \subseteq X^{\mathbb{C}}$ un espacio de Hilbert formado por funciones $f : X \longrightarrow \mathbb{C}$. Para cada función $K : X \times X \longrightarrow \mathbb{C}$ y cada elemento $x \in X$, consideramos la función parcial

$$\begin{aligned} K_x : X &\longrightarrow \mathbb{C} \\ y &\longmapsto K_x(y) := K(x, y) \end{aligned}$$

DEFINICIÓN 16. *Con las notaciones precedentes, sea* $\langle \cdot, \cdot \rangle : H \times H \longrightarrow \mathbb{C}$ *el producto interno del espacio de Hilbert* $H \subseteq X^{\mathbb{C}}$ *anterior. Una función* $K : X \times X \longrightarrow \mathbb{C}$ *se denomina* núcleo reproductor *de* $H$ *si satisface las siguientes propiedades:*

*i) Para cada* $x \in X$, $K_x \in H$.

*ii) Para cada* $f : X \longrightarrow \mathbb{C}$ *tal que* $f \in H$ *y para cada* $x \in X$, *se verifica:*

$$f(x) := \langle f, K_x \rangle.$$

La idea subyacente al núcleo reproductor es la idea de considerar el proceso de evaluación en un punto como elemento de un espacio de Hilbert. Así, con las mismas notaciones precedentes, si $H \subseteq X^{\mathbb{C}}$, tenemos el operador lineal sobre $X$ definido sobre $H$ por la evaluación en un punto $x \in X$.

$$\begin{aligned} L_x : H &\longrightarrow \mathbb{C} \\ f &\longmapsto f(x) \end{aligned}$$

DEFINICIÓN 17. *Un espacio de Hilbert* $(H, \langle \cdot, \cdot \rangle)$ *de funciones sobre un conjunto* $X$ *se denomina un* espacio de Hilbert con núcleo reproductor *si para cada* $x \in X$, *las funciones lineales* $L_x : H \longrightarrow \mathbb{C}$ *son operadores lineales acotados, esto es, si para cada* $x \in X$, *existe un número real* $M_x > 0$ *tal que*

$$\forall f \in H, |L_x(f)| = |f(x)| \leq M_x ||f||_H,$$

*donde* $||f||_H = \sqrt{\langle f, f \rangle}$ *es la norma de* $f$ *como elemento de* $H$.

En la literatura se suele referir a este tipo de espacios como *RKHS*, las siglas del término en inglés *Reproducing Kernel Hilbert Space*. Vamos ahora a verificar que las dos presentaciones de espacios de Hilbert con núcleo reproductor de las Definiciones 16 y 17 son, en realidad, la misma. Para ello, recuperaremos un resultado clásico como es el *Teorema de Representación de Riesz-Fréchet*. Así, dado un espacio de Hilbert $(H, \langle \cdot, \cdot \rangle)$ denotaremos por $H^*$ su dual como espacio de Banach, es decir, el espacio de Banach formado por todas las funciones lineales continuas definidas en $H$.

TEOREMA 2.1.1 (de Representación de Riesz-Fréchet). *Sea* $H$ *un espacio de Hilbert. Para todo* $L \in H^*$ *existe un* $f \in H$ *tal que*

$$L(g) := \langle g, f \rangle, \forall g \in H.$$

*Es más, si denotamos por* $L_f$ *al operador determinado por* $f \in H$, *podemos definir la siguiente función*

$$\begin{aligned} L : H &\longrightarrow H^* \\ f &\longmapsto L_f \end{aligned}$$

*con las siguientes propiedades*

*i)* $L$ *es una biyección.*

*ii) Preserva la norma, es decir,* $||f||_H = ||L_f||_{H^*}$.

*iii) Si consideramos el caso* $\mathbb{K} = \mathbb{R}$, $L$ *es lineal.*

*iv) Si consideramos el caso* $\mathbb{K} = \mathbb{C}$, $L$ *es sesquilineal.*

DEMOSTRACIÓN. Comencemos viendo la existencia del $f$ de la primera parte del enunciado. Sea $\mathcal{M} = ker(L)$ y suponemos que la codimensión de $\mathcal{M}$, $\dim(H) - \dim(\mathcal{M})$, es al menos 1. Si fuese 0, entonces $H = \mathcal{M}$ y para todo $g \in H$, $L(g) = 0$ por lo que bastaría tomar $f \equiv 0$.
Como hemos supuesto que la codimensión no es 0, $H \neq \mathcal{M}$ y por ello $0 \neq \mathcal{M}^{\perp}$. Podemos considerar pues un elemento $u \in \mathcal{M}^{\perp}$ de tal forma que $||u|| = 1$, pues basta elegir un elemento

no nulo cualquiera y dividir por su norma. Entonces, por ser $L$ lineal, tenemos que para todo $g \in H$

$$L(uL(g) - hL(u)) = L(uL(g)) - L(gL(u)) = L(g)L(u) - L(u)L(g) = 0,$$

por lo que $uL(g) - gL(u) \in \mathcal{M}$. Así

$$L(g) = L(g)||u|| = L(g)\langle u, u\rangle = \langle L(g)u, u\rangle = \langle L(g)u - L(u)g + L(u)g, u\rangle = \langle L(g)u - L(u)g, u\rangle + \langle L(u)g, u\rangle = \langle L(u)g, u\rangle.$$

La última igualdad viene por pertenecer $uL(g) - gL(u)$ a $\mathcal{M}$ y $u$ a $\mathcal{M}^{\perp}$. Para terminar, como $\langle L(u)g, u\rangle = \langle g, \overline{L(u)}u\rangle$, si tomamos $f = \overline{L(u)}u$ habremos probado la primera parte del resultado. Probemos que $L$ es biyectiva. Vista la primera parte del enunciado y tal y como hemos definido la aplicación, está claro que $L$ es sobreyectiva. Para probar la inyectividad, probaremos que $ker(L) = \{0\}$. Supongamos que existe $f \in H$, $f$ no nula, tal que $L(f) = L_f \equiv 0$. Entonces tendríamos que $L_f(f) = \langle f, f\rangle = 0$, pero por las propiedades del producto interno, entonces $f \equiv 0$, llegando a una contradicción. Por ello, $ker(L) = \{0\}$ y $L$ es inyectiva. Probamos ahora que la norma se preserva. Aplicando la *desigualdad de Cauchy-Schwarz* tenemos que

$$||L_f||_{H^*} = \sup_{\substack{g\in H\\ ||g||_H=1}} |L_f(g)| = \sup_{\substack{g\in H\\ ||g||_H=1}} |\langle g, f\rangle| \leq \sup_{\substack{g\in H\\ ||g||_H=1}} ||g||_H\, ||f||_H = ||f||_H.$$

Por otro lado

$$||f||_H^2 = \langle f, f\rangle = L_f(f),$$

luego

$$||f||_H = \frac{L_f(f)}{||f||_H} = L_f(\frac{f}{||f||_H}) \leq \sup_{\substack{g\in H\\ ||g||_H=1}} |L_f(g)| = ||L_f||_{H^*}.$$

Por último, probemos que $L$ es lineal o sesquilineal según $\mathbb{K}$ sea $\mathbb{R}$ o $\mathbb{C}$ respectivamente. Vamos a hacerlo para el caso complejo, pues el caso real es equivalente teniendo en cuenta que el conjugado de un número real es él mismo:

$$L(f+g) = L_{f+g} = \langle \cdot, f+g\rangle = \langle \cdot, f\rangle + \langle \cdot, g\rangle = L_f + L_g = L(f) + L(g).$$
$$L(\lambda f) = L_{\lambda f} = \langle \cdot, \lambda f\rangle = \overline{\lambda}\langle \cdot, f\rangle = \overline{\lambda}L_f = \overline{\lambda}L(f).$$

□

TEOREMA 2.1.2. *Sea $X$ un conjunto, $(H, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert, $H \subseteq X^{\mathbb{C}}$ formado por funciones de $X$ en $\mathbb{C}$. Son equivalentes:*

*i) Existe $K : X \times X \longrightarrow \mathbb{C}$ que es núcleo reproductor para $H$*

*ii) Para cada $x \in X$ el operador evaluación en $x$, $L_x : H \longrightarrow \mathbb{C}$ es continua de norma acotada.*

DEMOSTRACIÓN. Probemos ambas implicaciones:

$\Longrightarrow$) Si $H$ tiene un núcleo reproductor $K$, para todo $x \in X$ tenemos que para cualquier $f \in H$ se cumple

$$L_x(f) = f(x) = \langle f, K_x\rangle.$$

Aplicando la *desigualdad de Cauchy-Schwarz*

$$|L_x(f)| = |\langle f, K_x\rangle| \leq ||f||\, ||K_x|| = ||f||\sqrt{\langle K_x, K_x\rangle} = ||f||\sqrt{K(x,x)}.$$

Por tanto

$$||L_x|| = \sup_{||f||\neq 0} \frac{|L_x(f)|}{||f||} \leq \sup_{||f||\neq 0} \frac{||f||\sqrt{K(x,x)}}{||f||} = \sqrt{K(x,x)},$$

y entonces $L_x$ está acotado y por tanto es continuo.

$\Longleftarrow$) Si $L_x$ es continuo, entonces podemos aplicar el Teorema de Riesz-Fréchet: existe un $f_x \in H$ tal que para todo $g \in H$

$$L_x(g) = \langle g, f_x\rangle,$$

y como $L_x(g) = g(x)$, tenemos que

$$g(x) = \langle g, f_x\rangle.$$

Por tanto, si consideramos

$$K : X \times X \longrightarrow \mathbb{C}$$
$$(x, y) \mapsto f_x(y),$$

$K$ es el núcleo reproductor que buscamos.

□

## 2.2. El Teorema de Aronszajn-Moore

El resultado aparece en el trabajo de Aronszajn (cf. [**Ar, 50**]), aunque el autor se lo atribuye a E. H. Moore, de ahí el nombre con el que es comúnmente conocido: *Teorema de Aronszajn-Moore.* Podemos ver el resultado como una especie de recíproco de lo contado en la sección anterior. Allí hemos definido los espacios de Hilbert con núcleo reproductor como aquellos que poseen una función $K : X \times X \longrightarrow \mathbb{C}$ que satisface los requerimientos de la Definición 16. Ahora procederemos de modo recíproco. Consideraremos ciertas clases de funciones $K : X \times X \longrightarrow \mathbb{C}$ que satisfacen dos sencillas propiedades y probaremos que entonces existe un único (salvo isometría) espacio de Hilbert $(H_K, \langle \cdot, \cdot \rangle)$ de funciones definidas en $X$ tal que $K$ es el único núcleo reproductor de $H_K$. Para probar el resultado, recuperaremos un poco de la terminología usual sobre funciones sesquilineales, simétricas y definidas positivas. De nuevo, como en la sección precedente, haremos el estudio para el caso complejo, aunque el resultado es igualmente cierto en el caso real.

Definición 18. *Sea $X$ un conjunto. Una función en valores complejos $K : X \times X \longrightarrow \mathbb{C}$ se denomina* simétrica *si para todo par $(x, y) \in X \times X$ se verifica*

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

*En el caso real, una función $K : X \times X \longrightarrow \mathbb{R}$ se dice simétrica si*

$$\forall x, y \in X, K(x, y) = K(y, x).$$

*Dada una función simétrica $K : X \times X \longrightarrow \mathbb{C}$ y un conjunto finito de puntos de $X$:*

$$X' = \{x_1, ..., x_n\},$$

*podemos asociar una matriz hermítica $K[X'] \in \mathcal{M}_n(\mathbb{C})$ dada por:*

$$K[X'] = (K(x_i, x_j))_{1 \leq i,j \leq n}.$$

*La matriz $K[X']$ se denomina la matriz* de Gram *de $K$ con respecto al conjunto finito $X' \subseteq X$.*

Definición 19. *Una matriz simétrica $M \in \mathcal{M}_n(\mathbb{C})$ se denomina* semidefinida positiva *si para todo $z \in \mathbb{C}^n$ se verifica*

$$zMz^* \geq 0,$$

*donde $z^*$ es el traspuesto conjugado del vector fila $z \in \mathbb{C}^n$. La matriz hermítica $M \in \mathcal{M}_n(\mathbb{C})$ se denomina* definida positiva *si es semidefinida positiva y además*

$$\forall z \in \mathbb{C}^n \setminus \{0\}, zMz^* > 0.$$

Como propiedades características de las matrices definidas y semidefinidas positivas, podemos resumir las siguientes:

Proposición 2.2.1. *Sea $M \in \mathcal{M}_n(\mathbb{C})$ una matriz hermítica. Se tiene:*

*i) $M$ es semidefinida positiva si y solamente si todos sus valores propios (que son números reales por ser hermítica) son no negativos. Es decir, si $\lambda \in \mathbb{R}$ es un valor propio de $M$, entonces $\lambda \geq 0$.*

*ii) $M$ es definida positiva si y solamente si todos sus valores propios son números reales estrictamente positivos.*

*iii) $M$ es definida positiva si y solamente si la siguiente aplicación define un producto hermítico sobre $\mathbb{C}^n$ que le confiere estructura de espacio de Hilbert:*

$$\langle \cdot, \cdot \rangle_M : \mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C}$$
$$(x, y) \longmapsto xMy^*$$

Un enunciado idéntico se tiene en el caso real:

PROPOSICIÓN 2.2.2. *Sea $N \in \mathcal{M}_n(\mathbb{R})$ una matriz simétrica real. Se tiene:*

*i)* *$N$ es semidefinida positiva si y solamente si todos sus valores propios son no negativos.*

*ii)* *$N$ es definida positiva si y solament si es semidefinida positiva y $N \in GL(n, \mathbb{R})$. Equivalentemente, $N$ es definida positiva si y solamente si todos sus valores propios son números reales estrictamente positivos.*

*iii)* *$N$ es definida positiva si y solamente si la siguiente aplicación es un producto interno sobre $\mathbb{R}^n$ que induce una estructura de espacio de Hilbert en $\mathbb{R}^n$:*

$$\begin{aligned} \langle \cdot, \cdot \rangle_N : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto xNy^T \end{aligned}$$

Ambas proposiciones son resultados clásicos con pruebas elementales, con lo que omitiremos la demostración. Simplemente comentar sobre el punto *iii*) que puede ser el más llamativo para este trabajo, ya discutimos algo similar en el Ejemplo 1.1.3. La generalización de las condiciones de semidefinida o definida positiva son las siguientes en nuestro contexto:

DEFINICIÓN 20. *Sea $X$ un conjunto y $K : X \times X \longrightarrow \mathbb{K}$ una aplicación simétrica. Diremos que $K$ es* semidefinida positiva *si para cualquier conjunto finito $X' \subseteq X$, la matriz de Gram $K[X']$ es semidefinida positiva.*
*Diremos que $K$ es* definida positiva *si para cualquier conjunto finito $X' \subseteq X$, la matriz de Gram $K[X']$ es definida positiva.*

TEOREMA 2.2.3 (Aronszajn-Moore). *Sea $X$ un conjunto y sea $K : X \times X \longrightarrow \mathbb{K}$ una función simétrica y definida positiva sobre $X$. Entonces existe un único (salvo isometría) espacio de Hilbert con núcleo reproductor $H_K \subseteq \mathbb{K}^X$ tal que $K$ es un núcleo reproductor de $H_K$. Es decir, se cumple*

*i)* *para todo $x \in X$, $K_x \in H_K$,*

*ii)* *para todo $f \in H_K$ y $x \in X$, $f(x) = \langle f, K_x \rangle_{H_K}$.*

DEMOSTRACIÓN. Sea $H_0$ el espacio generado por $\{K_x : x \in X\}$. Consideramos

$$f = \sum_{i=1}^{s} \alpha_i K_{x_i}, \qquad g = \sum_{j=1}^{r} \beta_j K_{x_j},$$

con $\alpha_i, \beta_j \in \mathbb{K}$. Definimos en $H_0$ la función $\langle \cdot, \cdot \rangle_{H_0} : H_0 \longrightarrow \mathbb{K}$, dada por

$$\langle f, g \rangle_{H_0} = \sum_{i=1}^{s} \sum_{j=1}^{r} \alpha_i \overline{\beta_j} K(x_i, x_j),$$

y vamos a probar que efectivamente se trata de un producto interno:

- $\langle f, g \rangle_{H_0} = \overline{\langle g, f \rangle_{H_0}}$

$$\overline{\langle g, f \rangle_{H_0}} = \overline{(\sum_{i=1}^{s} \sum_{j=1}^{r} \beta_j \overline{\alpha_i} K(x_j, x_i))} = \sum_{i=1}^{s} \sum_{j=1}^{r} \overline{\beta_j} \alpha_i \overline{K(x_j, x_i)} = \sum_{i=1}^{s} \sum_{j=1}^{r} \alpha_i \overline{\beta_j} K(x_i, x_j) = \langle f, g \rangle_{H_0}.$$

- $\langle f, g + h \rangle_{H_0} = \langle f, g \rangle_{H_0} + \langle f, h \rangle_{H_0}$
  Sea
  $$h = \sum_{l=1}^{t} \gamma_l K_{x_l},$$
  y vamos a suponer $r \geq t$ ya que el otro caso sería análogo. De esta manera, podemos escribir $g + h$ como
  $$g + h = \sum_{j=1}^{r} (\beta_j + \gamma_j) K_{x_j},$$
  con $\gamma_j = 0$ para todo $j > t$.
  Entonces

$$\langle f, g + h \rangle_{H_0} = \sum_{i=1}^{s} \sum_{j=1}^{r} \alpha_i \overline{(\beta_j + \gamma_j)} K(x_i, x_j) = \sum_{i=1}^{s} \sum_{j=1}^{r} (\alpha_i \overline{\beta_j} + \alpha_i \overline{\gamma_j}) K(x_i, x_j) =$$

$$= \sum_{i=1}^{s}\sum_{j=1}^{r}\alpha_i\overline{\beta_j}K(x_i,x_j) + \sum_{i=1}^{s}\sum_{j=1}^{r}\alpha_i\overline{\gamma_j}K(x_i,x_j) = \langle f,g\rangle_{H_0} + \langle f,h\rangle_{H_0}.$$

- $\langle f,\lambda g\rangle_{H_0} = \overline{\lambda}\langle f,g\rangle_{H_0}$

$$\overline{\langle f,\lambda g\rangle} = \sum_{i=1}^{s}\sum_{j=1}^{r}\alpha_i\overline{\lambda\beta_j}K(x_i,x_j) = \overline{\lambda}\sum_{i=1}^{s}\sum_{j=1}^{r}\alpha_i\overline{\beta_j}K(x_i,x_j) = \overline{\lambda}\langle f,g\rangle.$$

- $\langle f,f\rangle_{H_0} \geq 0$ y $\langle f,f\rangle_{H_0} = 0 \iff f \equiv 0$

$$\langle f,f\rangle_{H_0} = \sum_{i=1}^{s}\sum_{j=1}^{s}\alpha_i\overline{\alpha_j}K(x_i,x_j) \geq 0,$$

  por ser $K$ definida positiva y en concreto, semidefinida positiva. Por la propia definición de definida positiva, la igualdad se da cuando los $\alpha_i$ son nulos. Por tanto, $f$ sería la función nula.

Por tanto, $(H_0, \langle\cdot,\cdot\rangle_{H_0})$ es un espacio preHilbert. Podemos considerar su completación, que será el $H_K$ que buscamos. Al tratarse de una completación de $H_0$, y como $H_0$ es el subespacio generado por los diferentes $K_x$, se va a cumplir que $K_x \in H_K$ para todo $x \in X$, probando así *i*). Nos falta probar el punto *ii*). Consideremos un $f \in H_0$. Al tratarse $H_K$ de la completación de $H_0$, bastará probarlo para funciones en el espacio original:

$$f = \sum_{i=1}^{s}\alpha_i K_{x_i} \Longrightarrow f(x) = \sum_{i=1}^{s}\alpha_i K(x_i,x),$$

entonces

$$\langle f,K_x\rangle_{H_K} = \sum_{i=1}^{s}\sum_{j=1}^{r}\alpha_i\overline{\beta_j}K(x_i,x_j).$$

Como $K_x$ es un generador de $H_0$, va a existir un índice $j_0$ tal que $x_{j_0}$ coincide con $x$ y por tanto $\beta_{j_0} = 1$ y $\beta_j = 0$ para el resto. Por tanto, el sumatorio de antes es igual a

$$\sum_{i=1}^{s}\alpha_i K(x_i,x_{j_0}) = \sum_{i=1}^{s}\alpha_i K(x_i,x) = f(x).$$

Nos falta probar la unicidad. Sea $H$ un espacio de Hilbert que cumple las condiciones del enunciado. Como $K_x \in H$ para todo $x \in X$, está claro que $H_0 \subset H$. También tenemos que

$$\langle K_x,K_y\rangle_H = K_y(x) = K(y,x)$$

$$\langle K_x,K_y\rangle_{H_K} = K_y(x) = K(y,x)$$

luego coinciden. Por la linealidad del producto interno, está claro entonces que para todo $f, g \in H_0$, $\langle f,g\rangle_{H_K} = \langle f,g\rangle_H$. Como tanto $H$ como $H_K$ son completaciones de $H_0$, por la unicidad de este, salvo isometría, $H_K$ es único. $\square$

### 2.3. Núcleos de Mercer. Separabilidad.

Definición 21. *Sea* $(X,d)$ *un espacio métrico. Un* núcleo de Mercer *es una aplicación* $K : X \times X \longrightarrow \mathbb{C}$ *que es simétrica, definida positiva y continua con respecto a la topología producto en* $X \times X$.

El teorema de Aronszajn-Moore tiene una forma suplementaria interesante en el caso de núcleos de Mercer.

TEOREMA 2.3.1. *Sea $(X,d)$ un espacio métrico y sea $K : X\times X \longrightarrow \mathbb{K}$ un núcleo de Mercer. Sea $(H_K, \langle\cdot,\cdot\rangle_K)$ el espacio de Hilbert asociado a este núcleo por el Teorema de Aronszajn-Moore. Definamos la constante asociada al núcleo $K$:*

$$C_K := \sup_{x\in X} \sqrt{|K(x,x)|}.$$

*Supongamos que $C_K < \infty$. Entonces, se tiene:*

*i)* $C_K := \sup_{x,t\in X} \sqrt{|K(x,t)|}$.

*ii)* $H_K \subseteq \mathscr{C}(X)$ *está formado por funciones continuas y la norma de la inclusión* $I_K : H_K \longrightarrow \mathscr{C}(X)$ *satisface:*

$$||I_K|| \leq C_K.$$

*iii) En particular, para cada* $f \in H_K$ *se tendrá*

$$||f||_\infty := \sup\{|f(x)| : x \in X\} \leq C_K||f||_K.$$

DEMOSTRACIÓN. i) Para probar la igualdad (es obvio que $C_K \leq \sup_{x,t\in X} \sqrt{|K(x,t)|}$) basta con observar que, por ser la matriz de Gram en $x$ y $t$ definida positiva:

$$K[\{x,t\}] := \begin{pmatrix} K(x,x) & K(x,t) \\ K(t,x) & K(t,t) \end{pmatrix},$$

entonces $|K(x,t)|^2 < K(x,x)K(t,t)$ y basta tomar supremos.

ii) Consideramos $f \in H_K$ y $x \in X$ y tendremos que por Cauchy-Schwarz

$$|f(x)| = |\langle f, K_x\rangle_K| \leq ||f||_K||K_x||_K = ||f||_K\sqrt{K(x,x)}. \tag{2.3.1}$$

Por otro lado, si $f \in H_K$ y $x, y \in X$ tendremos que, por Cauchy-Schwarz verifica:

$$|f(x) - f(y)| = |\langle f, K_x\rangle - \langle f, K_y\rangle| = |\langle f, K_x - K_y\rangle| \leq ||K_x - K_y||_K||f||_K.$$

Ahora, observamos que

$$||K_x - K_y||_K^2 = \langle K_x - K_y, K_x - K_y\rangle = \langle K_x, K_x\rangle + \langle K_y, K_y\rangle - \langle K_x, K_y\rangle - \langle K_y, K_x\rangle =$$
$$= K(x,x) + K(y,y) - (K(y,x) + K(x,y)).$$

Por ser $K$ continua, dado $x \in X$ y dado $\varepsilon > 0$, existirá $\delta > 0$ tal que si $y \in X$ es tal que $d(x,y) < \delta$, entonces

$$||K_x - K_y||_K^2 \leq |K(x,x) + K(y,y) - (K(y,x) + K(x,y))| \leq \varepsilon.$$

En particular, si $d(x,y) < \delta$ tendremos que

$$|f(x) - f(x)| \leq \varepsilon||f||_K,$$

lo que demuestra la continuidad de $f \in H_K$ en cualquier punto $x \in X$. Si consideramos $I_K : H_K \longrightarrow \mathscr{C}(X)$ la aplicación lineal continua dada por la inclusión, la desigualdad 2.3.1 nos dice que $\forall f \in H_K$

$$||I_K(f)||_\infty = \sup_{x\in X}|f(x)| \leq ||f||_K \sup_{x\in X}\sqrt{K(x,x)} = ||f||_K C_K,$$

con lo que queda probado tanto *ii*) como *iii*).

