UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Matemática

Tesis de Licenciatura

Aplicaciones de la teoría de núcleos reproductores

Manuel Enrique Nores

Director: Daniel Carando

Fecha de Presentación: 3 de octubre de 2025

Índice general

Agradecimientos

Me es difícil encontrar las palabras que logren expresar las emociones que me pasan por el corazón en este momento. Soy víctima de una de las grandes limitaciones del lenguaje: no todo lo que existe (y sobre todo, lo que se siente) puede ser dicho con palabras. Lo que sigue va a ser mi mejor intento de achicar esa brecha entre lo existente y lo expresable mediante palabras. Espero que el profundo cariño que siento por todas las personas que nombraré a continuación no se pierda en la traducción.

Le quiero agradecer primero a Dani por aceptar dirigir esta tesis. Yo creo que la palabra “dirigir” se queda extremadamente corta para lo que fue (y seguirá siendo) su rol en este proceso. Dani, fuiste un guía extremadamente presente en la abrumadora tarea de escribir una tesis de licenciatura. Frente a cualquier pregunta, por básica que sea (y créanme que las hubo bien básicas), tu respuesta siempre fue de absoluta calidez y pedagogía. Todas tus correcciones, que fueron precisas y expeditivas, fueron fundamentales en mi formación profesional. Espero con entusiasmo seguir trabajando con vos en el doctorado.

A Julián Fernández Bonder y a Martín Mansilla, por aceptar ser jurados de esta tesis. Sé que el tiempo necesario para juzgar el trabajo no es poco, por lo que les agradezco enormemente que se hayan hecho el tiempo para esto.

A la UBA, por todo lo que me dio en esta etapa. Por abrirme las puertas para recibir una formación científica de excelencia y de manera gratuita. Por permitirme también desempeñar mis primeros cargos docentes, donde también encontré un interés que estoy dispuesto a seguir explorando. En esta línea, también quiero agradecer a todos los docentes que estuvieron presentes a lo largo de este camino. Quisiera empezar nombrando a Carlos A. Fuentes, profesor