□

La primera observación que haremos sobre núcleos de Mercer es la condición de separabilidad en el caso en que $(X,d)$ sea compacto, cuya prueba hemos tomado de [**PR, 16**]

TEOREMA 2.3.2. *Sea $(X,d)$ un espacio métrico compacto y sea $K : X \times X \longrightarrow \mathbb{C}$ un núcleo de Mercer definido sobre $X$. Sea $H_K$ el espacio de Hilbert asociado a $K$ por el teorema de Aronszjan-Moore. Entonces, $H_K$ es un espacio de Hilbert separable y, por tanto, posee una base ortonormal numerable.*

DEMOSTRACIÓN. Observemos que $X \times X$ es también un espacio métrico compacto con la métrica

$$d^* : X \times X \longrightarrow \mathbb{R}$$
$$((x,y),(x',y')) \longmapsto max\{|d(x,y)|, |d(x',y')|\}.$$

Como $K : X \times X \longrightarrow \mathbb{C}$ es continua y $X \times X$ es compacto, $K$ es uniformemente continua, por lo que para todo $\varepsilon_n = \frac{1}{n} > 0$ existe $\delta_n > 0$ tal que para todo $(x, y), (x', y') \in X \times X$ se cumple

$$d^*((x, y), (x', y')) < \delta_n \Longrightarrow |K(x, y) - K(x', y')| = \frac{1}{n}.$$

Si consideramos el recubrimiento

$$X \subseteq \bigcup_{x \in X} B(x, \delta_n),$$

debe tener un subrecubrimiento finito al ser $X$ compacto. Entonces, para cada $n \in \mathbb{N}$, existe una colección finita $F_n \coloneqq \{x_1^{(n)}, ..., x_{m(n)}^{(n)}\} \subseteq X$ tal que

$$X \subseteq \bigcup_{x \in F_n} B(x, \delta_n).$$

En particular, para todo $(x, y) \in X \times X$ y para todo $n \in \mathbb{N}$ existe algún $j$, $1 \leq j \leq m(n)$, tal que

$$d^*((x, y), (x_j^{(n)}, y) = max\{d(x, x_j^{(n)}), d(y, y)\} = d(x, x_j^{(n)}) \leq \delta_n,$$

y por ello existe un $j$ tal que

$$|K_x(y) - K_{x_j^{(n)}}(y)| < \varepsilon_n = \frac{1}{n},$$

para todo $x \in X$. Por tanto, tenemos que

$$||K_x - K_{x_j^{(n)}}||_{H_K}^2 = \langle K_x - K_{x_j^{(n)}}, K_x - K_{x_j^{(n)}} \rangle_{H_K} =$$
$$\langle K_x, K_x \rangle_{H_K} - \langle K_{x_j^{(n)}}, K_x \rangle_{H_K} - \langle K_x, K_{x_j^{(n)}} \rangle_{H_K} + \langle K_{x_j^{(n)}}, K_{x_j^{(n)}} \rangle_{H_K},$$

y como $K$ es un núcleo reproductor tenemos que

$$||K_x - K_{x_j^{(n)}}||_{H_K}^2 = K(x, x) + K(x_j^{(n)}, x_j^{(n)}) - K(x, x_j^{(n)}) - K(x_j^{(n)}, x) \leq 2\varepsilon_n = \tfrac{2}{n}.$$

Es decir, para cada $x \in X$ y para cada $n \in \mathbb{N}$ existe un $x_j \in F_n$ tal que

$$||K_x - K_{x_j^{(n)}}||_{H_K} \leq \frac{\sqrt{2}}{\sqrt{n}},$$

y por ello, $\{K_x : x \in X\}$ está en la clausura del conjunto

$$\bigcup_{n \in \mathbb{N}} \{K_x : x \in F_n\},$$

el cual es un conjunto numerable. Como $H_K$ es la clausura del $\mathbb{C}$-espacio vectorial generado por $\{K_x : x \in X\}$, tenemos que

$$H_K \subseteq \overline{\bigcup_{n \in \mathbb{N}} \{K_x : x \in F_n\}},$$

por lo que los conjuntos coinciden y $\bigcup_{n \in \mathbb{N}} \{K_x : x \in F_n\}$ es denso y numerable, probando así que $H_K$ es separable. □

Así, en el caso de espacios métricos compactos, los núcleos de Mercer no sólo definen un espacio de Hilbert formado por funciones continuas sino que, además, ese espacio de Hilbert será separable y, por tanto, o bien tiene dimensión finita o bien es isomorfo a $\ell^2$.

Observación 2.3.3. En el trabajo [**OS, 17**] se discute una caracterización de la condición de separabilidad de los espacios de Hilbert con núcleo reproductor asociados a espacios topológicos metrizables. Su contenido y su resumen de resultados anteriores escapa ampliamente el alcance de este trabajo. En todo caso, se conocen ejemplos de espacios de Hilbert con núcleo reproductor que no son separables, como los descritos por S. Canu, X. Mary y A. Rakotomamonjy en [**CMR, 03**].

El teorema de Aronszajn-Moore (ver Teorema 2.2.3) nos garantiza que dado un núcleo $K : X \times X \longrightarrow \mathbb{K}$ tendremos un espacio de Hilbert con núcleo reproductor asociado. Por su parte, si el núcleo elegido es de Mercer, el Teorema 2.3.1 nos garantiza que ese espacio de Hilbert con núcleo reproductor está formado por funciones continuas. Veremos ahora algunos ejemplos clásicos de espacios de Hilbert con núcleo reproductor.

EJEMPLO 2.3.4. *Consideremos la esfera unidad compleja $S^{2n+1} \subseteq \mathbb{C}^{n+1}$ y consideremos la función:*

$$\begin{aligned} K \coloneqq K_d : S^{2n+1} \times S^{2n+1} &\longrightarrow \mathbb{C} \\ (x,y) &\longmapsto (\langle x,y\rangle_2)^d, \end{aligned}$$

*donde $\langle x,y\rangle_2$ es el producto hermítico canónico en $\mathbb{C}^{n+1}$. Es fácil verificar que $K_d$ es un núcleo de Mercer para dada $d \geq 1$. Consideremos $z \in S^{2n+1}$ un punto cualquiera y consideremos*

$$\begin{aligned} K_z : S^{2n+1} &\longrightarrow \mathbb{C} \\ y &\longmapsto \langle y,z\rangle_d = (\sum_{i=0}^{n} \overline{z_i} y_i)^d, \end{aligned}$$

*donde $z = (z_0, ..., z_n), y = (y_0, ..., y_n)$. Esta claro que $K_z \in H_d^{\mathbb{C}}$ para cada $z \in S^{2n+1}$. Supongamos que $z = e_0 = (1, 0, ..., 0) \in S^{2n+1}$. Sea $f \in H_d^{\mathbb{C}}$ un polinomio homogéneo de la forma*

$$f \coloneqq \sum_{|\mu|=d} a_\mu x_0^{\mu_0} ... x_n^{\mu_n}.$$

*Tendremos que $K_z = K_{e_0} = x_0^d$. Entonces, se tiene para la métrica de Bombieri-Weyl sobre $H_d^{\mathbb{C}}$:*

$$\langle f, K_{e_0}\rangle = \langle f, x_0^d\rangle = a_{(d,...,0)} = f(e_0).$$

*Consideremos $z \in S^{2n+1}$ un punto cualquiera. Como la acción de $U(n+1)$ sobre $S^{2n+1}$ es transitiva, existe $U \in U(n+1)$ tal que $Ue_0 = z$. Observamos que, entonces, para cada $y \in S^{2n+1}$*

$$\langle y, z\rangle_2 = \langle y, Ue_0\rangle = \langle U^* y, e_0\rangle_2,$$

*por ser $U(n+1)$ el grupo de isometrías lineales del producto hermítico canónico sobre $\mathbb{C}^{n+1}$. Por tanto, concluimos que*

$$(\langle y, z\rangle_2)^d = (\langle U^* y, e_0\rangle_2)^d.$$

*En particular,*

$$K_z(y) = K_{e_0}(U^* y) = (K_{e_0} \circ U^*)(y).$$

*Como $K_{e_0}$ y $K_z$ pueden verse como polinomios en $H_d^{\mathbb{C}}$ y tenemos la acción del grupo $U(n+1)$ sobre $H_d^{\mathbb{C}}$ descrita en el Ejemplo 1.1.4 anterior, concluimos que para cada $f \in H_d^{\mathbb{C}}$*

$$\langle f, K_z\rangle_d = \langle f, K_{e_0} \circ U^*\rangle_d = \langle f \circ U, e_0\rangle_d,$$

*donde hemos usado la invarianza de la métrica de Bombieri-Weyl por la acción de $U(n+1)$. En conclusión,*

$$\langle f, K_z\rangle_d = \langle f \circ U, e_0\rangle_d = (f \circ U)(e_0) = f(Ue_0) = f(z) \tag{2.3.2}$$

*Ahora, dado que $K_z \in H_d^{\mathbb{C}}$ para cada $z \in S^{2n+1}$ y se cumple 2.3.2, tenemos que el espacio de Hilbert asociado al núcleo de Mercer $K$ antes definido es un subespacio de Hilbert de $H_d^{\mathbb{C}}$. Quedaría por ver que $H_d^{\mathbb{C}} = H_K$. Para probarlo, consideremos el espacio ortogonal en $H_d^{\mathbb{C}}$ del $\mathbb{C}$-espacio vectorial generado por $\{K_z : z \in S^{2n+1}\}$. Esto es*

$$W = (\mathbb{C}\langle\{K_z : z \in S^{2n+1}\}\rangle)^\perp.$$

*Un polinomio homogéneo $f \in H_d^{\mathbb{C}}$ está en $W$ si y solamente si $f \perp K_z$ para cada $z \in S^{2n+1}$. Esto significa que*

$$0 = \langle f, K_z\rangle_d = f(z).$$

*Luego $f \in W$ si y solamente si $f|_{S^{2n+1}} \equiv 0$. Como $f$ es homogéneo de grado $d$, entonces para cualquier $x = (x_0, ..., x_n) =\in \mathbb{C}^{n+1} \setminus \{0\}$, tendremos*

$$f(x) = ||x||^d f(\frac{x_0}{||x||}, ..., \frac{x_n}{||x||}) = ||x||^d \cdot 0 = 0,$$

*porque $\frac{x}{||x||} \in S^{2n+1}$. Como los polinomios homogéneos se anulan en $0 \in C^{n+1}$, concluiremos que si $f \in W$, entonces $f$ se anula idénticamente en $\mathbb{C}^{n+1}$. Como $\mathbb{C}$ es un cuerpo de cardinal infinito, si un polinomio se anula idénticamente en $\mathbb{C}^{n+1}$, entonces $f$ es el polinomio nulo. Concluimos así que $W = \{0\}$, donde $0 \in H_d^{\mathbb{C}}$ es el polinomio nulo y*

$$H_d^{\mathbb{C}} = W \perp \mathbb{C}\langle\{K_z : z \in S^{2n+1}\}\rangle = \{0\} \perp \mathbb{C}\langle\{K_z : z \in S^{2n+1}\}\rangle,$$

*por lo que* $\{K_z : z \in S^{2n+1}\}$ *es un sistema generador de* $H_d^{\mathbb{C}}$ *y se tiene que* $H_d^{\mathbb{C}} \subseteq H_K$. *Hemos visto así que el espacio de Hilbert con núcleo reproductor asociado a* $K_d$ *es precisamente* $H_d^{\mathbb{C}}$ *con la métrica de Bombieri-Weyl.*

## 2.4. Un Breve Resumen sobre la Teoría de Mercer

En esta sección nos vamos a ocupar simplemente de resumir las principales contribuciones de J. Mercer en su trabajo seminal de 1909 (cf. [**Me, 09**]). Seguiremos una formalización más contemporánea como la que se expone en [**CZ, 07**]. Omitiremos las demostraciones de estos resultados por no hacer demasiado extenso este trabajo. Las demostraciones pueden verse en [**CZ, 07**]. Se trata, por tanto, de un breve resumen de las contribuciones de Mercer. Sea $(X, d)$ un espacio métrico compacto, sea $\nu$ una medida de Borel finita sobre $(X, d)$ y sea $L^2_\nu(X)$ el espacio de Hilbert descrito en el Ejemplo 1.1.12. Sea $K : X \times X \longrightarrow \mathbb{R}$ un núcleo de Mercer y $\mathscr{C}(X)$ el espacio de Banach de las funciones continuas a valores reales definidas sobre $X$. Definamos la correspondencia siguiente:

$$\begin{aligned} L_K : L^2_\nu(X) &\longrightarrow \mathscr{C}(X) \\ f &\longmapsto L_K(f), \end{aligned}$$

donde para cada $x \in X$,

$$(L_K(f))(x) := \int_X K(x,t) f(t) d\nu(t).$$

Definamos también la siguiente cantidad:

$$C_K := \sup\{\sqrt{K(x,x)} : x \in X\} < \infty.$$

Nótese que esta cantidad está bien definida por ser $K$ continua y $(X, d)$ compacto. La correspondencia $L_K$ anterior es una aplicación bien definida como se prueba en la siguiente Proposición.

PROPOSICIÓN 2.4.1. *Con las notaciones precedentes, si* $K$ *es un núcleo de Mercer, la correspondencia* $L_K$ *es una aplicación lineal. Esto es:*

*i)* $\forall f \in L^2_\nu(X), \quad L_K(f) \in \mathscr{C}(X)$.

*ii)* $L_K$ *es aplicación lineal.*

*Además, la norma de* $L_K$ *satisface* $||L_K|| \leq \sqrt{\nu(X)} C_K^2$.

DEMOSTRACIÓN. La única dificultad es ver que si $f \in L^2_\nu(X)$, entonces $L_K(f) \in \mathscr{C}(X)$. Ahora bien, dado $f \in L^2(X)$ y dados $x_1, x_2 \in X$ se tiene:

$$|(L_K(f))(x_1) - (L_K(f))(x_2)| = \left| \int_X (K(x_1,t) - K(x_2,t)) f(t) d\nu(t) \right|.$$

Por Cauchy-Schwarz, tendremos

$$\begin{aligned} |(L_K(f))(x_1) - (L_K(f))(x_2)| &\leq ||K_{x_1} - K_{x_2}||_{L^2_\nu(X)} ||f||_{L^2_\nu(X)} \leq \\ &\leq \sqrt{\nu(X)} \max_{t \in X}\{|K(x_1,t) - K(x_2,t)|\} \; ||f||_{L^2_\nu(X)} \end{aligned}$$

Como $(X, d)$ es compacto y $K$ es continua, entonces $K$ es uniformemente continua. Por tanto, $\forall \varepsilon > 0 \; \exists \delta > 0$ tal que $\forall x_1, x_2 \in X, t \in X$ si $d(x_1, x_2) < \delta$ entonces

$$|K(x_1,t) - K(x_2,t)| < \varepsilon.$$

Por tanto, si $d(x_1, x_2) < \delta$ se tiene

$$|(L_K(f))(x_1) - (L_K(f))(x_2)| \leq \sqrt{\nu(X)} \; ||f||_{L^2_\nu(X)} \; \varepsilon.$$

La acotación de la norma de $L_K$ también se sigue de Cauchy-Schwarz:

$$|(L_K(f))(x)| \leq \sqrt{\nu(X)} \sup_{t \in X} \sqrt{K(x,t)} \; ||f||_{L^2_\nu(X)},$$

luego

$$\sup_{x \in X} |(L_K(f))(x)| = ||L_K(f)||_{\mathscr{C}(X)} \leq \sqrt{\nu(X)} \; C_K^2 \; ||f||_{L^2_\nu(X)},$$

de donde concluimos

$$||L_K|| \leq \sqrt{\nu(X)} \; C_K^2.$$

□

El trabajo de Mercer expone un estudio del operador integral $L_K$ y su interacción con espacios de Hilbert. Así, como consecuencia del Teorema Espectral de operadores autoadjuntos, Mercer probaría, en nuestro lenguaje actual:

TEOREMA 2.4.2. *Con las notaciones precedentes, sea $K : X \times X \longrightarrow \mathbb{R}$ un núcleo de Mercer sobre un espacio métrico compacto $(X, d)$ dotado con una medida de Borel finita $\nu$. Entonces,*

*i) Existe un sistema ortonormal numerable $\{\phi_k : k > 1\} \subseteq L^2_\nu(X)$ de tal modo que cada $\phi_k$ es vector propio de $L_K$.*

*ii) Los valores propios asociados $\{\lambda_k : k > 1\}$ (ie. $L_K(\phi_k) = \lambda_k\phi_k, \forall k$) son números reales y verifican*
- *o bien $\{\lambda_k : k > 1\}$ es un conjunto finito,*
- *o bien $\lim_{k\to\infty} \lambda_k = 0$.*

*iii) Además, con las notaciones precedentes, $\lambda_k \geq 0$, $\forall k \geq 1$ y*

$$||L_K|| = \max_{k\geq 1}\{\lambda_k\} \leq \sqrt{\nu(X)}\, C_K^2.$$

La prueba puede seguirse en [**CZ, 07**], en concreto, en las páginas 57 y 58. El siguiente resultado, conocido como Teorema de Mercer, nos ofrece una versión explícita sobre la separabilidad de $H_K$. Hemos elegido enunciarlo en el caso en que $\nu$ sea una medida finita no degenerada (ie. $\nu(X) < \infty$, y para cada $U \subseteq X$ abierto, $\nu(U) > 0$).

TEOREMA 2.4.3 (de Mercer). *Con las notaciones del teorema precente, consideremos la familia*

$$\mathcal{F} = \{\sqrt{\lambda_k}\phi_k : \lambda_k > 0\}.$$

*Entonces $\mathcal{F}$ es un conjunto de vectores ortonormales de $H_K$. Además, si $\nu$ es no degenerada, la familia $\mathcal{F}$ es una base ortonormal del espacio de Hilbert $H_K$ asociada al núcleo de Mercer $K$. Más aun, si $\nu$ es no degenerada, definimos el operador*

$$\begin{aligned} L_K^{\frac{1}{2}} : L^2_\nu(X) &\longrightarrow H_K \\ \sum_{k\geq 1} a_k\phi_k &\longmapsto \sum_{k\geq 1} a_k\sqrt{\lambda_k}\phi_k. \end{aligned}$$

*Entonces, $H_K = L_K^{\frac{1}{2}}(L^2_\nu(X))$. Además, si $f \in H_K$, $f = L_K^{\frac{1}{2}} g$ con $g \in L^2_\nu(X)$, se tendrá $||f||_K = ||g||_{L^2_\nu(X)}$.*

Dejaremos también la prueba de este resultado para [**CZ, 07**], en las páginas 61 y 62 del Capítulo 4.

## 2.5. Error de Aproximación entre Espacios de Banach

DEFINICIÓN 22. *Sean $(B, ||\cdot||_B)$ y $(H, ||\cdot||_H)$ dos espacios de Banach con $H$ subespacio de $B$. Definimos el* $\mathbb{K}$-funcional *del par $(B, H)$ mediante:*

$$\begin{aligned} \mathbb{K} : B \times (0, \infty) &\longrightarrow \mathbb{R} \\ (a, t) &\longmapsto \mathbb{K}(a, t) := \inf\{||a - b||_B + t||b||_H : b \in H\}, \end{aligned}$$

*donde $||\cdot||_B$ es la norma de $B$ y $||\cdot||_H$ es la norma de $H$.*

Observemos que para cada $a \in B$ fijo, la función $\mathbb{K}(a, \cdot) : (0, \infty) \longrightarrow \mathbb{R}$ es una función continua no decreciente y $\mathbb{K}(a, t) \leq ||a||_B$ $\forall a \in B, \forall t \in (0, \infty)$ (basta con tomar $b = 0$ para tener esta desigualdad). Además, si $H$ es denso en $B$, $\lim_{t\to 0} \mathbb{K}(a, t) = 0$.

DEFINICIÓN 23. *Sean $0 < r < 1$ y dados $(B, ||\cdot||_B)$, $(H, ||\cdot||_H)$ y $\mathbb{K}$ como en la definición precedente. Llamaremos* espacio de interpolación $(B, H)_r$ *al conjunto definido por la igualdad siguiente:*

$$(B, H)_r := \{a \in B : \sup\{\frac{\mathbb{K}(a, t)}{t^r} : t > 0\} < \infty\}.$$

*Para cada $a \in (B, H)_r$ definimos su norma de $r$-aproximación mediante:*

$$||a||_r = \sup\{\frac{\mathbb{K}(a, t)}{t^r} : t > 0\}.$$

*Luego $(B, H)_r := \{a \in B : ||a||_r < \infty\}$ es una definición equivalente.*

TEOREMA 2.5.1 (de Aproximación). *Sean* $(B, ||\cdot||_B)$ *y* $(H, ||\cdot||_H)$ *dos espacios de Banach, con* $H$ *subespacio de* $B$.

*i) Supongamos que existe* $C_0 \in \mathbb{R}$ *una constante tal que para cada* $b \in H$, $||b||_B \leq C_0||b||_H$. *Sea* $0 < r < 1$ *un número real positivo. Si* $a \in (B,H)_r$, *entonces* $\forall R > 0$ *se tiene:*

$$d(a,R) := \inf\{||a-b||^2 : b \in H, ||b||_H \leq R\} \leq ||a||_r^{\frac{2}{1-r}} R^{-\frac{2r}{1-r}}.$$

*ii) Si dada una constante* $C > 0$ *y dado* $a \in B$, $d(a,R) \leq CR^{-\frac{2r}{1-r}}$, $\forall R \in \mathbb{R}$, *entonces,* $a \in (B,H)_r$ *y*

$$||a||_r \leq 2C^{\frac{1-r}{2}}.$$

DEMOSTRACIÓN. Comencemos considerando la función $f : (0,\infty) \longrightarrow \mathbb{R}$ dada mediante

$$f(t) := \frac{\mathbb{K}(a,t)}{t}.$$

Es una función continua porque $\mathbb{K}(a,\cdot)$ lo es y $t > 0$. Además, se tiene

$$\mathbb{K}(a,t) \leq ||a||.$$

$$\inf\{f(t) : t > 0\} = 0.$$

Fijemos $R > 0$. Si $\sup\{f(t) : t > 0\} \geq R$, entonces, para cada $\varepsilon > 0$, $\varepsilon < 1$, existe $t_{R,\varepsilon} \in (0,\infty)$ tal que

$$f(t_{R,\varepsilon}) = \frac{\mathbb{K}(a,t_{R,\varepsilon})}{t_{R,\varepsilon}} = (1-\varepsilon)R.$$

Por la definición del $\mathbb{K}$-funcional asociado a $(B,H)$, existirá $b_\varepsilon \in H$ tal que

$$||a-b_\varepsilon|| + t_{R,\varepsilon}||b_\varepsilon||_H \leq \frac{\mathbb{K}(a,t_{R,\varepsilon})}{1-\varepsilon}.$$

Luego

$$||b_\varepsilon||_H \leq \frac{\mathbb{K}(a,t_{R,\varepsilon})}{(1-\varepsilon)t_{R,\varepsilon}} = R,$$

$$||a-b_\varepsilon|| \leq \frac{\mathbb{K}(a,t_{R,\varepsilon})}{1-\varepsilon}.$$

Por la definición de $||a||_r$ se tiene

$$\frac{\mathbb{K}(a,t_{R,\varepsilon})}{t_{R,\varepsilon}^r} \leq ||a||_r.$$

Por tanto,

$$||a-b_\varepsilon|| \leq \left[\frac{\mathbb{K}(a,t_{R,\varepsilon})}{(1-\varepsilon)t_{R,\varepsilon}}\right]^{\frac{-r}{1-r}} \left[\frac{\mathbb{K}(a,t_{R,\varepsilon})}{(1-\varepsilon)t_{R,\varepsilon}^r}\right]^{\frac{1}{1-r}} \leq$$

$$\leq R^{\frac{-r}{1-r}} \left(\frac{1}{1-\varepsilon}\right)^{\frac{1}{1-r}} ||a||_r^{\frac{1}{1-r}}.$$

Por tanto,

$$d(a,R) \leq \inf_{0<\varepsilon<1}\{||a-b_\varepsilon||^2\} \leq ||a||_r^{\frac{2}{1-r}} R^{\frac{-2r}{1-r}},$$

lo que prueba el apartado $i)$ en este caso. Si, por el contrario, $\sup\{f(t) : t > 0\} < R$, entonces, para todo $\varepsilon > 0$ tal que

$$\varepsilon < 1 - \frac{1}{R}\sup\{f(u) : u > 0\},$$

y todo $t > 0$, existirá $b_{t,\varepsilon} \in H$ tal que

$$||a-b_{t,\varepsilon}|| + t||b_{t,\varepsilon}||_H \leq \frac{\mathbb{K}(a,t)}{1-\varepsilon}.$$

Esto implica, de nuevo, las dos desigualdades siguientes:

$$||b_{t,\varepsilon}||_H \leq \frac{\mathbb{K}(a,t)}{(1-\varepsilon)t} \leq \frac{1}{1-\varepsilon}\sup\{f(u) : u > 0\} < R,$$

$$||a-b_{t,\varepsilon}|| \leq \frac{\mathbb{K}(a,t)}{1-\varepsilon}.$$

De donde concluimos

$$\begin{aligned} d(a,R) &\leq \inf\{||a - b_{t,\varepsilon}||^2 : t > 0\} \leq \\ &\leq \inf\{\left(\frac{\mathbb{K}(a,t)}{1-\varepsilon}\right)^2 : t > 0\} \leq \\ &\leq \left(\inf\{\frac{||a||_r}{t^r} : t > 0\}\frac{1}{1-\varepsilon}\right)^2 = 0 \end{aligned}$$

Lo que prueba $i)$ en este segundo caso. Para la afirmación $ii)$ supongamos que $\forall k > 0$ se tiene

$$d(a,k) \leq CR^{\frac{-2r}{1-r}}.$$

Sea $t > 0$ y sea

$$R_t = \left(\frac{\sqrt{C}}{t}\right)^{(1-r)}.$$

Entonces, para cada $\varepsilon > 0$, podemos encontrar $b_{t,\varepsilon} \in H$ tal que

$$||b_{t,\varepsilon}||_H \leq R_t$$

$$||a - b_{t,\varepsilon}||^2 \leq CR_t^{\frac{-2r}{1-r}}(1+\varepsilon)^2.$$

Se sigue que

$$\mathbb{K}(a,t) \leq ||a - b_{t,\varepsilon}|| + t||b_{t,\varepsilon}||_H \leq \sqrt{C}R_t^{\frac{-r}{1-r}}(1+\varepsilon) + tR_t \leq 2(1+\varepsilon)C^{\frac{1-r}{2}}t^r.$$

Tomando el límite cuando $\varepsilon \longrightarrow 0$ concluimos

$$\mathbb{K}(a,t) \leq 2C^{\frac{1-r}{2}}t^r.$$

Por tanto

$$||a||_r = \sup_{t>0}\{\frac{\mathbb{K}(a,t)}{t^r}\} \leq 2C^{\frac{1-r}{2}} < \infty.$$

□

## 2.6. Teorema del Representante

TEOREMA 2.6.1 (del Representante). *Sea $X$ un conjunto, $K : X \times X \longrightarrow \mathbb{C}$ un núcleo reproductor y $H_K$ el espacio de Hilbert asociado. Sea $n \in \mathbb{N}$ y sea*

$$\mathcal{E}_n : (X \times \mathbb{C}^2)^n \longrightarrow \mathbb{R} \cup \{\infty\}$$

*una función de riesgo. Sea $h : [0,\infty) \longrightarrow \mathbb{R}$ una función de regularización creciente. Consideramos el problema de hallar el mínimo $\omega \in H_K$ del conjunto:*

$$\mathscr{S} = \{\mathcal{E}_n((x_1, y_1, f(x_1)), ..., (x_n, y_n, f(x_n))) + h(||f||) : f \in H_K\}.$$

*Entonces, si existe $\omega \in H_K$ que minimiza $\mathscr{S}$, es decir, si*

$$\mathcal{E}_n((x_1, y_1, \omega(x_1)), ..., (x_n, y_n, \omega(x_n))) + h(||\omega||) = min\ \mathscr{S},$$

*se cumple que existen $\lambda_1, .., \lambda_n \in \mathbb{C}$ y existe*

$$\hat{\omega} = \sum_{i=1}^n \lambda_i K_{x_i},$$

*de tal forma que $\hat{\omega}$ es un mínimo para $\mathscr{S}$.*

DEMOSTRACIÓN. Como $K$ sobre $H_K$ es un núcleo reproductor, se cumple que si definimos

$$\begin{aligned} \phi : X &\longrightarrow \mathbb{R}^X \\ x &\longmapsto K_x = K(x,\cdot) \end{aligned}$$

tenemos que

$$(\phi(x))(y) = K_x(y) = \langle K_x, K_y\rangle_{H_K} = \langle \phi(x), \phi(y)\rangle_{H_K}.$$

Considerando los $x_1, ..., x_n \in X$, cualquier $f \in H_K$ puede descomponerse como la suma de una parte que est en lo generado por los $\phi(x_i)$ y otra parte ortogonal a ello, es decir, podemos escribir

$$f = \sum_{i=1}^{n} \alpha_i \phi(x_i) + v,$$

con $\alpha_i \in \mathbb{R}$ y $v \in H_K$ satisfaciendo

$$\langle v, \phi(x_i)\rangle_{H_K} = 0,$$

para todo $j$ con $1 \leq j \leq m$. Si evaluamos $f$ en un punto $x_j$:

$$f(x_j) = \langle f, K_{x_j}\rangle_{H_K} = \langle f, \phi(x_j)\rangle_{H_K} = \langle \sum_{i=1}^{n} \alpha_i \phi(x_i) + v, \phi(x_j)\rangle_{H_K} = \sum_{i=1}^{n} \alpha_i \langle \phi(x_i), \phi(x_j)\rangle_{H_K},$$

por lo que el primer término de las funciones de $\mathscr{S}$ no depende de $v$. Para el segundo término, como $v$ es ortogonal a $\sum_{i=1}^{n} \alpha_i \phi(x_i)$, se cumple que

$$h(||f||) = h(||\sum_{i=1}^{n} \alpha_i \phi(x_i) + v||) = h\left(\sqrt{\langle \sum_{i=1}^{n} \alpha_i \phi(x_i) + v, \sum_{i=1}^{n} \alpha_i \phi(x_i) + v\rangle_{H_K}}\right) =$$

$$= h\left(\sqrt{\langle \sum_{i=1}^{n} \alpha_i \phi(x_i), \sum_{i=1}^{n} \alpha_i \phi(x_i)\rangle_{H_K} + \langle \sum_{i=1}^{n} \alpha_i \phi(x_i), v\rangle_{H_K} + \langle v, \sum_{i=1}^{n} \alpha_i \phi(x_i)\rangle_{H_K} + \langle v, v\rangle_{H_K}}\right) =$$

$$= h\left(\sqrt{||\sum_{i=1}^{n} \alpha_i \phi(x_i)||^2 + ||v||^2}\right).$$

Como $h$ es creciente

$$h\left(\sqrt{||\sum_{i=1}^{n} \alpha_i \phi(x_i)||^2 + ||v||^2}\right) \geq h\left(\sqrt{||\sum_{i=1}^{n} \alpha_i \phi(x_i)||^2}\right) = h\left(||\sum_{i=1}^{n} \alpha_i \phi(x_i)||\right).$$

Por tanto, la igualdad se alcanza cuando $v = 0$. Así, cualquier solución ha de ser de la forma

$$f = \sum_{i=1}^{n} \alpha_i \phi(x_i) = \sum_{i=1}^{n} \alpha_i K(x_i, \cdot),$$

tal y como queríamos probar. □

## 2.7. Problema de Interpolación y Función de Regularización

En la sección precedente hemos analizado el Teorema del Representante, que reduce ciertos problemas de optimización sobre espacios de Hilbert a subespacios de dimensión finita. Debido a la especial aplicación en Aprendizaje Computacional, el problema de optimización a tratar incluye una función de regularización (conocida como la regularización de Tikhonov (cf. [**AT, 77**])) $\Omega : H \longrightarrow \mathbb{R}$, donde $H$ es un espacio de Hilbert. Nuestra intención ahora es dar una caracterización de las funciones de regularización admisibles para garantizar un Teorema del Representante en dimensión finita. Seguiremos las líneas de trabajo de [**AMP, 09**], haciendo el estudio a través del problema de interpolación en espacios de Hilbert.

PROBLEMA 2 (de interpolación en espacios de Hilbert). *Sea $H$ un espacio de Hilbert y $\Omega : H \longrightarrow \mathbb{R}$ una función de regularización. Dado $z = ((x_1, y_1), ..., (x_n, y_n)) \in (H \times \mathbb{C})^n$ una instancia, llamaremos problema de interpolación en $H$ con instancia $z$ y regularizador $\Omega$ al problema de hallar $\overline{\omega} \in H$ tal que $\Omega(\overline{\omega})$ minimiza el siguiente conjunto*

$$\{\Omega(\omega) : \omega \in H, \langle x_i, \omega\rangle = y_i, 1 \leq i \leq n\}.$$

*Un problema de interpolación sobre $H$ con instancia $z$ y regularizador $\Omega$ se dice* satisfacible *si y solamente si admite solución en $H$.*

Como en [**AMP, 09**] no nos ocuparemos aquí de las condiciones suficientes de satisfacibilidad. Un caso particular del anterior problema de interpolación es el caso para espacios de Hilbert con núcleo reproductor.

PROBLEMA 3 (de interpolación en espacios de Hilbert con núcleo reproductor). *Sea $X$ un conjunto, $H \subseteq \mathbb{C}^X$ un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{C}$. Llamaremos problema de interpolación sobre $H$ con base en $X$ al siguiente: Dada una instancia $z = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{C})^n$ y dada una función de regularización $\Omega : H \longrightarrow \mathbb{R}$, llamaremos problema de interpolación en $H$ con instancia $z$ y función de regularización $\Omega : H \longrightarrow \mathbb{R}$ al problema de minimizar el siguiente conjunto:*

$$\{\Omega(\omega) : \omega \in H, \omega(x_i) = y_i, 1 \leq i \leq m\}.$$

Dado que nuestro problema aquí es caracterizar las regularizaciones de Tikhonov apropiadas para el Teorema del Representante, introducimos la siguienten noción:

DEFINICIÓN 24. *Sea $H$ un espacio de Hilbert y $\Omega : H \longrightarrow \mathbb{R}$ una función de regularización. Diremos que $\Omega$ es* admisible *para el problema de interpolación sobre $H$ si para todo $n \in \mathbb{N}$ y para toda instancia $z = ((x_1, y_1), ..., (x_n, y_n)) \in (H \times \mathbb{C})^n$ tal que el problema de interpolación asociado es satisfacible, se cumple que existe $\hat{\omega} \in Span(x_1, ..., x_n)$ tal que*

$$\Omega(\hat{\omega}) = min\{\Omega(\omega) : \omega \in H, \langle x_i, \omega\rangle = y_i, 1 \leq i \leq n\}. \tag{2.7.1}$$

Es decir, $\Omega : H \longrightarrow \mathbb{R}$ es admisible si y solamente si todo problema de interpolación satisfacible sobre $H$ con instancia $z = ((x_1, y_1), ..., (x_n, y_n)) \in (H \times \mathbb{C})^n$ y regularizador $\Omega$ satisface que existen $\lambda_1, ..., \lambda_n \in \mathbb{C}$ tales que

$$\hat{\omega} = \sum_{i=1}^{n} \lambda_i x_i,$$

es solución de la identidad 2.7.1 anterior.

DEFINICIÓN 25. *Sea $H$ un espacio de Hilbert y $\Omega : H \longrightarrow \mathbb{R}$ una función. Diremos que $\Omega$ es* diferenciable *si posee derivadas direccionales en todos los puntos $\omega \in H$. Es decir, diremos que es diferenciable si para cada $\omega \in H$ existe $\nabla_\omega \Omega \in H$ tal que para cada $z \in H$ se verifica que existe el siguiente límite y se satisface:*

$$\langle \nabla_\omega \Omega, z\rangle = \lim_{t\to 0} \frac{\Omega(\omega + tz) - \Omega(\omega)}{t}.$$

LEMA 2.7.1. *Sea $H$ un espacio de Hilbert y $\Omega : H \longrightarrow \mathbb{R}$ una función de regularización. Entonces, $\Omega$ es admisible para el problema de interpolación sobre $H$ si y solamente si satisface para cualquier $\omega \in H$*

$$\forall p \in H, \qquad \langle \omega, p\rangle = 0 \quad \Rightarrow \quad \Omega(\omega + p) \geq \Omega(\omega). \tag{2.7.2}$$

DEMOSTRACIÓN. Supongamos que $\Omega$ satisface 2.7.2. Consideramos una instancia arbitraria

$$z = ((x_1, y_1), ..., (x_n, y_n)) \in (H \times \mathbb{C})^n.$$

Sea $\mathscr{S}$ el subespacio generado por $\{x_1, ..., x_n\} \subseteq H$ y sea $\pi : H \longrightarrow \mathscr{S}$ la proyección ortogonal asociada. Supongamos que $\omega$ es una solución al problema de interpolación con instancia $z$ y regularización $\Omega$. Sea $\hat{\omega} = \pi(\omega) \in \mathscr{S}$ la proyeción ortogonal de $\omega$. Entonces

$$\omega = \hat{\omega} + p, \text{ con } \hat{\omega} \in \mathscr{S}, p \in \mathscr{S}^\perp.$$

Por la propiedad 2.7.2 tenemos que $\Omega(\omega) \geq \Omega(\hat{\omega})$ y, dado que $\omega$ minimiza $\Omega$ bajo la hipótesis $\langle x_i, \omega\rangle = y_i, 1 \leq i \leq n$, $\omega \in H$, entonces también $\Omega(\hat{\omega})$ satisface esa propiedad pues

- $\Omega(\hat{\omega}) \leq \Omega(\omega)$
- $\langle x_i, \omega\rangle = \langle x_i, \hat{\omega}\rangle + \langle x_i, p\rangle = \langle x_i, \hat{\omega}\rangle$.

Por tanto $\hat{\omega}$ es solución al problema de interpolación y $\Omega$ es admisible. Recíprocamente, sea $\omega \in H$ y consideremos el siguiente problema de interpolación:

$$min\{\Omega(t) : t \in H, \langle \omega, t\rangle = \langle \omega, \omega\rangle\}.$$

El problema es satisfacible, pues podemos tomar $t = \omega$ y por ser $\Omega$ admisible, al tratarse este nuevo problema de un caso particular del anterior, existe una solución en el subespacio generado por $\omega$. De hecho, el propio $\omega$ es solución. Además, dado cualquier $p \in \langle\omega\rangle^\perp$ tendremos que

$$\Omega(p + \omega) \geq \Omega(\omega),$$

dado que $\langle p + \omega, \omega\rangle = \langle \omega, \omega\rangle$. $\square$

En el caso de espacios de Hilbert con núcleo reproductor, el anterior lema se transforma en el siguiente:

LEMA 2.7.2. *Sea $X$ un conjunto, $H \subseteq \mathbb{C}^X$ un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{C}$ y $\Omega : H \longrightarrow \mathbb{R}$ una regularización. Son equivalentes:*

*i) Para cada instancia $z = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{C})^n$, si el siguiente problema de interpolación con instancia $z$ y regularización $\Omega$*

$$min\{\Omega(\omega) : \omega \in H, \omega(x_i) = y_i, 1 \leq i \leq n\}$$

*es satisfacible, entonces existen $\lambda_1, ..., \lambda_n \in \mathbb{C}$ tales que*

$$\hat{\omega} = \sum_{i=1}^{n} \lambda_i K_{x_i}$$

*satisface*

$$\Omega(\hat{\omega}) = min\{\Omega(\omega) : \omega \in H, \omega(x_i) = y_i, 1 \leq i \leq n\}.$$

*ii) Para cualquier conjunto finito $\{x_1, ...x_n\} \subseteq X$, para cualquier función $\omega \in Span\{K_{x_1}, ..., K_{x_n}\}$ y para cualquier $p \in (Span\{K_{x_1}, ..., K_{x_n}\})^{\perp}$ se verifica $\Omega(\omega + p) \geq \Omega(\omega)$.*

DEMOSTRACIÓN. La prueba es muy similar a la del lema anterior.

*i)* $\Rightarrow$ *ii)* Sea $\omega \in H$ y consideramos el problema

$$\min\{\Omega(t) : t \in H, t(x_i) = \langle \omega, K_{x_i} \rangle\},$$

que es un caso particular del descrito en el enunciado. Podemos tomar $t = \omega$ y así ver que el problema es satisfactible, por lo que sabemos que existe una solución de la forma

$$\hat{\omega} = \sum_{i=1}^{n} \lambda_i K_{x_i},$$

con $\lambda_i \in \mathbb{C}$. Sea $p \in (Span\{K_{x_1}, ..., K_{x_n}\})^{\perp}$ y entonces

$$\langle \hat{\omega} + p, K_{x_i} \rangle = \langle \hat{\omega}, K_{x_i} \rangle + \langle p, K_{x_i} \rangle = y_i$$

*ii)* $\Rightarrow$ *i)* Consideramos la función

$$\begin{aligned} \phi : X &\longrightarrow H \\ x &\longmapsto K_x. \end{aligned}$$

Consideramos además $\mathscr{S}$ como el subespacio generado por $\{\phi(x_1), ..., \phi(x_n)\} \subseteq H$ y $\pi : H \longrightarrow \mathscr{S}$ la proyección ortogonal. Supongamos que $\omega$ es una solución del problema y consideramos $\hat{\omega} = \pi(\omega)$. Entonces,

$$\omega = \hat{\omega} + p,$$

con $p \in \mathscr{S}^{\perp}$ y $\hat{\omega} \in \mathscr{S}$. Tenemos que $\Omega(\omega) = \Omega(\hat{\omega} + p) \geq \Omega(\hat{\omega})$. Además, se cumple que $\omega(x_i) = y_i$, por lo que

$$\omega(x_i) = \langle \omega, K_{x_i} \rangle = \langle \hat{\omega} + p, K_{x_i} \rangle = \langle \hat{\omega}, K_{x_i} \rangle + \langle p, K_{x_i} \rangle = \hat{\omega}(x_i) = y_i.$$

Como $\hat{\omega} \in Span\{K_{x_1}, ..., K_{x_n}\}$, hemos probado que $\Omega$ es admisible.

$\square$

TEOREMA 2.7.3 (Caracterización de las regularizaciones de Tikhonov). *Sea $H$ un espacio de Hilbert y $\Omega : H \longrightarrow \mathbb{R}$ una regularización de Tikhonov continua y diferenciable (en el sentido de la Definición 25). Supongamos que $dim(H) \geq 2$. Entonces, $\Omega$ es admisible para el problema de interpolación sobre $H$ si y solamente si existe una función monótona creciente $h : \mathbb{R}_+ \longrightarrow \mathbb{R}$ tal que*

$$\Omega(\omega) = h(||\omega||^2), \forall \omega \in H.$$

DEMOSTRACIÓN. Supongamos, en primer lugar, que $dim(H) < \infty$. Tendremos que $H \cong \mathbb{K}^d$ (con $\mathbb{K} = \mathbb{R} \vee \mathbb{C}$) con lo que podemos directamente suponer $H = \mathbb{K}^d$. Haremos la prueba para el caso $\mathbb{K} = \mathbb{R}$, aunque es igualmente válida para $\mathbb{K} = \mathbb{C}$.
Como $\Omega$ es diferenciable, para cada $\omega \in H$, existe $\nabla_\omega \Omega \in H$ tal que

$$\langle \nabla_\omega \Omega, p \rangle = 0$$

para todo $p \in \mathbb{R}^d$ tal que $\langle \omega, p \rangle = 0$. Fijemos $\omega_0 \in S^{d-1} \subseteq \mathbb{R}^d$ un vector de norma 1 ($||\omega_0|| = 1$). Consideremos $\omega \in \mathbb{R}^d$ un elemento cualquiera. Entonces, existe una matriz ortogonal $U \in \mathscr{O}(d)$ (en el caso complejo, una matriz unitaria $U \in U(d)$) tal que

$$\omega = ||\omega|| U \omega_0.$$

Además, existe una matriz antisimétrica $A \in \mathcal{M}_d(\mathbb{R})$ ($A^T = -A$) tal que $U = e^A$, donde

$$e^A = \sum_{k \leq 0}^{\infty} \frac{A^k}{k!},$$

es la matriz exponencial de $A$. Consideremos el camino en $\mathbb{R}^d$ dado mediante:

$$\begin{aligned} \gamma_U : [0,1] &\longrightarrow \mathbb{R}^d \\ t &\longmapsto \gamma_U(t) := ||\omega|| e^{tA} \omega_0. \end{aligned}$$

Tenemos que $\gamma_U(0) = ||\omega|| \omega_0$ y que $\gamma_U(1) = \omega$. Además, el camino satisface

$$\forall t \in [0,1], \qquad \langle \gamma_U(t), \gamma_U(t) \rangle = \langle \omega, \omega \rangle. \tag{2.7.3}$$

Al ser constante, derivando con respecto a $t$ la igualdad 2.7.3 obtendremos

$$\langle \gamma_U'(t), \gamma_U(t) \rangle = 0, \qquad \forall t \in [0,1].$$

Por tanto,

$$\frac{d\Omega(\gamma_U(t))}{dt} = \langle \nabla_{\gamma_U(t)} \Omega, \gamma_U'(t) \rangle = 0.$$

Por tanto, $\Omega(\gamma_U(t))$ es constante (a lo largo de ese camino $\gamma_U$). Por ello

$$\Omega(||\omega|| \omega_0) = \Omega(\omega), \qquad \forall \omega \in H.$$

Definiendo $h : \mathbb{R}_+ \longrightarrow \mathbb{R}$ mediante

$$h(z) = \Omega(\sqrt{z} \omega_0), \qquad \forall z \in \mathbb{R}_+,$$

tendremos que $h$ es no decreciente y satisface la hipótesis del enunciado. Para el caso $dim(H) = \infty$ el argumento es similar, pero usamos un camino como el siguiente

$$\begin{aligned} \gamma : [0,1] &\longrightarrow H \\ t &\longmapsto \gamma(t) = \frac{(1-t)\omega_0 + t\omega}{||(1-t)\omega_0 + t\omega||} ||\omega||. \end{aligned}$$

Este camino es diferenciable cuando $\omega$ no es linealmente independiente de $\omega_0$. Con los mismos argumentos llegaríamos a que $\Omega \circ \gamma$ es constante y se define $h$ de modo similar. Para los $\omega$ linealmente dependientes de $\omega_0$ usaremos el hecho de que $\Omega$ es continua. En cuanto al recíproco, si $\Omega(\omega) = h(||\omega||^2)$, claramente se satisface la tesis del Lema 2.7.1 y $\Omega$ es admisible. $\square$

CAPíTULO 3

# Interpretación del Teorema del Representante como Aproximante en Aprendizaje

## Índice



Como se indica en casi todas las referencias al Teorema del Representante, una de sus principales aplicaciones consiste en su utilización en Aprendizaje. Desafortunadamente, no hemos sido capaces de localizar un resultado conclusivo, válido para un contenido abstracto como el de [**CS, 02**], que permita mostrar la manera de usar este resultado para aproximar funciones de regresión. Es notable observar que ni la referencia clásica, ni [**CS, 02**] sean capaces de ligar el Teorema del Representante con los fundamentos matemáticos del Aprendizaje que ellos mismos pretender establecer. Así, la sección 6 del capítulo III de [**CS, 02**] introduce el Teorema del Representante aunque no explica en qué medida este resultado está ligado a las páginas que le preceden. Ante esta situación hemos optado por introducir un resultado original que liga el Teorema del Representante con los principios de Teoría del Aprendizaje tal y como está establecido en [**CS, 02**]. El resultado no es ni óptimo ni satisfactorio, dado que sólo establece condiciones suficientes de aplicabilidad del Teorema del Representante al Aprendizaje sin mostrar casos concretos en los que esa aplicabilidad es además factible. Es sólo una primera aproximación a las palabras usadas regularmente por los autores del ámbito. Este capítulo se dedica así a enunciar y dar una demostración bastante completa de este teorema original que no hemos podido encontrar en la literatura. Para poder enunciarlo, necesitamos introducir algunas nociones preliminares. En lo que sigue, $(X, d)$ será un espacio métrico, $I \subseteq \mathbb{R}$ un intervalo, $Z = X \times I \subseteq X \times \mathbb{R}$ y $\mathscr{B}(Z)$ la clase de los conjuntos de Borel del espacio producto $X \times I$. Supondremos $\rho : \mathscr{B} \longrightarrow [0, 1]$ una distribución de probabilidad definida sobre $\mathscr{B}(Z)$. Para cualquier función medible $f : X \longrightarrow \mathbb{R}$, consideramos el error cuadrático promedio de $f$ con respecto a la probabilidad $\rho$ dado mediante

$$\mathcal{E}(f) = \int_Z |y - f(x)|^2 d\rho.$$

El propósito de la teoría del Aprendizaje, conforme a los principios de [**CS, 02**] consiste en minimizar este error cuadrático promedio. Consideremos $\mathscr{B}(X)$ la clase de los conjuntos de Borel sobre $X$ y sea $\rho_X : \mathscr{B}(X) \longrightarrow [0, 1]$ la *distribución marginal de $\rho$ sobre $X$* a través de la proyección canónica $\pi : Z \longrightarrow X$. Es decir, para $E \in \mathscr{B}(X)$

$$\rho_X(E) := \rho(\pi^{-1}(E)) = \rho(E \times I).$$

Con respecto a $\rho_X$ podemos definir el espacio $L^2(X)$ de funciones cuadrado integrales. Es decir, las funciones Borel medibles $f : X \longrightarrow \mathbb{R}$ tales que

$$\int_X |f(x)|^2 d\rho_X < \infty.$$

Bajo la hipótesis de que el espacio métrico $(X, d)$ es separable (i.e, posee un subconjunto denso contable), el teorema de Desintegración de la Medida nos garantiza que existe una función $f_\rho : X \longrightarrow \mathbb{R}$ que minimiza $\mathcal{E}$:

$$\mathcal{E}(f_\rho) \leq \mathcal{E}(f), \quad \forall f \in L^2(X).$$

La función $f_\rho$ se denomina *función de regresión de $f$ sobre $X$*. Para que todo sea más conforme, la Teoría del Aprendizaje de [**CS, 02**] añade una hipótesis sobre la distribución de probabilidad $\rho$:

Definición 26. *Con las notaciones precedentes una distribución de probabilidad $\rho$ sobre $Z$ es* admisible *para su tratamiento en Teoría del Aprendizaje si la función de regresión $f_\rho$ es acotada sobre $X$. Es decir, $\rho$ admisible si*

$$||f_\rho||_\infty := sup\{|f_\rho(x)| : x \in X\} < \infty.$$

Bajo la hipótesis de la Teoría del Aprendizaje, si $\rho$ es una distribución de probabilidad admisible se tiene $f_\rho \in L^2(X)$ :

$$\int_X |f_\rho(x)|^2 d\rho_X \leq \int_X ||f_\rho||_\infty^2 d\rho_X = ||f_\rho||_\infty^2 \int_X d\rho_X = ||f_\rho||_\infty^2 < \infty.$$

Más aún, si $|| \cdot ||_{L^2(X)}$ es la norma en el espacio de Hilbert $L^2(X)$, tendremos que

$$||f_\rho||_{L^2(X)} \leq ||f_\rho||_\infty.$$

Por tanto, bajo la hipótesis de la Teoría del Aprendizaje, $f_\rho \in L^2(X)$ es la función que minimiza $\mathcal{E}$ sobre $L^2(X)$. De hecho, probamos, en la Proposición 3.2.3, que para cada $f \in L^2(X)$ se tiene

$$\mathcal{E}(f) - \mathcal{E}(f_\rho) = |\mathcal{E}(f) - \mathcal{E}(f_\rho)| = ||f - f_\rho||_{L^2(X)}^2.$$

Como, normalmente, $f_\rho$ no es fácilmente computable, se hace una aproximación de $f_\rho$ mediante un subespacio de Hilbert $H \subseteq L^2(X)$ con buenas propiedades de aproximación de $f_\rho$ que, al mismo tiempo, sean manipulables.

Definición 27. *Con las notaciones precedentes, sea $\delta > 0$ un número real positivo y $H \subseteq L^2(X)$ un subespacio de Hilbert. Diremos que $H$ posee una* $\delta$-aproximación de la función de regresión *$f_\rho : X \longrightarrow \mathbb{R}$ anterior si existe $f_H \in H$ tal que*

$$||f_\rho - f_H||_{L^2(X)}^2 < \delta.$$

En la Proposición 3.2.5 probaremos que dado un espacio de Hilbert $H$ que contiene una $\delta$-aproximación de $f_\rho$ y siendo $f_H \in H$ un elemento que minimiza $\mathcal{E}$ sobre $H$ se tiene:

$$||f_\rho - f_H||_{L^2(X)}^2 < \delta,$$
$$|\mathcal{E}(f_H) - \mathcal{E}(f_\rho)| < \delta.$$

En este sentido, $f_H \in H$ sería una buena aproximación de $f_\rho$ y, por tanto, una buena codificación de la función de regresión a través de elementos de $H$. No es el propósito de este Trabajo de Fin de Grado hacer una disquisición sobre la búsqueda de buenos espacios de Hilbert para aproximar la función de regresión. Nuestro objetivo es más modesto: *pretendemos aproximar $f_H$ a partir del Teorema del Representante.* Sin embargo, para que nuestra exposición no quedase corta en este aspecto, hemos descrito en la Subsección 3.2.4 una breve discusión sobre los ratios de Aproximación en el caso de espacios de Hilbert definidos por núcleos de Mercer. A partir de aquí nos ocupamos solamente del caso en que $H$ sea un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$ y que $H$ contenga una $\delta$-aproximación de $f_\rho$. Nuestro objetivo será calcular $f_H$ o, en su defecto, aproximar $f_H$ y aquí es donde interviene el Teorema del Representante.

Para ello, sea $z = ((x_1, y_1), ..., (x_n, y_n)) \in Z^n$ una observación empírica de un evento donde $n \in \mathbb{N}, n > 1$. Para cada $f \in H$, llamaremos error empírico de $f$ en $z$ a la cantidad

$$\mathcal{E}_z(f) = \frac{1}{n}\sum_{i=1}^{n} |y_i - f(x_i)|^2.$$

No queda claro, a priori, que el error empírico tenga relación con $\mathcal{E}(f)$ para clases amplias de funciones. Por ello, dado $\epsilon > 0$ y dado $\mathcal{F} \subseteq H$ definimos:

$$H_\rho^{(n)}(\mathcal{F}, \varepsilon) := \{z \in Z^n : \sup_{f\in\mathcal{F}} |\mathcal{E}_z(f) - \mathcal{E}(f)| < \epsilon\}.$$

Las grandes contribuciones de V. N. Vapnik y A. Ya. Chervonenkis (cf. [**CV, 71**]), para $\mathcal{F}$ formado por funciones características de conjuntos de Borel de $X$, o de autores como D. Pollard (cf. [**Po, 84**]), con el uso de la esperanza de números de recubrimiento, consisten en mostrar clases $\mathcal{F}$ para las cuales $H_\rho^{(n)}(\mathcal{F}, \varepsilon) \neq \emptyset$. Más aún, los teoremas de estos autores son de la forma

$$Prob\left(H_\rho^{(n)}(\mathcal{F}, \varepsilon)\right) \geq 1 - N_\rho(\mathcal{F}, \varepsilon, n)e^{-\frac{\epsilon^2 m}{D}},$$

para una cierta constante $D$ y una cierta función $N_\rho(\mathcal{F}, \mathcal{E}, n)$ usualmente denominada *número de recubrimiento*. En la Subsección 3.2.6 mostramos un ejemplo de una cota inferior de este tipo basada en el método combinatorio de Pollard y ciertas disgresiones estudiadas en [**PS, 20**]. Para concluir nuestras notaciones, si $x = (x_1, ..., x_n) \in X^n$, denotaremos por $G_x$ a la matriz de Gram asociada al núcleo $K$ y a la lista de puntos de $x$. Es decir,

$$G_x = \begin{pmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{pmatrix} = (K(x_i, x_j))_{1\leq i,j\leq n},$$

Denotamos por $({G_x}^{\frac{1}{2}})$ a la matriz raíz cuadrada de $G_x$, por $\mu({G_x}^{\frac{1}{2}})$ su número de condicionamiento y por $||G_x^{\frac{1}{2}}||_F$ su norma de Frobenius. Finalmente, dado $z \in Z^n$, denotamos por $f_z \in H$ a un elemento que minimiza el error empírico $\mathcal{E}_z$. Nuestro primer resultado es el siguiente, que acota la norma en $H$ del $f_z$ dado por el Teorema del Representante.

TEOREMA 3.0.1. *Con las notaciones precedentes, sea $(X, d)$ un espacio métrico, $H$ un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$. Sea $z = ((x_1, y_1), ..., (x_n, y_n)) \in Z^n$ y sea*

$$W_z := Span(\{K_{x_1}, ..., K_{x_n}\}),$$

*el subespacio vectorial de $H$ generado por $\{K_{x_1}, ..., K_{x_n}\}$. Entonces, existe $f_z \in W_z$ tal que*

*i)*

$$\mathcal{E}_z(f_z) = \min\{\mathcal{E}_z(f) : f \in H\}$$

*ii)*

$$||f_z||_H \leq \frac{\mu(G_x^{\frac{1}{2}})}{||G_x^{\frac{1}{2}}||_F}||y||_2.$$

Este resultado (Teorema 3.1.1) se prueba en la Sección 3.1. Utilizando todas estas herramientas, probamos finalmente el siguiente enunciado en la Sección 3.2.

TEOREMA 3.0.2. *Sea $(X, d)$ un espacio métrico separable, $I \subseteq \mathbb{R}$ un intervalo, $Z = X \times I$ y $\rho$ una distribución de probabilidad sobre los subconjuntos de Borel de $Z$. Supongamos que $\rho$ es admisible para el aprendizaje y que la norma $||f_\rho||_\infty$ de la función de regresión es finita. Sean $R, \varepsilon, \delta \in \mathbb{R}$ números reales con $0 < \varepsilon, 0 < \delta < 1$. Sea $H$ un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$ que contiene una $\delta$-aproximación $f_H \in H$ de $f_\rho$. Supongamos que $R$ satisface las siguientes propiedades:*

*i)* $H_g^{(n)}(B_H(0, R), \varepsilon) \neq \emptyset$*, donde* $B_H(0, R)$ *es la bola en* $H$ *de centro* $0$ *y radio* $R$*.*
*ii)* $||f_\rho||_\infty \leq R - 1$.

*Supongamos que existe $z = ((x_1, y_1), ..., (x_n, y_n)) \in H_g^{(n)}(B_H(0, R), \varepsilon)$ tal que*

$$\mu({G_x}^{\frac{1}{2}})||y||_2 \leq R||{G_x}^{\frac{1}{2}}||_F.$$

*Entonces, existe* $f_z \in W_z = Span(\{K_{x_1}, .., K_{x_n})$ *tal que*

$$||f_z - f_\rho||^2_{L^2(X)} \leq 2\varepsilon + \delta.$$

$$|\mathcal{E}(f_z) - \mathcal{E}(f_\rho)| \leq 2\varepsilon + \delta.$$

## 3.1. Acotando la Norma del Aproximante del Teorema del Representante

No hemos encontrado en la literatura ninguna acotación de la norma del elemento óptimo que aparece en el Teorema del Representante (cf. Teorema 2.6.1 anterior). En esta Sección vamos a aportar una *contribución original* al Trabajo de Fin de Grado, escribiendo una acotación de esa norma de $f_z$ en función del condicionamiento de la matriz de Gram. Comencemos enunciando este resultado, para el cual necesitamos recordar algunas notaciones de las usadas en Capítulos precedentes. Consideremos un espacio métrico separable $(X, d)$ y $H \subseteq \mathbb{R}^X$ un espacio de Hilbert con núcleo reproductor de funciones definidas en $X$ a valores reales. Sea $K : X \times X \longrightarrow \mathbb{R}$ el núcleo reproductor de $H$. Dado $x = (x_1, ..., x_n) \in X^n$ una lista de $n$ puntos de $X$, llamaremos matriz de Gram asociada a $x$ a la matriz

$$G_x = \begin{pmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{pmatrix} \in \mathscr{M}_n(\mathbb{R}).$$

Por hipótesis, $G_x$ es una matriz simétrica, semidefinida positiva.
Sea $z = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{R})^n$ una observación empírica y definamos sobre $H$ la función de error empírico cuadrático en media:

$$\begin{aligned} \mathcal{E}_z : H &\longrightarrow \mathbb{R} \\ f &\longmapsto \frac{1}{n}\sum_{i=1}^{n} |y_i - f(x_i)|^2 \end{aligned}$$

Dado $z$, consideremos la familia de funciones $K_{x_1}, ..., K_{x_n} \in H$ y sea $W_z = Span(\{K_{x_1}, ..., K_{x_n}\})$ el subespacio vectorial de $H$ generado por las funciones $\{K_{x_1}, ..., K_{x_n}\}$. Nótese que la función $\mathcal{E}_z$ anterior puede representarse como una función de riesgo

$$\begin{aligned} \mathcal{E} : (X \times \mathbb{R}^2)^n &\longrightarrow \mathbb{R} \\ ((x_1, y_1, u_1), ..., (x_n, y_n, u_n)) &\longmapsto \tfrac{1}{n}\textstyle\sum_{i=1}^{n} |y_i - u_i|^2, \end{aligned}$$

y que para cada $f \in H$

$$\mathcal{E}_z(f) = \mathcal{E}\left((x_1, y_1, f(x_1)), ..., (x_n, y_n, f(x_n))\right).$$

Por tanto, aplicando el Teorema del Representante (cf. Teorema 2.6.1 anterior), dado $z \in (X \times \mathbb{R})^n$, existe $f_z \in W_z$ tal que

$$\mathcal{E}_z(f_z) = \min\{\mathcal{E}_z(f) : f \in H\}.$$

Llamaremos a $f_z$ el *aproximante de* $H$ *asociado a* $\mathcal{E}_z$. Nuestro objetivo en esta sección consiste en probar el siguiente resultado

TEOREMA 3.1.1. *Con las notaciones precedentes, se tiene:*

$$||f_z||_H \leq \frac{\mu({G_x}^{\frac{1}{2}})}{||{G_x}^{\frac{1}{2}}||_F} ||y||_2,$$

*donde*

*i)* $||y||_H$ *es la norma en* $H$ *con respecto a su producto interno.*
*ii)* ${G_x}^{\frac{1}{2}}$ *es una raíz cuadrada de la matriz de Gram* $G_x$*.*
*iii)* $||A||_F$ *es la norma de Frobenius de cualquier matriz* $A \in \mathscr{M}_n(\mathbb{R})$*.*
*iv)* $\mu(A)$ *es el condicionamiento de* $A \in \mathscr{M}_n(\mathbb{R})$*, definido para* $A$ *del modo siguiente:*

$$\mu(A) = ||A||_F ||A^+||_2,$$

*donde* $A^+$ *es la pseudoinversa de Moore-Penrose de* $A$ *y* $||B||_2$ *es la norma de* $B$ *como operador lineal para cualquier* $B \in \mathscr{M}_n(\mathbb{R})$*.*

**3.1.1. Proyección Ortogonal, SVD y Pseudoinversa de Moore-Penrose.** Comencemos revisando algunas propiedades de Álgebra Lineal elemental en espacios de Hilbert reales de dimensión finita. Consideremos $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert de dimensión finita sobre $\mathbb{R}$. Sea $W \subseteq \mathbb{R}^n$ un subespacio vectorial, también de dimensión finita y sea $\mathcal{B} = \{w_1, ..., w_r\}$ una base ortonormal con respecto a $\langle \cdot, \cdot \rangle$ y definimos la siguiente aplicación lineal:

$$\begin{aligned} \pi_W : \mathbb{R}^n &\longrightarrow W \\ v &\longmapsto \sum_{i=1}^{r} \langle v, w_i \rangle w_i. \end{aligned}$$

Esta aplicación lineal se conoce como *proyección ortogonal sobre* $W$ y satisface las siguientes propiedades:

Proposición 3.1.2. *Con las notaciones anteriores, para cada* $v \in \mathbb{R}^n$ *se tiene que*

$$||v - \pi_W(v)|| = \min\{||v - w|| : w \in W\}.$$

Demostración. Es un resultado clásico, esencialmente conocido y basado en lo que significa el método de Gram-Schmidt y el Teorema del Reemplazamiento en espacios vectoriales de dimensión finita. Podemos proceder del modo siguiente:

- Consideramos $\mathcal{B} = \{w_1, ..., w_r\}$ la base ortonormal de $W$, dado por el enunciado.
- Por el Teorema del Reemplazamiento, ampliamos $\mathcal{B}$ hasta obtener una base $\tilde{\mathcal{B}}$ de $\mathbb{R}^n$ de la forma siguiente:

$$\tilde{\mathcal{B}} := \{w_1, ..., w_r, u_{r+1}, ..., u_n\}.$$

- Aplicando el método de Gram-Schmidt, obtendremos una base ortonormal $\mathcal{B}^*$ de $\mathbb{R}^n$ de la forma siguiente:

$$\mathcal{B}^* = \{w_1^*, ..., w_r^*, u_{r+1}^*, ..., u_n^*\}.$$

Pero, por la forma del método de Gram-Schmidt, la construcción permitirá concluir que

$$w_i^* = w_i, \quad 1 \leq i \leq r.$$

En otras palabras, existe una base ortonormal $\mathcal{B}^*$ de $\mathbb{R}^n$ que tiene la forma

$$\mathcal{B}^* = \{w_1, ..., w_r, u_{r+1}^*, ..., u_n^*\}.$$

Así, $\{u_{r+1}^*, ..., u_n^*\}$ es una base ortonormal del complementario ortogonal $W^\perp \subseteq \mathbb{R}^n$ de $W$. Más aún, para cada $v \in \mathbb{R}^n$ tendremos que

$$v = \sum_{i=1}^{r} \langle v, w_i \rangle w_i + \sum_{j=r+1}^{n} \langle v, u_j^* \rangle u_j^*.$$

Nótese que

$$\pi_W(v) := \sum_{i=1}^{r} \langle v, w_i \rangle w_i.$$

Por tanto, por ser $\mathcal{B}^*$ ortonormal,

$$||v - \sum_{i=1}^{r} \langle v, w_i \rangle w_i||^2 = \sum_{j=r+1}^{n} |\langle v, u_j^* \rangle|^2.$$

Es más, dado $w \in W$ un elemento cualquiera y supuesto $w = \sum_{i=1}^{r} \lambda_i w_i$, tenemos

$$||v - w||^2 = \sum_{i=1}^{r} |\langle v, w_i \rangle - \lambda_i|^2 + \sum_{j=r+1}^{n} |\langle v, u_j^* \rangle|^2.$$

En estas dos últimas igualdades hemos aplicado el Teorema 1.2.4. Claramente se tiene que

$$||v - w||^2 \geq \sum_{r+1}^{n} |\langle v, u_j^* \rangle|^2 = ||v - \pi_W(v)||^2,$$

con lo que se tiene el resultado. $\square$

Recordemos, por un momento, la descomposicón en valores singulares de matrices con coordenadas reales.

TEOREMA 3.1.3. *Sea* $A \in \mathscr{M}_{m\times n}(\mathbb{R})$ *una matriz de coordenadas reales con* $m$ *filas y* $n$ *columnas. Entonces, existe* $U \in \mathscr{O}(m)$, $V \in \mathscr{O}(n)$ *matrices ortogonales (respectivamente* $m \times m$ *y* $n \times n$*) y existen* $\sigma_1, ..., \sigma_r \in \mathbb{R}$ *números reales que satisfacen*

$$\sigma_1 \geq ... \geq \sigma_r \geq 0,$$

*tales que*

$$A = U \left(\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \sigma_r & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) V^T \tag{3.1.1}$$

*Además,* $r = rank(A)$ *es el rango de la matriz* $A$ *y* $\sigma_1, ..., \sigma_r$ *son únicos con esas propiedades. A la descomposición 3.1.1 se le llama* descomposición de $A$ en valores singulares *y, adicionalmente, se satisface*

*i)* $\sigma_1 = ||A||_2$ *es la norma* $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *como operador lineal.*
*ii)* $||A||_F^2 = \sum_{i=1}^r \sigma_i^2$ *es la norma de Frobenius de* $A$*.*

$$||A||_F := (Tr(A^T A))^{\frac{1}{2}},$$

*donde* $Tr$ *es la traza y* $A^T$ *es la matriz traspuesta de* $A$*.*
*iii) Los valores propios de la matriz* $A^T A$ *cumplen la siguiente igualdad*

$$\sigma_i = \sqrt{\lambda_i}$$

.

**3.1.2. La SVD de Matrices Simétricas Semidefinidas Positivas.** Supongamos ahora $A \in \mathscr{M}_n(\mathbb{R})$ una matriz simétrica semidefinida positiva (i.e, $u^T Au \geq 0, \forall u \in \mathbb{R}^n$). Entonces, su descomposición en valores singulares y su forma canónica de Jordan coinciden en el sentido siguiente: Existe $U \in \mathscr{O}(n)$ una matriz ortogonal y $\sigma_1 \geq ... \geq \sigma_r \geq 0$ números reales positivos tales que

$$A = U \left(\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \sigma_r & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T \tag{3.1.2}$$

Además, $\{\sigma_1, ..., \sigma_r\}$ son los valores propios de $A$.

DEFINICIÓN 28. *Sea* $A \in \mathscr{M}_n(\mathbb{R})$ *una matriz simétrica, semidefinida positiva que admite una descomposición en valores singulares como la descrita en 3.1.2. Definiremos*

*i) La raíz cuadrada de* $A$

$$A^{\frac{1}{2}} = U \left(\begin{array}{cccc|c} {\sigma_1}^{\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & {\sigma_r}^{\frac{1}{2}} & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T$$

*ii) La pseudoinversa de Moore-Penrose de A*

$$A^{+} = U \left(\begin{array}{cccc|c} {\sigma_1}^{-1} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & {\sigma_r}^{-1} & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T$$

*iii) La pseudoinversa de Moore-Penrose de* $A^{\frac{1}{2}}$

$$(A^{\frac{1}{2}})^{+} = U \left(\begin{array}{cccc|c} {\sigma_1}^{-\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & {\sigma_r}^{-\frac{1}{2}} & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T$$

LEMA 3.1.4. *Con las notaciones precedentes, sea* $A \in \mathscr{M}_n(\mathbb{R})$ *una matriz simétrica semidefinida positiva. Sea* $y \in \mathbb{R}^n$ *y sea* $\lambda = (\lambda_1, ..., \lambda_n) \in \mathbb{R}^n$ *dado por*

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \vdots \\ \lambda_n \end{pmatrix} = A^{+} \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}$$

*Entonces,*

*i) La proyección ortogonal de y sobre el subespacio* $Im(A)$ *generado por las columnas de la matriz A es* $A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$. *Es decir:*

$$||y - A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} ||_2^2 = \min\{||y - v||^2 : v \in Im(A)\}.$$

*ii) Se verifica la siguiente relación:*

$$A^{\frac{1}{2}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \vdots \\ \lambda_n \end{pmatrix} = (A^{\frac{1}{2}})^{+} \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}$$

DEMOSTRACIÓN. Se trata simplemente de verificar ambas propiedades

i) Comencemos calculando la distancia en $|| \cdot ||_2$ de $y$ a $A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$:

$$I_\lambda = ||y - A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} ||^2 = ||y - U \left(\begin{array}{cccc|c} {\sigma_1}^{-\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & {\sigma_r}^{-\frac{1}{2}} & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T \lambda||^2,$$

pero

$$U \left(\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \sigma_r & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T \lambda = U \left(\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \sigma_r & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T U \left(\begin{array}{cccc|c} {\sigma_1}^{-1} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & {\sigma_r}^{-1} & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T y.$$

Por tanto,

$$A\begin{pmatrix}\lambda_1\\ \vdots\\ \vdots\\ \lambda_r\end{pmatrix}=U\left(\begin{array}{c|c} Id_r & 0\\ \hline 0 & 0\end{array}\right)U^Ty.$$

Como $U\in\mathscr{O}(n)$, entonces $U$ es una isometría para el producto escalar canónico en $\mathbb{R}^n$ (y lo mismo para $U^T$). Por tanto,

$$I_\lambda=||U^Ty-U^TA\begin{pmatrix}\lambda_1\\ \vdots\\ \vdots\\ \lambda_n\end{pmatrix}||^2=||U^Ty-\left(\begin{array}{c|c} Id_r & 0\\ \hline 0 & 0\end{array}\right)U^Ty||^2.$$

Considerando $y'=U^Ty$, tendremos que

$$\left(\begin{array}{c|c} Id_r & 0\\ \hline 0 & 0\end{array}\right)y'=\pi_1(y'),$$

donde $\pi_1:\mathbb{R}^n\longrightarrow\mathbb{R}^r\times\{0\}^{n-r}$ es la proyección ortogonal sobre el subespacio $\mathbb{R}^r\times\{0\}^{n-r}$. En otras palabras, el punto $(y_1',...,y_r',0,...,0)\in\mathbb{R}^r\times\{0\}^{n-r}$ verifica que

$$||y'-\left(\begin{array}{c|c} Id_r & 0\\ \hline 0 & 0\end{array}\right)y'||^2\leq||y'-v||^2,\qquad\forall v\in\mathbb{R}^r\times\{0\}^{n-r},$$

y además,

$$||y'-\left(\begin{array}{c|c} Id_r & 0\\ \hline 0 & 0\end{array}\right)y'||^2=\sum_{r+1}^{n}|y_i'|^2.$$

Sea $v\in\mathbb{R}^n$ un punto cualquiera y consideremos

$$||y-Av||^2=||U^Ty-\left(\begin{array}{cccc|c}\sigma_1&0&\cdots&0&0\\ 0&\ddots&&\vdots&\vdots\\ \vdots&&\ddots&0&\vdots\\ 0&\cdots&0&\sigma_r&0\\ \hline 0&\cdots&\cdots&0&0\end{array}\right)U^Tv||^2=||y'-\left(\begin{array}{cccc|c}\sigma_1&0&\cdots&0&0\\ 0&\ddots&&\vdots&\vdots\\ \vdots&&\ddots&0&\vdots\\ 0&\cdots&0&\sigma_r&0\\ \hline 0&\cdots&\cdots&0&0\end{array}\right)v'||^2,$$

pero

$$\left(\begin{array}{cccc|c}\sigma_1&0&\cdots&0&0\\ 0&\ddots&&\vdots&\vdots\\ \vdots&&\ddots&0&\vdots\\ 0&\cdots&0&\sigma_r&0\\ \hline 0&\cdots&\cdots&0&0\end{array}\right)v'\in\mathbb{R}^r\times\{0\}^{n-r},$$

con lo que

$$||y-Av||^2=||y'-\left(\begin{array}{cccc|c}\sigma_1&0&\cdots&0&0\\ 0&\ddots&&\vdots&\vdots\\ \vdots&&\ddots&0&\vdots\\ 0&\cdots&0&\sigma_r&0\\ \hline 0&\cdots&\cdots&0&0\end{array}\right)v'||\geq\sum_{r+1}^{n}|y_i'|^2,$$

lo que significa que

$$||y - Av||^2 \geq I_\lambda = ||y - A \begin{pmatrix} \lambda_1 \\ \vdots \\ \vdots \\ \lambda_n \end{pmatrix} ||^2.$$

Con ello, queda probado el apartado.

ii) La segunda propiedad es una mera comprobación que haremos a continuación. Por las definiciones antes realizadas tendremos

$$A^{\frac{1}{2}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \vdots \\ \lambda_n \end{pmatrix} = A^{\frac{1}{2}} A^+ \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix},$$

pero

$$A^{\frac{1}{2}} A^+ = U \left(\begin{array}{cccc|c} {\sigma_1}^{\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & {\sigma_r}^{\frac{1}{2}} & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T U \left(\begin{array}{cccc|c} {\sigma_1}^{-1} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & {\sigma_r}^{-1} & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T =$$

$$= U \left(\begin{array}{cccc|c} {\sigma_1}^{-\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & {\sigma_r}^{-\frac{1}{2}} & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 \end{array}\right) U^T = (A^{\frac{1}{2}})^+.$$

Luego

$$A^{\frac{1}{2}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \vdots \\ \lambda_n \end{pmatrix} = (A^{\frac{1}{2}})^+ \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}$$

$\square$

**3.1.3. Demostración del Teorema 3.1.1.** Retomamos las notaciones introducidas al principio de la Sección. Así $H$ es un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$. Dado $z = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{R})^n$, consideramos la matriz de Gram $G_x = (K(x_i, x_j))_{1 \leq i,j \leq n}$ y su raíz cuadrada ${G_x}^{\frac{1}{2}}$, definida en la subsección precedente (porque $G_x$ es cuadrada, simétrica y semidefinida positiva). Sea ${G_x}^+$ la pseudoinversa de Moore-Penrose de $G_x$, definida también en la subsección precedente. Definimos

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \vdots \\ \lambda_n \end{pmatrix} = {G_x}^+ \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n \tag{3.1.3}$$

Finalmente, definimos la función

$$f_\lambda = \sum_{i=1}^{n} \lambda_i K_{x_i} \in W_z = Span(\{K_{x_1}, ..., K_{x_n}\}) \subseteq H. \tag{3.1.4}$$

Esta función verifica las siguientes propiedades:

PROPOSICIÓN 3.1.5. *Con las notaciones precedentes, se tiene:*

*i) La función $f_\lambda$ minimiza $\mathcal{E}_z$ sobre $W_z$. Es decir, $\forall f \in W_z$*

$$\mathcal{E}_z(f_\lambda) = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_\lambda(x_i)|^2 = \min\{\mathcal{E}_z(f) : f \in W_z\}.$$

*ii) La función $f_\lambda$ minimiza el error empírico $\mathcal{E}_z$ sobre $H$. Es decir,*

$$\mathcal{E}_z(f_\lambda) = \min\{\mathcal{E}_z(f) : f \in H\}.$$

*En particular $f_\lambda = f_z \in W_z$ es uno de los aproximantes determinados por el Teorema del Representante (Teorema 2.6.1).*

*iii)* $||f_\lambda||_H = ||f_z||_H = ||(G_x{}^{\frac{1}{2}})^+ y||$.

DEMOSTRACIÓN. Probemos las propiedades una tras otra.

i) Comencemos observando lo siguiente: Sea $\theta = (\theta_1, ..., \theta_n) \in \mathbb{R}^n$ y $f_\theta = \sum_{i=1}^n \theta_i K_{x_i} \in W_z$. Tendremos que:

$$\sum_{i=1}^{n}|y_i - f_\theta(x_i)|^2 = \sum_{i=1}^{n}|y_i - \sum_{j=1}^{n}\theta_j K_{x_j}(x_i)|^2 = \sum_{i=1}^{n}|y_i - \sum_{j=1}^{n}K(x_j, x_i)\theta_j|^2.$$

Consideramos el vector

$$G_x\theta = \begin{pmatrix} K(x_1,x_1) & \cdots & \cdots & K(x_1,x_n) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ K(x_n,x_1) & \cdots & \cdots & K(x_n,x_n) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \vdots \\ \theta_n \end{pmatrix}.$$

De las anteriores igualdades concluimos, por ser $G_x$ simétrica, que

$$\sum_{i=1}^{n}|y_i^2 - f_\theta(x_i)|^2 = ||y - G_x\theta||_2^2,$$

donde $||v||_2 = (\sum_{i=1}^n |v|^2)^{\frac{1}{2}}$ es la norma euclídea canónica de cualquier vector $v \in \mathbb{R}^n$. Como $G_x \in \mathscr{M}_n(\mathbb{R})$ es cuadrada, simétrica y semidefinida positiva, el Lema 3.1.4 precedente implica que si definimos

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \vdots \\ \lambda_n \end{pmatrix} = G_x{}^+ \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix},$$

entonces, el vector $G_x\lambda$ minimiza la distancia euclídea al punto $y \in \mathbb{R}$. En particular, para todo $\theta \in \mathbb{R}^n$ se tiene que

$$||y - G_x\theta||_2^2 \geq ||y - G_x\lambda||_2^2.$$

Por tanto,

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - f_\theta(x_i)|^2 = \frac{1}{n}||y - G_x\theta||^2 \geq \frac{1}{n}||y - G_x\lambda||^2 = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_\lambda(x_i)|^2.$$

Luego $f_\lambda$ minimiza $\mathcal{E}_z$ sobre $W_z$.

ii) El Teorema del Representante implica que dado $z \in (X \times \mathbb{R})^n$, existe $g \in W_z = Span(\{K_{x_1}, ..., K_{x_n}\})$ tal que

$$\mathcal{E}_z(g) = \min\{\mathcal{E}_z(f) : f \in H\}.$$

Por el apartado *i)* precedente, $f_\lambda$ satisface

$$\mathcal{E}_z(f_\lambda) = \min\{\mathcal{E}_z(g) : g \in W_z\} = \min\{\mathcal{E}_z(f) : f \in H\}.$$

En particular, $f_\lambda$ es uno de los aproximantes (llamados $f_z$) que minimizan $\mathcal{E}_z$ sobre $H$ y que pertenecen a $W_z$. Escribimos $f_z = f_\lambda$.

iii) Es un resultado de mera comprobación:

$$||f_\lambda||_H^2 = \langle\sum_{i=1}^n \lambda_i K_{x_i}, \sum_{i=1}^n \lambda_i K_{x_i}\rangle = \sum_{i=1}^n\sum_{j=1}^n \lambda_i\langle K_{x_i}, K_{x_j}\rangle\lambda_j = \sum_{i=1}^n\sum_{j=1}^n \lambda_i K(x_i,x_j)\lambda_j.$$

Por tanto, podemos escribir

$$||f_\lambda||_H^2 = (\lambda_1,...,\lambda_n)G_x\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}.$$

Ahora, recordemos que si $G_x$ es simétrica y semidefinida positiva, su raíz cuadrada ${G_x}^{\frac{1}{2}}$ es también simétrica. Así es, puesto que si

$$G_x = U\left(\begin{array}{ccc|c}\sigma_1 & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & \sigma_n & 0\\ \hline 0 & \cdots & 0 & 0\end{array}\right)U^T,$$

con $U \in \mathscr{O}(n)$ ortogonal y $\sigma_1 \geq ... \geq \sigma_n > 0$. Entonces

$${G_x}^{\frac{1}{2}} = U\left(\begin{array}{ccc|c}{\sigma_1}^{\frac{1}{2}} & \cdots & 0 & 0\\ \vdots & \ddots & & \vdots\\ 0 & \cdots & {\sigma_n}^{\frac{1}{2}} & 0\\ \hline 0 & \cdots & 0 & 0\end{array}\right)U^T,$$

y es fácil verificar que su traspuesta verifica:

$$({G_x}^{\frac{1}{2}})^T = ({G_x}^{\frac{1}{2}}).$$

Por otro lado, ${G_x}^{\frac{1}{2}}$ es, propiamente, una raíz cuadrada de $G_x$ en el sentido

$$({G_x}^{\frac{1}{2}})({G_x}^{\frac{1}{2}}) = G_x,$$

lo cual es fácil de comprobar a partir de las definiciones. Por tanto, ambas propiedades implican

$$({G_x}^{\frac{1}{2}})^T({G_x}^{\frac{1}{2}}) = ({G_x}^{\frac{1}{2}})({G_x}^{\frac{1}{2}}) = G_x.$$

En particular,

$$||f_\lambda||_H^2 = (\lambda_1,...,\lambda_n)({G_x}^{\frac{1}{2}})^T({G_x}^{\frac{1}{2}})\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix} =$$

$$= \langle {G_x}^{\frac{1}{2}}\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}, {G_x}^{\frac{1}{2}}\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}\rangle = ||{G_x}^{\frac{1}{2}}\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}||_2^2.$$

Por el apartado *iii*) del Lema 3.1.4 concluimos

$$||f_\lambda||_H^2 = ||{G_x}^{\frac{1}{2}}\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}||_2^2 = ||({G_x}^{\frac{1}{2}})^+\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}||_2^2$$

□

Finalmente, el Teorema 3.1.1 es una forma particular del siguiente Corolario, que procederemos a probar:

COROLARIO 3.1.6. *Con las notaciones de la Proposición precedente, se tiene:*

*i)* $||f_\lambda||_H \leq ||({G_x}^{\frac{1}{2}})^+||_2 \, ||\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}||_2$

*ii)* $||f_\lambda||_H \leq \frac{\mu({G_x}^{\frac{1}{2}})}{||{G_x}^{\frac{1}{2}}||_F}||y||_2.$

DEMOSTRACIÓN. La propiedad *i)* se sigue del apartado *iii)* de la Proposición anterior, mientras que el apartado *ii)*, es decir, el Teorema 3.1.1 del comienzo de la Sección, se tiene porque

$$\mu(G_x^{\frac{1}{2}}) = ||G_x^{\frac{1}{2}}||_F||(G_x^{\frac{1}{2}})^+||_2.$$

□

## 3.2. Breve Introducción a los Fundamentos Matemáticos del Aprendizaje

**3.2.1. La Función de Regresión en Aprendizaje.** En el ao 2002, F. Cucker y S. Smale publicaron su trabajo [**CS, 02**], con el que pretendían exponer los fundamentos matemáticos de la Teoría del Aprendizaje. Aunque el trabajo no es original (es una relectura de resultados y terminología) en su mayor parte, nos permitirá contextualizar la aplicación del Teorema del Representante del Capítulo anterior a la Teoría del Aprendizaje. Comencemos con unas pocas anotaciones preliminares. Sea $(X, d)$ un espacio métrico y denotamos por $Z = X \times \mathbb{R}$ con la topología producto. Sea $\mathcal{B}(Z)$ la clase de conjuntos de Borel de $Z$ y $\rho : \mathcal{B}(Z) \longrightarrow [0, 1]$ una distribución de probabilidad definida sobre $\mathcal{B}(Z)$. Dada $f : X \longrightarrow \mathbb{R}$ una función a valores reales definida en $X$ definiremos el error cuadrático medio de $f$ respecto a $\rho$ mediante:

$$\mathcal{E}(f) = \int_Z |y - f(x)|^2 d\rho.$$

El objetivo principal de la Teoría del Aprendizaje consiste en determinar una función $f_\rho : X \longrightarrow \mathbb{R}$ que minimiza el error cuadrático medio, es decir, algo similar a lo siguiente:

$$\mathcal{E}(f_\rho) = \min\{\mathcal{E}(f) : \mathcal{E}(f) < +\infty\}.$$

La función $f_\rho$ resulta, como veremos, difícil de explicitar , por tanto, el Aprendizaje Computacional busca construir a partir de una observación empírica $z = ((x_1, y_1), ..., (x_n, y_n)) \in (X \times \mathbb{R})^n$ una función $f_z : X \longrightarrow \mathbb{R}$ que sea una buena aproximación de $f_\rho$ y, al mismo tiempo, $\mathcal{E}(f_z)$ sea próximo a $\mathcal{E}(f_\rho)$. Aquí, los terminos $f_\rho$ y "buena aproximación" requieren un poco más de precisión, aspector que iremos desarrollando de manera breve, sin entrar en las pruebas de las mismas, aunque daremos referencias de los principales resultados que escribiremos.

**3.2.2. La Función de Regresión $f_\rho$.** En primer lugar, vamos a dar argumentos que garanticen la existencia de $f_\rho$. Para ello, consideramos $\rho$ una distribución de probabilidad en $\mathcal{B}(Z)$, con $Z = X \times \mathbb{R}$. La distribución marginal de $\rho$ sobre $X$ es la dada mediante:

$$\begin{aligned} \rho_X : \mathcal{B}(X) &\longrightarrow [0, 1] \\ A &\longmapsto f_X(A) := \rho(\pi^{-1}(A)) = \rho(A \times \mathbb{R}). \end{aligned}$$

Nos apoyaremos en un teorema clásico de Teoría de la Medida, llamado Teorema de Desintegración, que enunciaremos en la versión particular de nuestro caso:

TEOREMA 3.2.1 (de Desintegración). *Con las notaciones precedentes, supongamos que $(X, d)$ es un espacio métrico separable (i.e, posee un subconjunto denso contable), sea $Z = X \times \mathbb{R}$ y sea $\rho$ una distribución de probabilidad sobre $\mathcal{B}(Z)$ y $\pi : Z \longrightarrow X$ la proyección canónica. Entonces, existe una familia $\{\rho_{(x)} : x \in X\}$ de distribución de probabilidad sobre $\mathcal{B}(Z)$ que satisface las siguientes propiedades:*

*i) Para cada $B \in \mathcal{B}(Z)$, la siguiente función es medible*

$$\begin{aligned} X &\longrightarrow [0, 1] \\ x &\longmapsto \rho_{(x)}(B) \end{aligned}$$

*ii) La función $\rho_{(x)}$ "vive dentro de la fibra $\pi^{-1}(\{x\})$", salvo por un conjunto de medida nula en $X$ con respecto a $\rho_X$. Es decir, $\exists A \subseteq X$ tal que $\rho_X(A) = 0$ tal que $\forall x \in X \backslash A$ se verifican*

$$\rho_{(x)}(Z \backslash \pi^{-1}(\{x\})) = \rho_{(x)}(Z \backslash (\{x\} \times \mathbb{R})) = 0.$$

*(En particular, si $E \in \mathcal{B}(Z), \forall x \in X \backslash A$, $\rho_{(x)}(E) = \rho_{(x)}(E \cap \pi^{-1}(\{x\}))$ ).*

*iii) Para cada función medible $f : Z \longrightarrow [0, +\infty]$, se tiene*

$$\int_Z f(z) d\rho(z) = \int_{x \in X} \left( \int_{\pi^{-1}(\{x\})} f(x, y) d\rho_{(x)}(y) \right) d\rho_X(x).$$

La distribución de probabilidad $\rho_{(x)}$ es la distribución condicionada $\rho(\cdot|x)$ y se suele usar la notación:

$$d\rho(y|x) = \rho_{(x)}(y).$$

Un recuento de las propiedades del Teorema de Desintegración de la medida y su relación con la probabilidad condicionada puede encontrarse en [**CP, 97**].

DEFINICIÓN 29. *Con las notaciones e hipótesis del Teorema de Desintegración, llamaremos* función de regresión *a la función $f_\rho : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ definida mediante:*

$$x \longmapsto f_\rho(x) = \int_{\mathbb{R}} y d\rho(y|x).$$

Hipótesis de la Teoría del Aprendizaje: Supondremos que la función de regresión $f_\rho$ está bien definida y está acotada para la norma uniforme sobre $X$, esto es, $\exists M_\rho \in \mathbb{R}_+$ un número real positivo tal que

$$|f_\rho(x)| \leq M_\rho, \forall x \in X.$$

Escribiremos $||f_\rho||_\infty$ para indicar una cualquiera de esas cotas superiores. Notemos que si $(X, d)$ es compacto y separable, se tendrá

$$||f_\rho||_\infty = \sup\{|f_\rho(x)| : x \in X\}.$$

Pero, para nuestros propósitos, nos sirve cualquier cota superior.
Analicemos un poco las propiedades más inmediatas de $f_\rho$ tal y como se usan implicitamente en [**CS, 02**]. Fijado $x \in X$, consideremos la siguiente función, dada por la traslación por $-f_\rho(x)$:

$$\begin{aligned} \Psi_x : \mathbb{R} &\longrightarrow \mathbb{R} \\ y &\longmapsto y - f_\rho(x) \end{aligned}$$

PROPOSICIÓN 3.2.2. *Con las notaciones anteriores, se tiene:*

*i) La esperanza de $\Psi_x$ con respecto a la distribución $\rho(\cdot|x)$ del Teorema de Desintegración es 0, es decir,*

$$\int_{\mathbb{R}} \Psi_x(y) d\rho(y|x) = 0.$$

*ii) Sea $\sigma^2(x)$ la varianza de $\Psi_x$ con respecto a la distribución $\rho(\cdot|x)$ del Teorema de Desintegración, es decir,*

$$\sigma^2(x) = \int_{\mathbb{R}} (y - f_\rho(x))^2 d\rho(y|x).$$

*Entonces, la esperanza de $\sigma^2(x)$ con respecto a la distribución marginal $\rho_X$ es el error cuadrático medio de $f_\rho$, es decir:*

$$\mathcal{E}(f_\rho) = \sigma_\rho^2 = E_{f_\rho}[\sigma^2] = \int_X \sigma^2 d\rho_X.$$

DEMOSTRACIÓN. i) Es obvio por la definición de $f_\rho$.

$$E[\Psi_x] = \int_{\mathbb{R}} (y - f_\rho(x)) d\rho(y|x) = \int_{\mathbb{R}} y d\rho(y|x) - \int_{\mathbb{R}} f_\rho(x) d\rho(y|x) = f_\rho(x) - f_\rho(x) = 0.$$

ii) Es obvio por el Teorema de Desintegración de la Medida:

$$\sigma_\rho^2 = \int_X \sigma^2(x) d\rho_X = \int_{x\in X} \left( \int_{\pi^{-1}(\{x\})} (y - f_\rho(x))^2 d\rho(y|x) \right) d\rho_X(x) =$$

$$= \int_Z (y - f_\rho(x))^2 d\rho(x, y) = \mathcal{E}(f_\rho).$$

□

La función de regresión $f_\rho$ minimiza el error cuadrático medio, con respecto a $\rho$, como consecuencia de la siguiente proposición. Retomemos el espacio de Hilbert $L^2(X)$ definido en el Ejemplo 1.1.12 del Capítulo 1. La Hipótesis del Aprendizaje de [**CS, 02**] citada anteriormente implica que $f_\rho \in L^2(X)$:

$$||f_\rho||_{L^2} = \left(\int_X |f_\rho(x)|^2 d\rho_X(x)\right)^{\frac{1}{2}} \leq \left(M_\rho^2 \int_X d\rho_X(x)\right)^{\frac{1}{2}} = M_\rho,$$

para cualquier $M_\rho$ tal que $||f_\rho||_\infty \leq M_\rho$. La siguiente Proposición es crucial para entender el papel de $f_\rho$ en el análisis del error cuadrático medio.

PROPOSICIÓN 3.2.3. *Con las notaciones precedentes, si $f \in L^2(X)$ es una función medible, entonces*

$$\mathcal{E}(f) = ||f - f_\rho||^2_{L^2(X)} + \sigma_\rho^2.$$

*En particular, $f_\rho$ minimiza $\mathcal{E}$ sobre $L^2(X)$, esto es,*

$$\mathcal{E}(f_\rho) = \min\{\mathcal{E}(f) : f \in L^2(X)\}.$$

DEMOSTRACIÓN. Es de mera comprobación siguiendo [**CS, 02**]:

$$\begin{aligned}(3.2.1) \quad \mathcal{E}(f) &= \int_Z (y - f(x))^2 d\rho = \int_Z (y - f_\rho(x) + f_\rho(x) - f(x))^2 d\rho = \\ &= \int_Z (y - f_\rho(x))^2 d\rho + \int_Z (f(x) - f_\rho(x))^2 d\rho + 2\int_Z (y - f_\rho(x))(f_\rho(x) - f(x)) d\rho.\end{aligned}$$

Analicemos los tres términos de la expresión 3.2.1:

- En primer lugar, tenemos
$$\int_Z (y - f_\rho(x))^2 d\rho = \mathcal{E}(f_\rho) = \sigma_\rho^2.$$
- En segundo lugar, tenemos
$$\int_Z (f(x) - f_\rho(x))^2 d\rho = ||f - f_\rho||^2_{L^2(X)}.$$
- Por último, tenemos que el Teorema de Desintegración de la Medida:
$$\begin{aligned} I &= \int_Z (y - f_\rho(x))(f_\rho(x) - f(x)) d\rho(x, y) = \\ &= \int_{x\in X} \left(\int_{\pi^{-1}(\{x\})} (y - f_\rho(x))(f_\rho(x) - f(x)) d\rho(y|x)\right) d\rho_X(x) = \\ &= \int_{x\in X} (f_\rho(x) - f(x)) \left(\int_{\pi^{-1}(\{x\})} (y - f_\rho(x)) d\rho(y|x)\right) d\rho_X(x), \end{aligned}$$
porque $f_\rho(x) - f(x)$ no depende de $y$. Retomando $\Psi_x(y) = y - f_\rho(x)$ introducido en el Lema anterior y la propiedad $i)$
$$E[\Psi_x] = \int_{\mathbb{R}} \Psi_x(y) d\rho(y|x) = 0,$$
concluimos
$$I = \int_{x\in X} (f_\rho(x) - f(x)) E[\Psi_x] d\rho_X(x) = 0.$$

Con ello probamos la Proposición y claramente

$$\mathcal{E}(f_\rho) = \min\{\mathcal{E}(f) : f \in L^2(X)\}.$$

□

**3.2.3. Aproximaciones de la Función de Regresión por RKHS.** En Teoría del Aprendizaje se da por asumida la imposibilidad de calcular $f_\rho$ por lo que, bajo la Hipótesis del Aprendizaje, se reemplaza el cálculo de $f_\rho$ por el cálculo de una "decente aproximación". La Proposición de la Subsección precedente permite llegar a la siguiente conclusión:

COROLARIO 3.2.4. *Sea $(X, d)$ un espacio métrico separable, $Z = X \times \mathbb{R}$, $\rho$ una distribución de probabilidad definida en $\mathcal{B}(Z)$ y sea $M_\rho \in \mathbb{R}_+$ tal que $||f_\rho||_\infty \leq M_\rho$. Sea $H \subseteq L^2(X)$ un espacio de Hilbert, subespacio de $L^2(X)$. Para cada $f \in H$ son equivalentes:*

*i) La función $f$ minimiza la distancia a $f_\rho$, es decir*

$$||f - f_\rho||^2_{L^2(X)} = \min\{||g - f||^2_{L^2(X)} : g \in H\}.$$

*ii) La función $f$ minimiza $\mathcal{E}$ sobre $H$*

$$\mathcal{E}(f) = \min\{\mathcal{E}(g) : g \in H\}.$$

DEMOSTRACIÓN. Es inmediato a partir de la Proposición 3.2.3

$$\mathcal{E}(g) = ||g - f_\rho||^2_{L^2(X)} + \sigma^2_\rho.$$

Por tanto, minimizar $\mathcal{E}$ sobre $H$ es equivalente a minimizar la distancia a $f_\rho$. $\square$

A partir de aquí, el propósito de la Teoría del Aprendizaje se bifurca en dos tareas distintas:

i) **Problema de Aproximación**. Con las hipótesis precedentes, dado $\delta > 0$, hallar $H \subseteq L^2(X)$ un espacio de Hilbert tal que

$$dist_{L^2(X)}(f_\rho, H) < \delta.$$

ii) **Problema de minimizar el error en H**. Con las hipótesis precedentes, dado $\varepsilon > 0$, hallar $f_H \in H$ tal que

$$\mathcal{E}(f_H) = \min\{\mathcal{E}(g) : g \in H\}.$$

En [**CS, 02**] y en sus referencias se exhiben diversos espacios de Hilbert $H$ tales que $d_{L^2(X)}(f_\rho, H) < \delta$. No es el propósito de este trabajo enfrentar la búsqueda de tales espacios de Hilbert. Por tanto, introducimos la siguiente noción:

DEFINICIÓN 30. *Sea $(X, d)$ un espacio métrico separable, $Z = X \times \mathbb{R}$ y $\rho$ una distribución de probabilidad sobre $\mathcal{B}(Z)$. Supongamos que existe $M_\rho \in \mathbb{R}_+$ tal que $||f_\rho||_\infty \leq M_\rho$. Un espacio de Hilbert $H \subseteq L^2(X)$ se denomina $\delta$-aproximación de la regresión de $\rho$, con $\delta > 0$ un número real positivo si $\exists f \in H$ tal que*

$$||f_\rho - f||^2_{L^2(X)} < \delta.$$

PROPOSICIÓN 3.2.5. *Con las notaciones de la Definición precedente, sea $\delta > 0$ un número real positivo y $H$ un espacio de Hilbert que es una $\delta$-aproximación de la regresión de $\rho$. Sea $f_H \in H$ un elemento que minimiza $\mathcal{E}$ sobre $H$. Entonces, se tiene:*

$$||f_H - f_\rho||^2_{L^2(X)} < \delta$$

$$|\mathcal{E}(f_H) - \mathcal{E}(f_\rho)| < \delta.$$

DEMOSTRACIÓN. Es obvio por lo discutido anteriormente. Sea $f \in H$ una función tal que $||f_\rho - f||^2 < \delta$. Entonces:

$$||f_H - f_\rho||^2_{L^2(X)} \leq ||f - f_\rho||^2_{L^2(X)} < \delta.$$

Y por ello:

$$\mathcal{E}(f_H) = ||f_H - f_\rho||^2_{L^2(X)} + \mathcal{E}(f_\rho).$$

$\square$

**3.2.4. Caracterizando el Error de Aproximación en Espacios de Hilbert con Núcleo de Mercer.** Con las notaciones de la subsección anterior, suponemos $(X, d)$ un espacio métrico compacto, $Z = X \times \mathbb{R}$ dotado con una distribución de probabilidad $\rho$ y $f_\rho : X \longrightarrow \mathbb{R}$ la función de regresión asociada. El siguiente resultado resume una caracterización de la existencia de aproximación en espacios de Hilbert $H_K$ definidos por un núcleo de Mercer $K : X \times X \longrightarrow \mathbb{R}$.

COROLARIO 3.2.6. *Con las notaciones precedentes, supongamos que la probabilidad $\rho_X$ es una medida de Borel no degenerada (ie, para cada abierto no vacío $U \subseteq X, \rho_x(U) > 0$). Supongamos que, conforma a la hipótesis de admisibilidad del aprendizaje existe $M_\rho \in \mathbb{N}$ tal que la función de regresión $f_\rho$ satisface:*

$$|f_\rho(X)| \leq M_\rho, \qquad \forall x \in X.$$

*Sea $\nu = \rho_X$ y sea $L^2_\nu(X)$ el espacio de Hilbert definido en el Ejemplo 1.1.12. Sea $\theta > 0$ un número real positivo y sea $r = \frac{\theta}{2+\theta} < 1$. Sea $K : X \times X \longrightarrow \mathbb{R}$ un núcleo de Mercer sobre $X$ y $H_K$ el espacio de Hilbert asociado. Para cada $R > 0$, sea $B_K(0, R$ la bola cerrada de centro 0 y radio $R$. Sea*

$$||f_\rho||_r := \sup\{\frac{\mathbb{K}(f_\rho, t)}{t^r} : t > 0\},$$

*donde*

$$\mathbb{K}(f_\rho, t) = \inf\{||f_\rho - f||_{L^2_\nu(X)} + t||f||_K : f \in H_K\}.$$

*Entonces, si $||f_\rho||_r < \infty$ se tiene:*

$$d(f_\rho, B_K(0, R))_{L^2_\nu(X)} \leq ||f_\rho||_r^{\frac{\theta}{r}} R^{-\theta}.$$

*Más aún, en ese caso, tomando $\delta := ||f_\rho||^{\frac{\theta}{r}}$, $H_K$ contiene una $\delta$-aproximación de $f_\rho$.*

DEMOSTRACIÓN. Como la distribución $\nu = \rho_X$ es no degenerada, el Teorema de Mercer nos garantiza que existe un sistema ortonormal $\{\phi_k : k \geq 1\}$ de $L^2_\nu(X)$ formado por vectores propios de $L_K$. Además, la familia de valores propios $\{\lambda_k : k \geq 1\}$ asociada verifica:

i) $\sup\{|\lambda_k| : k \geq 1\} = ||L_K|| \leq \sqrt{\nu(X)} C_K^2$.
ii) La familia $\{\sqrt{\lambda_k}\phi_k : \lambda_k > 0\}$ es una base ortonormal del espacio de Hilbert $H_K$ asociado al núcleo de Mercer $K$.

Sea $f \in H_K$ un elemento cualquiera. Entonces $f$ admite una expansión en la base ortonormal de $H_K$:

$$f = \sum_{\lambda_k > 0} a_k \sqrt{\lambda_k}\phi_k,$$

donde $a_k = \langle f, \sqrt{\lambda_k}\phi_k\rangle_K$, Ahora, como $\{\phi_k : k \geq 1\}$ es un sistema ortonormal en $L^2_\nu(X)$ tendremos

$$\begin{aligned} ||f||^2_{L^2_\nu(X)} &= \langle \sum_{\lambda_k>0} a_k \sqrt{\lambda_k}\phi_k, \sum_{\lambda_k>0} a_k \sqrt{\lambda_k}\phi_k\rangle_{L^2(X)} = \\ &= \sum_{\lambda_k>0} a_k^2 |\lambda_k| \leq \sup\{|\lambda_k| : \lambda_k > 0\}(\sum_{\lambda_k>0} a_k^2) \leq \\ &\leq \sqrt{\nu(X)} C_K^2 (\sum_{\lambda_k>0} a_k^2). \end{aligned}$$

Como $\{\sqrt{\lambda_k}\phi_k : \lambda_k > 0\}$ es una base ortonormal de $H_K$

$$||f||^2_K = \langle \sum_{\lambda_k>0} a_k \sqrt{\lambda_k}\phi_k, \sum_{\lambda_k>0} a_k \sqrt{\lambda_k}\phi_k\rangle_K = \sum_{\lambda_k>0} a_k^2.$$

Por tanto, habremos concluido

$$||f||_{L^2_\nu(X)} \leq C_K ||f||_K.$$

Por el Teorema de Aproximación, concluiremos que si

$$||f_\rho||_r = \sup_{t>0}\{\frac{\mathbb{K}(a, t}{t^r}\} < \infty,$$

entonces

$$d(f_\rho, R) \leq ||f_\rho||_r^{\frac{2}{1-r}} R^{\frac{-2r}{1-r}}.$$

Tomando $r = \frac{\theta}{2+\theta}$ tenemos la equivalencia citada. $\square$

**3.2.5. Error Empírico y Error Cuadrático Medio.** La siguiente etapa de la Teoría del Aprendizaje pretende inferir la función de regresión $f_\rho$ a partir de funciones que optimizen el error empírico cuadrático en promedio a partir de una observación empírica $\underline{z} \in Z^n$. La idea se retrotrae a los trabajos de V. N. Vapnik and A. Ya. Chervonenkis ([**CV, 71**]) y a métodos combinatorios como el método descrito por D. Pollard en [**Po, 84**]. Para formalizarlo, retomaremos nuestra definición de $\mathcal{E}_z$ para $z = ((x_1, y_1), ..., (x_n, y_n)) \in Z^n$, con $Z = X \times Y$. Supongamos $(X, d)$ métrico separable, $Z = X \times Y$, $\rho$ una distribución de probabilidad en $Z$. Supongamos que $f_\rho : X \longrightarrow \mathbb{R}$ es la función de regresión y $M_\rho \in \mathbb{R}_+$ tal que $||f_\rho||_\infty \leq M_\rho$. Sea $H$ un espacio de Hilbert y definamos

$$\begin{aligned} \mathcal{E}_z : H &\longrightarrow \mathbb{R}_+ \\ f &\longmapsto \frac{1}{n}\sum_{i=1}^{n} |y_i - f(x_i)|^2 \end{aligned}$$

Supongamos que $H$ es una $\delta$-aproximación de la regresión de $\rho$, supongamos $f_H \in H$ tal que $||f_H - f_\rho||_{L^2(X)} < \delta$ y $f_H$ minimiza $\mathcal{E}$ sobre $H$. Sea $B = B_H(0, R) = \{g \in H : ||g||_H \leq R\}$ la bola cerrada de centro 0 y radio $R$ en $H$. Denotemos por $f_{B,z} \in B$ un elemento que minimiza $\mathcal{E}_z$ sobre $B$, es decir,

$$\mathcal{E}_z(f_{B,z}) = \min\{\mathcal{E}_z(g) : g \in B_H(0, R)\},$$

y definimos para $\varepsilon > 0, \varepsilon \in \mathbb{R}$

$$H_\rho^{(n)}(B, \varepsilon) = \{z \in Z^n : \sup_{f \in B} |\mathcal{E}_z(f) - \mathcal{E}(f)| \leq \varepsilon\}.$$

PROPOSICIÓN 3.2.7. *Con las hipótesis y notaciones precedentes, si* $0 < \delta < 1$ *y si* $R \geq M_\rho + 1$, *se tiene para* $B = B_H(0, R)$ *y* $z \in H_g^{(n)}(B, \varepsilon)$*:*

*i)* $f_H \in B$.
*ii)* $||f_{B,z} - f_\rho||^2_{L^2(X)} \leq 2\varepsilon + \delta$.
*iii)* $|\mathcal{E}(f_{B,z}) - \mathcal{E}(f_\rho)| \leq 2\varepsilon + \delta$.

DEMOSTRACIÓN. i) Tenemos que $H \subseteq L^2(X)$ es un espacio de Hilbert, subespacio de $L^2(X)$. Entonces

$$||f_H||_H = ||f_H||_{L^2(X)} \leq ||f_H - f_\rho||_{L^2(X)} + ||f_\rho||_{L^2(X)},$$

pero

$$||f_\rho||_{L^2(X)} \leq ||f||_\infty \leq M_\rho,$$

luego

$$||f_H||_H \leq \delta + M_\rho < M_\rho + 1 \leq R.$$

Por ello $f_H \in B$.

iii)

$$||f_{B,z} - f_\rho||_{L^2(X)} \leq ||f_{B,z} - f_H||_{L^2(X)} + ||f_H - f_\rho||_{L^2(X)}.$$

Como $||f_H - f_\rho||_{L^2(X)} \leq \delta$, solo nos queda por analizar el primer sumando. Observemos que

- Como $f_H \in B$
$$0 \leq \mathcal{E}_z(f_H) - \mathcal{E}_z(f_{B,z}).$$
- Como $f_H$ minimiza $\mathcal{E}$ sobre $H$ tendremos:
$$0 \leq \mathcal{E}(f_{B,z}) - \mathcal{E}(f_H).$$

Entonces, tendremos

$$\begin{aligned} |\mathcal{E}(f_{B,z}) - \mathcal{E}(f_H)| &= \mathcal{E}(f_{B,z}) - \mathcal{E}(f_H) = \\ &= \mathcal{E}(f_{B,z}) - \mathcal{E}_z(f_{B,z}) + \mathcal{E}_z(f_{B,z}) - \mathcal{E}(f_H). \end{aligned}$$

Como $\mathcal{E}_z(f_{B,z}) \leq \mathcal{E}_z(f_H)$ quedará

$$\begin{aligned} |\mathcal{E}(f_{B,z}) - \mathcal{E}(f_H)| &\leq \mathcal{E}(f_{B,z}) - \mathcal{E}_z(f_{B,z}) + \mathcal{E}_z(f_H) - \mathcal{E}(f_H) \leq \\ &\leq |\mathcal{E}(f_{B,z}) - \mathcal{E}_z(f_{B,z}) + \mathcal{E}_z(f_H) - \mathcal{E}(f_H)| \leq \\ &\leq |\mathcal{E}(f_{B,z}) - \mathcal{E}_z(f_{B,z})| + |\mathcal{E}_z(f_H) - \mathcal{E}(f_H)| \leq 2\varepsilon, \end{aligned}$$

porque $z \in H_g(B, \varepsilon)$. En conclusión

$$|\mathcal{E}(f_\rho) - \mathcal{E}(f_{B,z})| \leq 2\varepsilon + \delta.$$

Pero, además,

$$|\mathcal{E}(f_\rho) - \mathcal{E}(f_{B,z})| = ||f_\rho - f_{B,z}||^2_{L^2(X)},$$

por lo que hemos probado también el apartado ii).

□

**3.2.6. El Método Combinatorio de Pollard en Versión** [**PS, 20**]**.** Dado un espacio de Hilbert $H$ con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$, siendo $(X, d)$ un espacio métrico separable, dada $B = B_H(0, R)$ una bola en $H$ de centro 0 y radio $R$ y dado $\varepsilon > 0$, hemos introducido en la Subsección precedente la clase $H_g(B, \varepsilon)$. Este conjunto viene dado por

$$H_\rho^{(n)}(B, \varepsilon) = \{z \in Z^n : \sup_{f \in B} |\mathcal{E}_z(f) - \mathcal{E}(f)| \leq \varepsilon\}. \tag{3.2.2}$$

La principal contribución de Vapnik, Chervonenkis o Pollard (así como los Teoremas $B$ o $C$ de [**CS, 02**]) consiste en probar que $H_\rho(B, \varepsilon)$ no solamente es no vacío sino que es altamente probable en $Z^n$ con la probabilidad producto castesiano $\rho^n$, inducida en $Z^n$ por la probabilidad $\rho$ en $Z$. Esta serie de resultados permiten estimar cotas inferiores de $\rho^n$ $(H_\rho^{(n)}(B, \varepsilon))$ en función de los diversos elementos involucrados. El resultado que expondremos a continuación es la interpretación en [**PS, 20**] del método combinatorio de Pollard. Se trata de un ingrediente instrumental en el contexto de este trabajo por lo que no entraremos en la prueba.

Teorema 3.2.8. *Sea $(X, d)$ un espacio métrico separable, $I \subseteq \mathbb{R}$ un intervalo compacto, $Z = X \times I$ y $\rho$ una distribución de probabilidad sobre $Z$. Sea $f_\rho$ la función de regresión asociada y supongamos $M_\rho$ una cota superior de $||f_\rho||_\infty$.*
*Sea $H_K$ un espacio de Hilbert de funciones a valores reales sobre $X$, con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$. Sea $R := M_\rho + 1$ y $0 < \delta < 1$, de tal modo que $H_K$ contiene una $\delta$-aproximación $f_H$ de $f_\rho$. Sea $B \in \mathbb{R}$ un número real positivo, $n \in \mathbb{N}$ un entero y $\varepsilon > 0$, de tal modo que se verifica:*

- $\frac{log(2)B^2}{2n} < \varepsilon$.
- *$\forall f \in B_H(0, R)$, $\forall y \in I$ se tiene*

$$|y - f(x)|^2 \leq \frac{B}{2}.$$

*Entonces, si $H_\rho^{(n)}(B(0, R), \varepsilon)$ es el conjunto definido en 3.2.2 se tiene:*

$$\rho^n(H_\rho^{(n)}(B(0, R), \varepsilon)) \geq 1 - 2\left(\frac{\varepsilon}{16}\right)^{-n} (n(M_\rho + 1)E(R) + \frac{\varepsilon}{18})^n e^{\frac{-m\varepsilon^2}{64B^2}},$$

*donde $E(R)$ es la esperanza siguiente:*

$$E(R) = E_{B_H(0, M_\rho + 1)}[||G_x^{\frac{1}{2}}||_F].$$

**3.2.7. La Prueba del Teorema Fundamental del Capítulo.** Consideremos $I \subseteq \mathbb{R}$ un intervalo compacto, $(X, d)$ un espacio métrico separable, $Z = X \times I$ y $\rho$ una distribución de probabilidad sobre $X \times I$. Sea $f_\rho$ la función de regresión.
Sea $H \subseteq \mathbb{R}^X$ un espacio de Hilbert con núcleo reproductor $K : X \times X \longrightarrow \mathbb{R}$ que contiene una $\delta$-aproximación $f_H \in H$ de $f_\rho$, con $0 < \delta < 1$. Sean $R \in \mathbb{R}_+$ y $\varepsilon > 0$ dos números reales positivos. Definamos

$$H_\rho^{(n)}(B_H(0, R), \varepsilon) = \{z \in Z^n : \sup_{f \in B_H(0,R)} |\mathcal{E}_z(f) - \mathcal{E}(f)| \leq \varepsilon\}.$$

Sea $z \in H_\rho(B_H(0, R), \varepsilon)$, $z = ((x_1, y_1), ..., (x_n, y_n))$. Supongamos que

i) $||f_\rho||_\infty \leq R - 1$.
ii) Si $G_x = (K(x_i, x_j))_{1 \leq i,j \leq n}$ es la matriz de Gram en el punto $\underline{x} = (x_1, ..., x_n)$, por hipótesis tenemos

$$\frac{\mu(G_x)^{\frac{1}{2}}}{||G_x^{\frac{1}{2}}||_F} ||y||_2 \leq R.$$

Entonces, se tiene:

- Sea $f_{B_H(0,R),z} \in B_H(0,R)$ un elemento que minimiza el error empírico en $B_H(0,R)$

$$\mathcal{E}_z(f_{B_H(0,R),z}) = \min\{\mathcal{E}_z(g) : g \in B_{H(0,R),z}\},$$

y sea $f_z \in W_z = Span\{K_{x_1}, ..., K_{x_n}\}$ el elemento que minimiza $\mathcal{E}_z$ sobre $H$. Entonces, por el Teorema 3.1.1.

$$||f_z||_H \leq \frac{\mu(G_x)^{\frac{1}{2}}}{||G_x{}^{\frac{1}{2}}||_F}||y||_2 \leq R.$$

Luego $f_z \in B_H(0,R)$ y se tendrá $f_{B_H(0,R),z} = f_z$.

Por otro lado, sabemos que $0 < \delta < 1$, luego si $f_H \in H$ es una $\delta$-aproximación de $f_\rho$ sabemos que

$$||f_H||_H \leq ||f_\rho||_{L^2(X)} \leq ||f_\rho||_\infty + 1 \leq R.$$

Por tanto, $f_H \in B(0,R)$ con las hipótesis del Teorema. Además

$$||f_H - f_\rho||^2_{L^2(X)} < \delta$$

$$|\mathcal{E}(f_H) - \mathcal{E}(f_\rho)| < \delta.$$

Consideremos ahora

- $0 \leq \mathcal{E}(f_z) - \mathcal{E}(f_H)$, porque $\mathcal{E}(f_H)$ es mínimo de $\mathcal{E}$ sobre $H$.
- $0 \leq \mathcal{E}_z(f_H) - \mathcal{E}_z(f_z)$, porque $\mathcal{E}_z(f_z)$ es mínimo de $\mathcal{E}_z$ sobre $H$.

Entonces

$$\begin{aligned}|\mathcal{E}(f_z) - \mathcal{E}(f_H)| &= \mathcal{E}(f_z) - \mathcal{E}(f_H) = \\ &= \mathcal{E}(f_z) - \mathcal{E}_z(f_z) + \mathcal{E}_z(f_z) - \mathcal{E}(f_H) \leq \\ &\leq \mathcal{E}(f_z) - \mathcal{E}_z(f_z) + \mathcal{E}_z(f_H) - \mathcal{E}(f_H)\end{aligned}$$

porque $\mathcal{E}_z(f_z) \leq \mathcal{E}_z(f_H)$.

Entonces

$$|\mathcal{E}(f_z) - \mathcal{E}(f_H)| \leq |\mathcal{E}(f_z) - \mathcal{E}_z(f_z)| + |\mathcal{E}_z(f_H) - \mathcal{E}(f_H)|.$$

Pero como $z \in H_\rho^{(n)}(B_H(0,R), \varepsilon)$ y $f_z, f_H \in B_H(0,R)$, concluimos que

$$|\mathcal{E}(f_z) - \mathcal{E}_z(f_z)| \leq \varepsilon$$

$$|\mathcal{E}(f_H) - \mathcal{E}_z(f_H)| \leq \varepsilon.$$

Finalmente,

$$|\mathcal{E}(f_z) - \mathcal{E}(f_\rho)| \leq |\mathcal{E}(f_z) - \mathcal{E}(f_H)| + |\mathcal{E}(f_H) - \mathcal{E}(f_\rho)| \leq 2\varepsilon + \delta.$$

Como $|\mathcal{E}(f_z) - \mathcal{E}(f_\rho)| = ||f_z - f_\rho||^2_{L^2(X)}$, el Teorema se sigue.

APÉNDICE A

# Resultados Adicionales

## A.1. Pruebas de Algunos Resultados Elementales sobre Espacios de Hilbert

Comenzamos los resultados de este Apéndice con una prueba de la clásica *desigualdad de Cauchy-Schwarz.*

TEOREMA A.1.1 (Cauchy-Schwarz). *Sea* $(H, \langle\cdot,\cdot\rangle)$ *un espacio preHilbert sobre* $\mathbb{K}$. *Entonces, para cada* $x, y \in H$ *se cumple*

$$|\langle x, y\rangle| \leq \sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}.$$

DEMOSTRACIÓN. Por la no negatividad del producto interno, tenemos que

$$\langle \lambda x - \mu y, \lambda x - \mu y\rangle \geq 0$$

para todo $x, y \in H$ y $\lambda, \mu \in \mathbb{C}$. Desarrollando el producto interno obtenemos que

$$\langle \lambda x - \mu y, \lambda x - \mu y\rangle = \langle \lambda x, \lambda x\rangle + \langle \lambda x, -\mu y\rangle + \langle -\mu y, \lambda x\rangle + \langle -\mu y, -\mu y\rangle =$$
$$|\lambda|^2\langle x, x\rangle - \lambda\overline{\mu}\langle x, y\rangle - \overline{\lambda}\mu\langle y, x\rangle + |\mu|^2\langle y, y\rangle \geq 0,$$

por lo que se cumple

$$|\lambda|^2\langle x, x\rangle + |\mu|^2\langle y, y\rangle \geq \lambda\overline{\mu}\langle x, y\rangle + \overline{\lambda}\mu\langle y, x\rangle.$$

Como $\langle x, y\rangle \in \mathbb{C}$, podremos escribirlo como

$$\langle x, y\rangle = re^{i\theta},$$

con $r = |\langle x, y\rangle|$ y $\theta = arg(\langle x, y\rangle)$. Si tomamos $\lambda = \sqrt{\langle y, y\rangle}e^{-i\theta}$, $\mu = \sqrt{\langle x, x\rangle}$ y desarrollamos ambos lados de la desigualdad tenemos

- En el lado izquierdo
$$|\lambda|^2\langle x, x\rangle + |\mu|^2\langle y, y\rangle = \langle y, y\rangle\langle x, x\rangle + \langle x, x\rangle\langle y, y\rangle = 2\langle x, x\rangle\langle y, y\rangle.$$
teniendo en cuenta que $|\lambda| = \sqrt{\langle y, y\rangle}$.
- En el lado derecho

$$\lambda\overline{\mu}\langle x, y\rangle + \overline{\lambda}\mu\langle y, x\rangle = \sqrt{\langle y, y\rangle}e^{-i\theta}\sqrt{\langle x, x\rangle}\ |\langle x, y\rangle|e^{i\theta} + \sqrt{\langle y, y\rangle}e^{i\theta}\sqrt{\langle x, x\rangle}\ |\langle x, y\rangle|e^{-i\theta} = 2\langle x, x\rangle\langle y, y\rangle|\langle x, y\rangle|$$

Por tanto, tenemos que

$$2\langle x, x\rangle\langle y, y\rangle \geq 2\sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}|\langle x, y\rangle|,$$

y suponiendo que $x, y \neq 0$ llegamos a que

$$\sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle} \geq |\langle x, y\rangle|,$$

tal y como queríamos probar. Si $x$ o $y$ fuesen 0, ambos lados de la desigualdad serían 0, luego el resultado sigue siendo valido. □

Por completitud, incluimos una prueba del teorema 1.1.2 del capítulo 1 que enunciamos y probamos a continuación.

TEOREMA A.1.2. *Sea* $(H, \langle\cdot,\cdot\rangle)$ *un espacio preHilbert. Definimos la siguiente función*

$$||\cdot|| : H \longrightarrow \mathbb{R}_+$$
$$x \mapsto \sqrt{\langle x, x\rangle}.$$

*Entonces,* $(H, ||\cdot||)$ *es un* $\mathbb{K}$*-espacio vectorial normado. Es decir,* $||\cdot||$ *satisface las siguientes propiedades*

*i)* $||x|| \geq 0$, $\forall\ x \in H$.
*ii)* $||x|| = 0 \iff x = 0$.
*iii)* $||\lambda x|| = |\lambda|||x||$, $\forall\ x \in H$, $\lambda \in \mathbb{K}$.

*iv)* $||x+y|| \leq ||x|| + ||y||$, $\forall\ x, y \in H$.

DEMOSTRACIÓN. Necesitamos probar las cuatro propiedades que tiene que cumplir una función norma:

- $||x|| \geq 0$
  Basta usar la definición de $||\cdot||$ para ver que $||x|| = \langle x, x\rangle \geq 0$ por las propiedades del producto interno.
- $||x|| = 0 \iff x = 0$
  Probamos fácilmente la implicación a la derecha viendo que $||x|| = \sqrt{\langle x, x\rangle} = 0 \Rightarrow ||x||^2 = \langle x, x\rangle = 0 \Rightarrow x = 0$, por ser el producto interno definido positivo. La otra implicación es trivial.
- $||\lambda x|| = |\lambda|\ ||x||$

$$||\lambda x|| = \langle \lambda x, \lambda x\rangle = \lambda\overline{\lambda}\langle x, x\rangle = |\lambda|\ ||x||.$$

- $||x+y|| \leq ||x|| + ||y||$
  Antes de probar esta propiedad, notemos que dado un $\lambda \in \mathbb{C}$, se cumple que $|Re(\lambda)| \leq \lambda$, con $Re(\lambda)$ denotando la parte real de $\lambda$. Vemos esto fácilmente considerando un $z \in \mathbb{C}$ de la forma $z = a + bi$. Entonces

$$|z| = \sqrt{a^2+b^2} \geq \sqrt{a^2} = |a|.$$

  Dicho esto, tenemos que

$$||x+y||^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = ||x||^2 + ||y||^2 + 2Re(\langle x, y\rangle) \leq$$
$$\leq ||x||^2 + ||y||^2 + 2|Re(\langle x, y\rangle)| \leq ||x||^2 + ||y||^2 + 2|\langle x, y\rangle| \leq ||x||^2 + ||y||^2 + 2||x||\ ||y|| =$$
$$= (||x|| + ||y||)^2,$$

  probando así la cuarta condición. Notemos que la última desigualdad viene dada por la *desigualdad de Cauchy-Schwarz.*

□

Probamos a continuación la caracterización de los espacios normados que son preHilbert a través de la *Ley del Paralelogramo*, que corresponde al teorema 1.1.7 del capítulo 1

TEOREMA A.1.3 (Ley del Paralelogramo). *Sea $\mathbb{K} = \mathbb{R} \vee \mathbb{C}$ y sea $(V, ||\cdot||)$ un $\mathbb{K}$-espacio vectorial normado. Entonces, la norma $||\cdot||$ sobre $V$ es la inducida por un producto interno $\langle\cdot,\cdot\rangle$ sobre $V$ y una estructura de espacio preHilbert $(V, \langle\cdot,\cdot\rangle)$ si y solamente si la norma $||\cdot||$ satisface la siguiente propiedad (conocida como* Ley del Paralelogramo*):*

$$\forall x, y \in V \qquad ||x+y||^2 + ||x-y||^2 = 2||x||^2 + 2||y||^2.$$

DEMOSTRACIÓN. Probamos el resultado para el caso $\mathbb{K} = \mathbb{C}$ demostrando ambas implicaciones

$\Longrightarrow$) Si utilizamos la norma definida en el teorema 1.1.2.

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

$\Longleftarrow$) Si una norma satisface la Ley del Paralelogramo, entonces la siguiente ecuación, conocida como Identidad de Polarización, define un producto interno

$$\langle x, y\rangle = \frac{1}{4}(||x+y||^2 - ||x-y||^2 + i||x+iy||^2 - i||x-iy||^2).$$

Lo primero que tenemos que probar es que efectivamente $\langle x, x\rangle = ||x||^2$. Tenemos que

$$\langle x, x\rangle = \frac{1}{4}(||2x||^2 + i||x+ix||^2 - i||x-ix||^2) = \frac{1}{4}(4||x||^2 + i||x||^2|1+i|^2 - i||x||^2|1-i|^2).$$

Como $|1-i| = |1+i|$, vemos fácilmente que efectivamente $\langle x, x\rangle = ||x||^2$. Por tanto, sabemos que el $\langle\cdot,\cdot\rangle$ que hemos definido cumple que es positivo y nulo solo en el 0, es decir, las propiedades *iii)* y *iv)* del producto interno. Probemos la simetría. Tenemos que

$$\langle x, y\rangle = \frac{1}{4}(||x+y||^2 - ||x-y||^2 + i||x+iy||^2 - i||x-iy||^2).$$

$$\overline{\langle y, x\rangle} = \frac{1}{4}(||y+x||^2 - ||y-x||^2 - i||y+ix||^2 + i||y-ix||^2).$$

Está claro que $||x-y|| = ||y-x||$, luego las partes reales coinciden. También se cumple que

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

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

De forma análoga, vemos que $||x+iy||^2 = ||y-ix||^2$ y por tanto la parte imaginaria también coincide. Con ello, probamos la propiedad *ii*).
Para probar la linealidad, antes debemos probar la siguiente igualdad:

$$\forall x,y,z \in V, ||x+y+z||^2 = ||x+y||^2 + ||y+z||^2 + ||x+z||^2 - ||x||^2 - ||y||^2 - ||z||^2.$$

La demostración de esta igualdad consiste en aplicar la Ley del Paralelogramo sucesivamente. Los términos subrayados son donde la aplicaremos:

$$\begin{aligned}
\underline{||x+y+z||^2} &= ||x+y||^2 + ||z||^2 - ||x+y-z||^2 + \underline{||x+y||^2 + ||z||^2} = \\
&= ||x+y||^2 + ||z||^2 - \frac{1}{2}||x+y-z||^2 + \underline{\frac{1}{2}||x+y+z||^2} = \\
&= ||x+y||^2 + ||x+z||^2 + ||z||^2 + ||y||^2 - \frac{1}{2}(\underline{||x-y+z||^2 + ||x+y-z||^2}) = \\
&= ||x+y||^2 + ||x+z||^2 - ||x||^2 - ||-y+z||^2 + \underline{||y||^2 + ||z||^2} = \\
&= ||x+y||^2 + ||x+z||^2 - ||x||^2 + \frac{1}{2}||y+z||^2 - \frac{1}{2}||-y+z||^2 = \\
&= ||x+y||^2 + ||y+z||^2 + ||x+z||^2 - ||x||^2 - ||y||^2 - ||z||^2.
\end{aligned}$$

Ahora vamos a probar que

$$\forall x,y,z \in V, \langle x+y, z\rangle = \langle x,z\rangle + \langle y,z\rangle.$$

Para ello, nos vamos a fijar en que

$$\langle x,y\rangle = \frac{1}{4}(||x+y||^2 - ||x-y||^2 + i||x+iy||^2 - i||x-iy||^2) = \frac{1}{4}\sum_{k=0}^{3} i^k ||x+i^k y||^2,$$

y en que $\sum_{k=0}^{3} i^k = 1 + i - 1 - i = 0$. Así:

$$\begin{aligned}
\langle x+y, z\rangle &= \sum_{k=0}^{3} i^k ||x+y+i^k z||^2 = \\
&= \sum_{k=0}^{3} i^k (||x+y||^2 + ||x+i^k z||^2 + ||y+i^k z||^2 - ||x||^2 - ||y||^2 - ||i^k z||^2) = \\
&= \sum_{k=0}^{3} i^k (||x+i^k z||^2 + ||y+i^k z||^2) = 4\langle x,z\rangle + 4\langle y,z\rangle.
\end{aligned}$$

Probar que $\langle x, y+z\rangle = \langle x,y\rangle + \langle x,z\rangle$ se realiza de forma análoga. Falta probar que dado un $\lambda \in \mathbb{C}$ tenemos que $\langle \lambda x, y\rangle = \lambda\langle x,y\rangle$. Una vez hayamos visto eso, probar que $\langle x, \lambda y\rangle = \overline{\lambda}\langle x,y\rangle$ será solo cuestión de aplicar la simetría del producto interno complejo. Con lo que hemos visto hasta ahora, está claro que para todo $n \in \mathbb{Z}$ se cumple que $\langle nx, y\rangle = n\langle x,y\rangle$.
Sea $p,q \in \mathbb{Z}, q \neq 0$. Tenemos que

$$q\langle \frac{p}{q}x, y\rangle = q\langle p\frac{x}{q}, y\rangle = p\langle q\frac{x}{q}, y\rangle = p\langle x,y\rangle,$$

por lo que $\langle \frac{p}{q}x, y\rangle = \frac{p}{q}\langle x,y\rangle$. Hasta ahora hemos probado la linealidad para $\lambda \in \mathbb{Q}$. Si probamos la linealidad de la unidad imaginaria $i$, sumado a lo ya visto, habremos probado la linealidad para $\mathbb{Q}+\mathbb{Q}[i]$. Efectivamente:

$$\begin{aligned}4\langle ix,y\rangle &= ||ix+y||^2+i||ix+iy||^2-||ix-y||^2-i||ix-iy||^2=\\ &=||i(x-iy)||^2+i||i(x+y)||^2-||i(x+iy)||^2-i||i(x-y)||^2=\\ &=|i|\;||x-iy||^2+i|i|\;||x+y||^2-|i|\;||x+iy||^2-i|i|\;||x-y||^2=\\ &=||x-iy||^2+i||x+y||^2-||x+iy||^2-i||x-y||^2=i(-i||x-iy||^2+||x+y||^2+i||x+iy||^2-||x-y||^2)=\\ &=4i\langle x,y\rangle\end{aligned}$$

Ahora consideremos un $c\in\mathbb{C}$ y una sucesión $(q_n)_n$ en $\mathbb{Q}+\mathbb{Q}[i]$ tal que $q_n\longrightarrow c$. Por la desigualdad de Cauchy-Schwarz tenemos que

$$|\langle q_nx,y\rangle-\langle cx,y\rangle|\leq|q_n-c|\;||x||\;||y||,$$

luego $\langle q_nx,y\rangle\longrightarrow\langle cx,y\rangle$. Pero a su vez, $\langle q_nx,y\rangle=q_n\langle x,y\rangle\longrightarrow c\langle x,y\rangle$, por lo que hemos probado que la linealidad es para todo $\lambda\in\mathbb{C}$ y finalmente que $\langle\cdot,\cdot\rangle$ que hemos definido es un producto interno. Por tanto, $V$ es un espacio preHilbert.

Para el caso real ($\mathbb{K}=\mathbb{R}$), la prueba es idéntica salvo que la forma bilineal simétrica a utilizar cuando $(V,||\cdot||)$ es un $\mathbb{R}$-espacio vectorial normado que satisface la *Identidad de Polarización* es la siguiente:

$$\langle\cdot,\cdot\rangle:V\times V\longrightarrow\mathbb{R},$$

dada por

$$\forall x,y\in V,\qquad\langle x,y\rangle:=\frac{1}{4}(||x+y||^2-||x-y||^2).$$

□

Presentamos a continuación la demostración del teorema 1.1.9.

TEOREMA A.1.4. *Si $(H,\langle\cdot,\cdot\rangle)$ es un espacio preHilbert, el producto interno es una función continua con respecto a la topología producto en $H\times H$ inducida por la topología del espacio normado $(H,||\cdot||)$.*

DEMOSTRACIÓN. Vamos a probar que si una sucesión de elementos $x_n$ converge a $x$, también converge en norma, es decir $||x_n||\longrightarrow||x||$. Tenemos por la desigualdad triangular que

$$||x||\leq||x-x_n||+||x_n||,$$

$$||x_n||\leq||x_n-x||+||x||.$$

Como $||x-x_n||=||x_n-x||$, tenemos que

$$-||x-x_n||\leq||x||-||x_n||\leq||x-x_n||,$$

luego $|\;||x||-||x_n||\;|\leq||x-x_n||$ y como $x_n\longrightarrow x$ está claro que $||x_n||\longrightarrow||x||$ cuando $n\longrightarrow\infty$. Por otro lado,

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

$$\leq||x_n||\;||y_n-y||+||x_n-x||\;||y||,$$

aplicando *Cauchy-Schwarz* en la última desigualdad. Como $||x_n||\longrightarrow||x||$ se cumple que

$$|\langle x_n,y_n\rangle-\langle x,y\rangle|\longrightarrow 0$$

si $n\longrightarrow\infty$. □

Probaremos a continuación la Proposición 1.1.11 del Capítulo 1.

PROPOSICIÓN A.1.5. *El espacio preHilbet $(\ell^2,\langle\cdot,\cdot\rangle)$ introducido en el ejemplo 1.1.8 es completo y, por tanto, un espacio de Hilbert.*

DEMOSTRACIÓN. Vamos a probar que $\ell^2$ es completo y por tanto es un espacio de Hilbert con el producto interno que definimos antes. Sea $(x_n)_n$ una sucesión de Cauchy en $\ell^2$. Vamos a probar que existe $x_0\in\ell^2$ con

$$x_0=\lim_{n\to\infty}x_n.$$

Para cada $n \in \mathbb{N}$ escribimos

$$\begin{aligned} x_1 &= (z_1^1, z_2^1, ..., z_k^1, ...) \\ x_2 &= (z_1^2, z_2^2, ..., z_k^2, ...) \\ &\vdots \\ x_n &= (z_1^n, z_2^n, ..., z_k^n, ...) \end{aligned}$$

Vamos a ver que dado $k \in \mathbb{N}$, $\lim_{n\to\infty} z_k^n$ existe en $\mathbb{C}$. Para ello basta probar que $(z_k^n)_n$ es una sucesión de Cauchy en $\mathbb{C}$. Sea $\varepsilon > 0$. Vamos a probar que existe $n_0 \in \mathbb{N}$ tal que para todo $n, m \geq n_0$ se cumple

$$|z_k^n - z_k^m| < \varepsilon.$$

Sabemos que $(x_n)_n$ es una sucesión de Cauchy, luego para el mismo $\varepsilon$ existe un $n_1 \in \mathbb{N}$ tal que

$$||x_n - x_m||^2 < \varepsilon,$$

para $n, m \geq n_1$. Es decir

$$\sum_{i=1}^{\infty} |z_i^n - z_i^m|^2 < \varepsilon.$$

En concreto, si tomamos $n_0 = n_1$ tenemos que

$$|z_k^n - z_k^m|^2 \leq \sum_{i=1}^{\infty} |z_i^n - z_i^m|^2 < \varepsilon,$$

probando así que $(z_k^n)_n$ es de Cauchy. Si llamamos $z_k^0$ a $\lim_{n\to\infty} z_k^n$, tenemos que

$$x_0 := (z_1^0, z_2^0, ..., z_k^0, ...).$$

Falta probar que $x_0 \in \ell^2$ y que efectivamente $x_0 = \lim_{n\to\infty} x_n$.
Sea $\varepsilon > 0$. Sabemos que $(x_n)_n$ es de Cauchy, luego existe un $n_0 \in \mathbb{N}$ tal que si $n, m \geq n_0$ entonces

$$||x_n - x_m||^2 < \varepsilon \Longrightarrow \sum_{k=1}^{\infty} |z_k^n - z_k^m|^2 < \varepsilon.$$

Si fijamos $l \in \mathbb{N}$,

$$\sum_{k=1}^{l} |z_k^n - z_k^m|^2 < \varepsilon$$

y haciendo tender $n \longrightarrow \infty$ tenemos que

$$\sum_{k=1}^{l} |z_k^0 - z_k^m|^2 \leq \varepsilon$$

para todo $l \in \mathbb{N}$. La sucesión de sumas parciales es creciente y está acotada por $\varepsilon$, luego su lmite

$$\sum_{k=1}^{\infty} |z_k^0 - z_k^m|^2 \tag{A.1.1}$$

está acotado por $\varepsilon$. La ecuación A.1.1 se cumple para todo $m \geq n_0$, por lo que para $m = n_0$

$$\sum_{k=1}^{\infty} |z_k^0 - z_k^{n_0}|^2 \leq \varepsilon.$$

Si definimos

$$y := (z_1^0 - z_1^{n_0}, z_2^0 - z_2^{n_0}, ..., z_k^0 - z_k^{n_0}, ...),$$

entonces por la ecuación anterior $y \in \ell^2$. Como tanto $y$ como $x_{n_0}$ pertenecen a $\ell^2$, entonces $y + x_{n_0} = x_0 \in \ell^2$.
Para probar que $x_0 = \lim_{n\to\infty} x_n$ basta ver que la ecuación A.1.1 la podemos escribir como

$$||x_0 - x_m||^2 \leq \varepsilon,$$

para todo $m \geq n_0$, que es la definición de

$$x_0 = \lim_{m \to \infty} x_m.$$

$\square$

A continuación, enunciamos y probamos la existencia de completado de un espacio de Hilbert citado como Teorema 1.1.13 del Capítulo 1.

TEOREMA A.1.6. *Sea $(H, \langle \cdot, \cdot \rangle)$ un espacio preHilbert sobre $\mathbb{K}$. Entonces, existe un espacio de Hilbert $(H_0, \langle \cdot, \cdot \rangle_{H_0})$ tal que existe una aplicación lineal continua $I : H \longrightarrow H_0$ verificando:*

*i) $I(H)$ es denso en $H_0$ para la topología inducida por la norma $|| \cdot ||_{H_0}$.*

*ii) $I$ es una isometría, es decir,*

$$\langle I(x), I(y) \rangle_{H_0} = \langle x, y \rangle_H, \ \forall \ x, y \in H.$$

*Más aún, $H_0$ es único con estas propiedades salvo isomorfismo lineal isométrico. Es decir, dados $(H', \langle \cdot, \cdot \rangle_{H'})$ un espacio de Hilbert y una aplicación lineal $L : H \longrightarrow H'$ que satisface las propiedades i) y ii), entonces existe una isometría lineal $\varphi : H_0 \longrightarrow H'$, que además cumple que $\varphi |_{I(H)}: I(H) \longrightarrow J(H)$ es isomorfismo.*

DEMOSTRACIÓN. Comencemos defininiendo el conjunto

$$H' = \{(x_n)_n : (x_n)_n \text{ es una sucesión de Cauchy en } H\}$$

y la relación $\sim$ dada por

$$(x_n)_n \sim (y_n)_n \iff \lim_{n \to \infty} ||x_n - y_n|| = 0.$$

Probemos que se trata de una relación de equivalencia

- Reflexividad: Como $\lim_{n\to\infty} ||x_n - x_n|| = \lim_{n\to\infty} 0 = 0$, tenemos que $(x_n)_n \sim (x_n)_n$.
- Simetría: Sabemos que $\lim_{n\to\infty} ||x_n - y_n|| = 0$, pero además $\lim_{n\to\infty} ||x_n - y_n|| = \lim_{n\to\infty} ||y_n - x_n||$ al tratarse de una norma, luego $(y_n)_n \sim (x_n)_n$.
- Transitividad: Supongamos que $(x_n)_n \sim (y_n)_n$ y $(y_n)_n \sim (z_n)_n$. Tenemos que
$$||x_n - z_n|| = ||x_n - y_n + y_n - z_n|| \leq ||x_n - y_n|| + ||y_n - z_n||,$$
luego $\lim_{n\to\infty} ||x_n - z_n|| = 0$, es decir $(x_n)_n \sim (z_n)_n$.

En particular, $\sim$ a parte de ser una relación de equivalencia cumple que si $(x_n)_n, (y_n)_n \in H'$ entonces

$$[(x_n)_n] = [(y_n)_n] \text{ o } [(x_n)_n] \cap [(y_n)_n] = \emptyset.$$

Veamoslo: supongamos que $[(x_n)_n] \cap [(y_n)_n] \neq \emptyset$, es decir, existe una sucesión $(z_n)_n$ tal que $(x_n)_n \equiv (z_n)_n$ e $(y_n)_n \sim (z_n)_n$. Supongamos además que $[(x_n)_n] \neq [(y_n)_n]$ luego existe una sucesión $(u_n)_n$ tal que $(x_n)_n \sim (u_n)_n$ e $(y_n)_n \nsim (u_n)_n$. Podría ser al revés, pero sería un caso análogo. Entonces sabemos que

$$\lim_{n \to \infty} ||x_n - u_n|| = 0.$$

Por la desigualdad triangular tenemos que

$$||y_n - u_n|| = ||y_n - z_n + z_n - x_n + x_n - u_n|| \leq ||y_n - z_n|| + ||z_n - x_n|| + ||x_n - u_n||,$$

por lo que $\lim_{n\to\infty} ||y_n - u_n|| = 0$ llegando a contradicción.
Vamos a definir

$$H_0 = \{[(x_n)_n] : (x_n)_n \in H'\}.$$

Probemos ahora que si $(x_n)_n, (y_n)_n \in H'$ entonces $\lim_{n\to\infty}\langle x_n, y_n \rangle$ existe. Observemos primero que

$$|\langle x_n, y_n \rangle - \langle x_m, y_m \rangle| \leq |\langle x_n - x_m, y_n \rangle| + |\langle x_m, y_n - y_m \rangle| \leq ||x_n - x_m|| \, ||y_n|| + ||x_m|| \, ||y_n - y_m||$$

y como $(x_n)_n, (y_n)_n$ son de Cauchy, $(||x_n||)_n, (||y_n||)_n$ están acotadas y $||x_n - x_m||, ||y_n - y_m||$ convergen a 0 cuando $n, m$ tienden a $\infty$. Por tanto, $(\langle x_n, y_n \rangle)_n$ es de Cauchy y como $\mathbb{C}$ es completo, la sucesión converge.
Ahora vamos a definir una serie de operaciones en $H_0$ y probar que están bien definidas. Sea $[(x_n)_n], [(y_n)_n] \in H_0$ y sea $\alpha \in \mathbb{C}$. Tenemos que

- $[(x_n)_n] + [(y_n)_n] = [(x_n + y_n)_n]$.

- $\alpha[(x_n)_n] = [(\alpha x_n)_n]$.
- $\langle [(x_n)_n], [(y_n)_n] \rangle_{H_0} = \lim_{n\to\infty} \langle x_n, y_n \rangle_H$.

Consideramos $(x_n)_n \sim (x'_n)_n$ e $(y_n)_n \sim (y'_n)$. Tenemos que se cumple

$$|\langle x_n, y_n \rangle_H - \langle x'_n, y'_n \rangle_H| \leq ||x_n - x'_n||_H ||y_n||_H + ||x'_n||_H ||y_n - y'_n||_H.$$

Como $(x_n)_n \sim (x'_n)$, $\lim_{n\to\infty} ||x_n - x'_n|| = 0$ y lo mismo para $(y_n)_n \sim (y'_n)_n$. Además $(||x'_n||)_n$ y $(||y_n||)_n$ estn acotadas, por lo que

$$\lim_{n\to\infty} |\langle x_n, y_n \rangle - \langle x'_n, y'_n \rangle| = 0.$$

Así que el producto interno está bien definido. De forma similar

$$||(x_n + y_n) - (x'_n + y'_n)||_H \leq ||x_n - x'_n||_H + ||y_n - y'_n||_H \overset{n\to\infty}{\longrightarrow} 0$$

$$||\alpha x_n - \alpha x'_n||_H \leq |\alpha| \; ||x_n - x'_n||_H \overset{n\to\infty}{\longrightarrow} 0$$

Por lo que la suma y la multiplicación por un escalar también están bien definidos.
Si consideramos el conjunto $H_0$ con las operaciones que acabamos de definir, podemos concluir que $H_0$ es un espacio preHilbert. Sencillos argumentos de computación permiten verificar que $H_0$ es un espacio vectorial y con el producto interno definido antes es un espacio preHilbert, por lo que lo omitiremos en esta prueba.
Probaremos ahora que $H_0$ es completo y por tanto es un espacio de Hilbert. Sea $(X^{(n)} \in H_0)_{n\in\mathbb{N}}$ una sucesión de Cauchy. Debemos probar que tiene un límite $X \in H_0$ cuando $n$ tienda a infinito. Cada $X^{(n)}$ es una clase de equivalencia de las sucesiones de Cauchy en $H$. Supongamos que $X^{(n)} = [(x_l^{(n)})_{l\in\mathbb{N}}]$. Construiremos nuestra sucesión $X = [(x_n)_{n\in\mathbb{N}}]$ eligiendo cada $x_n$ como un $x_{l_n}^{(n)}$ con $l_n$ cada vez mayor cuando $n$ aumenta. Para ello:

> $(x_l^{(1)})_{l\in\mathbb{N}}$ es de Cauchy $\Rightarrow \exists\, l_1 \in \mathbb{N}$ tal que $l \geq l_1 \Rightarrow ||x_l^{(1)} - x_{l_1}^{(1)}||_H < 1$. Escogemos $x_1 = x_{l_1}^{(1)}$.
> $(x_l^{(2)})_{l\in\mathbb{N}}$ es de Cauchy $\Rightarrow \exists\, l_2 > l_1$ tal que $l \geq l_2 \Rightarrow ||x_l^{(2)} - x_{l_2}^{(2)}||_H < \frac{1}{2}$. Escogemos $x_2 = x_{l_2}^{(2)}$.
>
> $$\vdots$$
>
> $(x_l^{(n)})_{l\in\mathbb{N}}$ es de Cauchy $\Rightarrow \exists\, l_n > l_{n-1}$ tal que $l \geq l_n \Rightarrow ||x_l^{(n)} - x_{l_n}^{(n)}||_H < \frac{1}{n}$. Escogemos $x_n = x_{l_n}^{(n)}$.

Debemos probar ahora que esta sucesión efectivamente es de Cauchy. Sea $\epsilon > 0$. Por la desigualdad triangular tenemos que:

$$||x_n - x_m||_H = ||x_{l_n}^{(n)} - x_{l_m}^{(m)}||_H = ||x_{l_n}^{(n)} - x_l^{(n)} + x_l^{(n)} - x_l^{(m)} + x_l^{(m)} - x_{l_m}^{(m)}||_H \leq$$

$$\leq ||x_{l_n}^{(n)} - x_l^{(n)}||_H + ||x_l^{(n)} - x_l^{(m)}||_H + ||x_l^{(m)} - x_{l_m}^{(m)}||_H =$$

$$= ||x_{l_n}^{(n)} - x_l^{(n)}||_H + \Big( ||x_l^{(n)} - x_l^{(m)}||_H - ||X^{(n)} - X^{(m)}||_{H_0} \Big) + ||X^{(n)} - X^{(m)}||_{H_0} + ||x_l^{(m)} - x_{l_m}^{(m)}||_H,$$

para cualquier $l \in \mathbb{N}$.

- Si $l \geq l_n$, el primer término de la expresión anterior es más pequeño que $\frac{1}{n}$.
- Por definición, $||X^{(n)} - X^{(m)}||_{H_0} = \lim_{n\to\infty} ||x_l^{(n)} - x_l^{(m)}||_H$. Por tanto, existe un número natural $N_{n,m}$ tal que el segundo término es menor que $\frac{\epsilon}{4}$ cuando $l \geq N_{n,m}$.
- Por hipótesis, la sucesión $(X^{(n)})_{n\in\mathbb{N}}$ es de Cauchy. Por tanto, existe un número natural $\tilde{N}$ tal que el tercer término es más pequeño que $\frac{\epsilon}{4}$ cuando $n, m \geq \tilde{N}$.
- Finalmente, si $l \geq l_m$, entonces el último termino es menor que $\frac{1}{m}$.

Escogemos un número natural $N \geq max\{\tilde{N}, \frac{4}{\epsilon}\}$. De esta forma, tomando $n, m \geq N$, garantizamos que se cumpla el tercer punto. Es más, como $n, m \geq N \geq \frac{4}{\epsilon}$, tenemos que $\frac{1}{n}, \frac{1}{m} < \frac{\epsilon}{4}$. Si escogemos un $l$ natural tal que es mayor que $max\{N_{n,m}, l_n, l_m\}$ se cumplirá también el segundo punto, por lo que cada uno de los cuatro términos es menor que $\frac{\epsilon}{4}$. Así, podemos concluir que $||x_n - x_m||_H < \epsilon$ y por tanto la sucesión es de Cauchy.
Falta probar que $X = \lim_{n\to\infty} X^{(n)}$. Sea $\epsilon > 0$. Por definición:

$$||X - X^{(n)}||_{H_0} = \lim_{n\to\infty} ||x_m - x_m^{(n)}||_H = \lim_{n\to\infty} ||x_{l_m}^{(m)} - x_m^{(n)}||_H.$$

Aplicando la desigualdad triangular:

$$||x^{(m)}_{l_m} - x^{(n)}_m||_H \leq ||x^{(m)}_{l_m} - x^{(n)}_{l_n}||_H + ||x^{(n)}_{l_n} - x^{(n)}_m||_H.$$

- Como la sucesión $(x_n = x^{(n)}_{l_n})_{n\in\mathbb{N}}$ es de Cauchy, existe un natural $N'$ tal que el primer término de la anterior expresión es menor que $\frac{\epsilon}{2}$ cuando $n, m \geq N'$.
- Por la construcción de $l_n$, el segundo término es menor que $\frac{1}{n}$ cuando $m \geq l_n$.

Con un argumento idéntico al anterior, escogemos un número natural $N \geq max\{N', \frac{2}{\epsilon}\}$. Vamos a ver que $||X - X^{(n)}|| < \epsilon$ cuando $n \geq N$. Para ello, si $n \geq N$ y consideramos un $m$ mayor que $max\{N', l_n\}$, cada término de la expresión anterior es menor que $\frac{\epsilon}{2}$, probando que la sucesión tiene límite y que por ello $H_0$ es completo. Podemos definir ahora la aplicación

$$\begin{aligned} I : H &\longrightarrow H_0 \\ x &\longmapsto [(x)_n], \end{aligned}$$

es decir, a cada elemento $x \in H$ le lleva a la sucesión cuyos elementos son el propio $x$, que claramente es una sucesión de Cauchy. Además, al tratarse $I$ de una aplicacin cociente, sabemos que es continua. Probemos que $I$ es lineal y conserva el producto interno. Tenemos que

$$\langle I(x), I(y)\rangle_{H_0} = \langle [(x)_n], [(y)_n]\rangle_{H_0} = \lim_{n\to\infty} \langle x, y\rangle_H = \langle x, y\rangle_H,$$

probando así la segunda propiedad. Para probar que es lineal, basta ver que

$$I(x + y) = [(x + y)_n] = [(x)_n] + [(y)_n] = I(x) + I(y).$$

Falta probar que $I(H)$ es denso para finalizar con la primera parte del enunciado. Sea $X = [(x_n)_n] \in H_0$. Vamos a probar que la sucesión $(I(x_m))_m$ converge en $H_0$ a $X$. Observamos que

$$||X - I(x_m)||_{H_0} = ||[(x_n)_n - [(x_m)_n]||_{H_0} = ||[(x_n - x_m)_n||_{H_0} = \langle [(x_n - x_m)_n], [(x_n - x_m)_n]\rangle_{H_0} =$$
$$= \lim_{n\to\infty} \langle x_n - x_m, x_n - x_m\rangle = \lim_{n\to\infty} ||x_n - x_m||_H,$$

límite que es igual a 0 cuando $m \longrightarrow \infty$ al tratarse $(x_n)_n$ una sucesión de Cauchy.
Sea $H$ un espacio preHilbert y sean $H_0$ y $H'$ dos completaciones, es decir, existen $I : H \longrightarrow H_0$ y $J : H \longrightarrow H'$ isometrías lineales con $I(H)$ y $J(H)$ densos en $H_0$ y $H'$ respectivamente. Queremos ver que existe una isometría $\varphi : H_0 \longrightarrow H'$.
Sea $x^0 \in H_0$. Como $I(H)$ es denso en $H_0$, para cada $n \in \mathbb{N}$, existe un punto $x^0_n \in B(x, \frac{1}{n}) \cap I(H)$. Claramente la sucesión $(x^0_n)_{n\in\mathbb{N}}$ converge a $x^0$ en $H_0$. Para cada $n \in \mathbb{N}$, definimos $x_n = I^{-1}(x^0_n)$, que están bien definidos por ser $I$ una isometría y por tanto inyectiva. Por otro lado, definimos $x'_n = J(x_n) \in H'$. Como $I$ y $J$ son isometrías, para todo $n, m \in \mathbb{N}$:

$$\langle x'_n, x'_m\rangle = \langle x^0_n, x^0_m\rangle.$$

Como $(x^0_n)_n$ converge en $H_0$, es de Cauchy y por la igualdad anterior, está claro que $(x'_n)_n$ también es de Cauchy y por ser $H'$ completo, la sucesión converge a algún $x' \in H'$.
Consideramos $(\bar{x}^0_n)_n$ otra sucesión en $H_0$ convergente a $x^0$. Para cada $\varepsilon > 0$ existe un $m_\varepsilon \in \mathbb{N}$ tal que $d(x^0_n, x^0), d(\bar{x}^0_n, x^0) < \frac{\varepsilon}{2}$ cuando $n \geq m_\varepsilon$, luego

$$d(x^0_n, \bar{x}^0_n) \leq d(x^0_n, x^0) + d(x^0, \bar{x}^0_n) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

si $n \geq m_\varepsilon$.
Sea $\bar{x}_n = I^{-1}(\bar{x}^0_n) \in H$ y $\bar{x}'_n = J(\bar{x}_n) \in H'$. De nuevo, como $I$ y $J$ son isometrías:

$$d(x'_n, \bar{x}'_n) = d(x^0_n, \bar{x}^0_n),$$

para cada $n \in \mathbb{N}$. Por ello, $d(x'_n, \bar{x}'_n) < \varepsilon$ si $n \geq m_\varepsilon$. Sea $\varepsilon > 0$. Como $(x'_n)_n$ converge a $x'$ en $H'$, existe un $k_\varepsilon$ tal que $d(x'_n, x') < \frac{\varepsilon}{2}$ cuando $n \geq k_\varepsilon$. Entonces:

$$d(\bar{x}'_n, x') \leq d(\bar{x}'_n, x'_n) + d(x'_n, x') < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

cuando $n \geq max\{k_\varepsilon, m_{\varepsilon/2}\}$ y $(\bar{x}'_n)_n$ converge a $x'$.
Por tanto, la aplicación $\varphi : H_0 \longrightarrow H'$ dada por $\varphi(x^0) = x'$ está bien definida: no importa que sucesión en $H_0$ convergente a $x^0$ tomemos para definir $x'$.
Es más, si $x^0 \in H_0$, puedo tomar $x^0_n = x^0$ para todo $n \in \mathbb{N}$ y entonces $\varphi(x^0) = J(I^{-1}(x^0))$ probando así que $\varphi \mid_{H_0} = J \circ I^{-1}$. Por tanto, como $I$ y $J$ son isometrías, son inyectivas y

al restringir el dominio, también sobreyectivas, por lo que efectivamente $\varphi$ restringido es un isomorfismo.
Falta probar que $h$ es una isometría. Sean $x^0, y^0$ dos elementos de $H_0$ y sean $(x_n^0)_n$ e $(y_n^0)_n$ dos sucesiones en $H^0$ tales que convergen a $x^0, y^0$ respectivamente. Entonces la sucesión $((x_n^0, y_n^0))_n$ converge a $(x^0, y^0)$ en $H_0 \times H_0$. Como la métrica $d$ es continua en $H_0 \times H_0$, $(d(x_n^0, y_n^0))_n$ converge a $d(x^0, y^0)$ en $\mathbb{R}$.
Como antes, para cada $n \in \mathbb{N}$, sea $x_n = I^{-1}(x_n^0) \in H$ y sea $x'_n = J(x_n) \in H'$, tal que $(x'_n)_n$ converge a $h(x^0) \in H'$. Análogo para $y'_n$. Por el mismo argumento que antes, tenemos que $(d(x'_n, y'_n))_n$ converge a $d(x', y')$. Luego $d(x_n^0, y_n^0) = d(x_n, y_n) = d(x'_n, y'_n)$ para cada $n \in \mathbb{N}$. Así:

$$d(h(x^0), h(y^0)) = \lim_{n\to\infty} d(x'_n, y'_n) = \lim_{n\to\infty} d(x_n^0, y_n^0) = d(x^0, y^0),$$

por lo que $h$ es isometría. □

Continuamos, de nuevo por completitud, con la demostración del siguiente teorema, que corresponde en el trabajo al teorema 1.1.14

TEOREMA A.1.7. *Sea $(V, ||\cdot||)$ un espacio normado sobre $\mathbb{K}$ y $L \in Hom_{\mathbb{L}}(V, \mathbb{K})$ una función lineal. Entonces, $L$ es continua si y solamente si es de norma acotada.*

DEMOSTRACIÓN. Probemos ambas implicaciones

$\Longrightarrow$) Supongamos que $L$ es continua. En particular, $L$ es continua en el 0, luego para $\varepsilon = 1 > 0$ existe un $\delta > 0$ tal que si $||x|| = ||x - 0|| \leq \delta$ entonces

$$||L(x)|| = ||L(x) - L(0)|| \leq 1.$$

Si $x \neq 0$, entonces

$$\left|\left|\frac{\delta x}{||x||}\right|\right| = \delta\frac{||x||}{||x||} = \delta,$$

por lo que

$$\left|\left|L\left(\frac{\delta x}{||x||}\right)\right|\right| \leq 1.$$

Para todo $x \in H$ tenemos que

$$\delta\frac{||L(x)||}{||x||} \leq 1,$$

por lo que

$$||L(x)|| \leq \frac{1}{\delta}||x||.$$

Efectivamente $L$ está acotado.

$\Longleftarrow$) Supongamos que $L$ está acotado. Entonces existe $M > 0$ tal que para todo $x \in H$ tenemos que

$$||L(x)|| \leq M||x||.$$

Sea $\varepsilon > 0$ y definimos $\delta = \frac{\varepsilon}{M}$. Entonces, si $||x - y|| < \delta$ tenemos que

$$||L(x) - L(y)|| = ||L(x - y)|| \leq M||x - y|| < M\delta = M\frac{\varepsilon}{M} = \varepsilon.$$

Por tanto, $L$ es continua.

□

## A.2. Pruebas de Algunos Resultados Sobre Bases Ortonormales

Seguidamente incluimos una demostración elemental del Teorema 1.2.1 del capítulo 1 por completitud:

TEOREMA A.2.1. *El complementario ortogonal de un subconjunto de un espacio de Hilbert es un subespacio lineal cerrado.*

Demostración. Sea $H$ un espacio de Hilbert y sea $A$ un subconjunto de $H$. Vamos a probar que $A^\perp$ es un subespacio lineal. Sean $y, z \in A^\perp$ y sean $\lambda, \mu \in \mathbb{C}$. Tenemos que ver que

$$\langle x, \lambda y + \mu z \rangle = 0, \ \forall x \in A.$$

La sesquilinealidad del producto interno nos dice que

$$\langle x, \lambda y + \mu z \rangle = \overline{\lambda} \langle x, y \rangle + \overline{\mu} \langle x, z \rangle = 0,$$

por ser $y, z$ ortogonales a $x$. Por tanto, $A^\perp$ es un subespacio lineal.
Para probar que $A^\perp$ es cerrado, tenemos que probar que para toda suceción $(y_n)$ convergente en $A^\perp$, el límite está en $A^\perp$. Por el Teorema 1.1.9, el producto interno es continuo, luego

$$\langle x, y \rangle = \langle x, \lim_{n\to\infty} y_n \rangle = \lim_{n\to\infty} \langle x, y_n \rangle = 0,$$

pues $\langle x, y \rangle = 0$ para todo $x \in A$ e $y_n \in A^\perp$. Por tanto, $y \in A^\perp$ y $A^\perp$ es cerrado. $\square$

Lo siguiente que hacemos es probar el resultado del teorema 1.2.2.

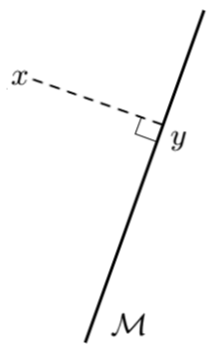
Teorema A.2.2. *Sea $\mathcal{M}$ un subespacio lineal cerrado de un espacio de Hilbert $H$.*

*a) Para cada $x \in H$ existe un único punto $y$ tal que*

$$||x - y|| = \min_{z \in \mathcal{M}} ||x - z||.$$

*b) El punto $y \in \mathcal{M}$ que cumple a) es el único elemento de $\mathcal{M}$ con la propiedad de que $(x - y) \perp \mathcal{M}$.*

Demostración. Visualmente el resultado puede parecer obvio, si nos imaginamos $\mathcal{M}$ como una recta, $x$ como un punto fuera de la recta e $y$ el punto de $\mathcal{M}$ más próximo a x, como podemos ver en la siguiente imagen.




La demostración realmente no es trivial. Empezaremos por probar la existencia del $y$ descrito en el apartado $a)$. Sea $d$ la distancia de $x$ a $\mathcal{M}$,

$$d = \inf\{||x - z|| : z \in \mathcal{M}\}.$$

Tal y como hemos definido $d$, existe una sucesión de elementos $y_n \in \mathcal{M}$ tal que

$$\lim_{n\to\infty} ||x - y_n|| = d.$$

Entonces, para todo $\varepsilon > 0$, existe un $N$ tal que

$$||x - y_n|| \leq d + \varepsilon,$$

cuando $n \geq N$. Veamos que la sucesión $(y_n)$ es de Cauchy. Si aplicamos la ley del paralelogramo a $x - y_n$ y $x - y_m$, tenemos que

$$||2x - y_m - y_n||^2 + ||y_m - y_n||^2 = 2||x - y_m||^2 + 2||x - y_n||^2.$$

Como $\mathcal{M}$ es lineal, tenemos que $\frac{y_m + y_n}{2} \in \mathcal{M}$ y por la definición de $d$ se cumple que

$$||x - \frac{y_m + y_n}{2}|| \geq d.$$

Por tanto, para todo $m, n \geq N$,

$$||y_m - y_n||^2 = 2||x - y_m||^2 + 2||x - y_n||^2 - ||2x - y_m - y_n||^2 \leq 4(d + \varepsilon)^2 - 4d^2 \leq 4\varepsilon(2d + \varepsilon).$$

Efectivamente $(y_n)$ es de Cauchy. Como un espacio de Hilbert es completo, existe un $y$ tal que $\lim_{n\to\infty} y_n = y$ y como $\mathcal{M}$ es cerrado, entonces $y \in \mathcal{M}$. La norma es una función continua, luego

$$||x-y|| = \lim_{n\to\infty} ||x-y_n|| = d.$$

Ahora probamos la unicidad del vector $y \in \mathcal{M}$ que minimiza $||x-y||$. Supongamos que existen dos vectores $y$ e $y'$ de tal forma que ambos minimizan la distancia a $x$, es decir

$$||x-y|| = d, \qquad ||x-y'|| = d.$$

Aplicando la ley del paralelogramo a $x-y$ y $x-y'$ tenemos que

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

Como $\frac{y+y'}{2} \in \mathcal{M}$, $||x - \frac{y+y'}{2}|| \geq d$

$$||y-y'|| = 2||x-y||^2 + 2||x-y'||^2 - ||2x-y-y'||^2 = 4d^2 - 4||x - \frac{y+y'}{2}||^2 \leq 0.$$

Por tanto $||y-y'|| = 0$ y entonces $y = y'$, probando así la unicidad y con ello el apartado $a$).

Vamos con el apartado $b$). Debemos probar que el $y$ que minimiza la distancia con $x$ cumple que $(x-y) \perp \mathcal{M}$. Como $y$ minimiza la distancia, tenemos que para todo $\lambda \in \mathbb{C}$ y $z \in \mathcal{M}$ se cumple

$$||x-y||^2 \leq ||x-y+\lambda z||^2.$$

Si expandimos el lado derecho de la desigualdad, tenemos que

$$||x-y||^2 \leq ||x-y+\lambda z||^2 = \langle x-y+\lambda z, x-y+\lambda z\rangle =$$
$$\langle x-y, x-y\rangle + \langle x-y, \lambda z\rangle + \langle \lambda z, x-y\rangle + \langle \lambda z, \lambda z\rangle = ||x-y||^2 + \overline{\lambda}\langle x-y, z\rangle + \lambda\overline{\langle x-y, z\rangle} + |\lambda|^2||z||^2,$$

por lo que

$$0 \leq \overline{\lambda}\langle x-y, z\rangle + \overline{(\overline{\lambda}\langle x-y, z\rangle)} + |\lambda|^2||z||^2 = 2Re(\overline{\lambda}\langle x-y, z\rangle) + |\lambda|^2||z||^2,$$

y entonces

$$-2Re(\overline{\lambda}\langle x-y, z\rangle) \leq |\lambda|^2||z||^2.$$

Podemos escribir $\langle x-y, z\rangle = |\langle x-y, z\rangle|e^{i\theta}$. Si escogemos $\lambda = -\varepsilon e^{i\theta}$ con $\varepsilon > 0$, entonces

$$-2Re(-\varepsilon e^{-i\theta}|\langle x-y, z\rangle|e^{-i\theta}) =\leq |-\varepsilon e^{i\theta}|^2||z||^2,$$
$$-2(-\varepsilon|\langle x-y, z\rangle|) \leq \varepsilon^2||z||^2,$$
$$2|\langle x-y, z\rangle| \leq \varepsilon||z||^2.$$

Tomando el límite cuando $\varepsilon \to 0^+$, tenemos que $\langle x-y, z\rangle = 0$, luego $(x-y) \perp \mathcal{M}$ tal y como queríamos probar.
Finalmente, vamos a probar que $y$ es el único elemento de $\mathcal{M}$ con la propiedad de que $(x-y) \perp \mathcal{M}$. Suponemos que existe otro elemento así en $\mathcal{M}$ y llamemosle $y'$. Entonces para cualquier $z \in \mathcal{M}$ se cumple que

$$\langle x-y, z\rangle = 0, \qquad \langle x-y', z\rangle = 0.$$

Como $\langle x-y, z\rangle = 0$, entonces $-\langle x-y, z\rangle = \langle y-z, z\rangle = 0$ y tenemos que

$$\langle y-x, z\rangle + \langle x-y', z\rangle = \langle y-x+x-y', z\rangle = \langle y-y', z\rangle = 0.$$

Tomando $z = y-y' \in \mathcal{M}$, tenemos que $\langle y-y', y-y'\rangle = 0$ y por las propiedades del producto interno, concluimos que $y-y' = 0$ y por tanto $y = y'$. $\square$

A continuación recordamos el Teorema 1.2.4 del Capítulo 1 y damos su demostración.

TEOREMA A.2.3. *Si $U = \{u_i : i \in I\}$ es un subconjunto ortonormal de un espacio de Hilbert $H$, entonces las siguientes condiciones son equivalentes:*

*i)* $\langle x, u_i\rangle = 0$ *para todo* $i \in I$ *implica* $x = 0$;
*ii)* $x = \sum_{i\in I}\langle x, u_i\rangle u_i$ *para todo* $x \in H$;
*iii)* $||x||^2 = \sum_{i\in I}|\langle x, u_i\rangle|^2$ *para todo* $x \in H$;
*iv)* $[U] = \{\sum_{u\in U} c_u u : c_u \in \mathbb{C}$ *y* $\sum_{u\in U} c_u u$ *converge incondicionalmente*$\} = H$;
*v)* $U$ *es un conjunto ortonormal maximal.*

DEMOSTRACIÓN. Probaremos que $i) \Rightarrow ii) \Rightarrow iii) \Rightarrow iv) \Rightarrow v) \Rightarrow i)$:

$i) \Rightarrow ii)$ La propiedad $i)$ implica que $U^\perp = 0$. Para probar que $i)$ implica $ii)$, vamos a probar que

$$x - \sum_{i\in I} \langle x, u_i\rangle u_i \in U^\perp.$$

Sea $u_{i_0} \in U$. Tenemos que por la continuidad del producto interno y por la ortonormalidad de $U$:

$$\langle x - \sum_{i\in I}\langle x, u_i\rangle u_i, u_{i_0}\rangle = \langle x, u_i\rangle - \langle \sum_{i\in I}\langle x, u_i\rangle u_i, u_{i_0}\rangle =$$
$$= \langle x, u_{i_0}\rangle - \sum_{i\in I}\langle u_i, x\rangle\langle u_i, u_{i_0}\rangle = \langle x, u_{i_0}\rangle - \langle x, u_{i_0}\rangle = 0.$$

$ii) \Rightarrow iii)$

$$||x||^2 = ||\sum_{i\in I}\langle u_i, x\rangle u_i||^2 = \sum_{i,j\in I}\langle\langle u_i, x\rangle u_i, \langle u_j, x\rangle u_j\rangle = \sum_{i,j\in I}\langle u_i, x\rangle\overline{\langle u_j, x\rangle}\langle u_i, u_j\rangle =$$
$$= \sum_{i\in I}\langle u_i, x\rangle\overline{\langle u_i, x\rangle}\langle u_i, u_i\rangle = \sum_{i\in I}|\langle u_i, x\rangle|^2||u_i||^2 = \sum_{i\in I}|\langle u_i, x\rangle|^2.$$

$iii) \Rightarrow iv)$ Si $\langle u_i, x\rangle = 0$ para todo $i \in I$, $||x||^2 = 0$, luego $x = 0$ y $U^\perp = \{0\rangle$. Entonces, $[U]^\perp = 0$ y finalmente $[U] = H$.

$iv) \Rightarrow v)$ Si $[U] = H$, $U^\perp = \{0\rangle$ cumpliéndose así $ii)$, por lo que cualquier $v \in H$ puede escribirse como $v = \sum_{i\in I} c_i u_i$ con $c_i = \langle u_i, v\rangle$. Entonces, si $v \perp U$, $c_i = 0$ para todo $i \in I$, por lo que $v = 0$. Por tanto $U \cup \{v\}$ no es ortonormal, probando así que no existe un subespacio $V$ ortonormal con $U \subset V$.

$v) \Rightarrow i)$ Sea $U$ un conjunto ortonormal maximal y sea $x \neq 0 \in H$ tal que $\langle u_i, x\rangle = 0$. Entonces, podemos considerar $V = U \cup \frac{x}{||x||}$ que sería un conjunto ortogonal con $U \subset V$ llegando a contradicción con el hecho de que $U$ es maximal.

□

Probamos finalmente el Teorema 1.2.5 del Capítulo 1.

TEOREMA A.2.4. *Asumiendo el* lema de Zorn*, todo espacio de Hilbert $H$ posee una base ortonormal.*

DEMOSTRACIÓN. Enunciamos el *lema de Zorn* como recordatorio: *"Todo conjunto parcialmente ordenado no vacío en el que toda cadena (subconjunto totalmente ordenado) tiene una cota superior, contiene al menos un elemento maximal".* Si $H = \{0\}$, la afirmación es cierta considerando $U = \emptyset$, así que supondremos que $H \neq \{0\}$. Definimos un orden parcial $\leq$ para los subconjuntos ortonormales de $H$ por inclusión, es decir, $U \leq V$ si y solo si $U \subset V$. Si consideramos $\{U_\alpha : \alpha \in \mathcal{A}\}$ como un conjunto de subconjuntos ortogonales totalmente ordenado donde $U_\alpha \leq U_\beta$ o $U_\beta \leq U_\alpha$, entonces $\cup_{\alpha\in\mathcal{A}} U_\alpha$ es un conjunto ortonormal y es una cota superior de $\{U_\alpha : \alpha \in \mathcal{A}\}$. El lema de Zorn nos asegura la existencia de un elemento maximal para el conjunto de subconjuntos ortogonales de $H$, por lo que ese elemento cumple la condición $v)$ del Teorema 1.2.4 y por tanto $H$ posee una base ortonormal. □

# Bibliografía


[Ar, 50] N. Aronszajn, *"Theory of reproducing kernels"*, Transactions of the American Mathematical Society, Vol. 68, No. 3 (May, 1950), pp. 337-404.

[AMP, 09] A. Argyriou, C. A. Micchelli and M, Pontil, *"When is there a Representer Theorem? Vector versus matrix regularizers"*, Journal of Machine Learning Research 10, 2009, pp. 2507-2529.

[AT, 77] V. Y. Arsenin and A. N. Tikhonov, *"Solutions of ill pased problems"*, V. H. Winston and Sons (distributed by Wiley), 1977, F. John, Translation Editor.

[BCSS, 98] L. Blum, F. Cucker, M. Shub and S. Smale, *"Complexity and real computation"*, Springer, New York, 1998.

[CMR, 03] S. Canu, X. Mary and A. Rakotomamonjy, *"Functional learning through kernels, advances in learning theory: methods, models and applications"*, NATO Science Series, Vol. 190, IOS Press, 2003, pp. 89-110.

[CP, 97] J.T. Chang and D. Pollard, *"Conditioning as disintegration"*, Statistica Neerlandica, 51, pp. 287-317.

[CS, 02] F. Cucker and S. Smale, *"On the mathematical foundations of Learning"*, Bulletin of the American Mathematical Society. Vol.39, 2002, pp. 1-49.

[CV, 71] A. Ya. Chervonenkis and V. N. Vapnik, *"On uniform convergence of the frequencies of events to their probabilities"*, Theory Probab. Appl, Vol. 16, pp. 264-280.

[CZ, 07] F. Cucker and D. Zhou *"Learning theory: an approximation theory viewpoint"*, Cambridge Monographs on Applied and Computational Mathematics, 2007.

[De, 06] J. P. Dedieu, *"Points fixes, zéros et la méthode de Newton"*, Springer, Berlín, 2006.

[Gi, 18] F. Girosi, *"An equivalence between sparse approximation and support vector machines"*, in Neural Computation, vol. 10, no. 6, pp. 1455-1480.

[HN, 01] J. K. Hunter and B. Nachtergaele, *"Applied analysis"*, World Scientific (2001).

[HSS, 01] R. Herbrich, B. Scholkopf and A. J. Smola, *"A generalised Representer Theorem"*, Proc. of the 14th Annual Conf. on Computational Learning Theory, 2001.

[KW, 70] G. S. Kimeldorf and G. Wahba, *"A correspondence between Bayesian estimation on stochastic processes and smoothing by splines"*, The Annals of Mathematical Statistics, Vol. 41, No. 2 (Apr., 1970), pp. 495-502.

[Me, 09] J. Mercer, *"Functions of positive and negative type, and their connection the theory of integral equations"*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, Vol. 209, 1909, pp. 415-446.

[OS, 17] H. Owhadi and C. Scovel, *"Separability of reproducing kernel spaces"*, Proceedings of the American Mathematical Society, Vol. 145, No. 5 (May, 2017), pp. 2131-2138.

[Po, 84] D. Pollard, *"Convergence of stochastic processes"*, Springer Series in Statistics, Springer-Verlag New York, 1984.

[PR, 16] V. I. Paulsen and M. Raghupathi, *"An introduction to the theory of reproducing kernel Hilbert spaces"*, Cambridge Univ. Press, 2016.

[PS, 20] L. M. Pardo, D. Sebastián, *"A gentle walk on correct test sequences"*. Manuscrito, U. Cantabria, 2020.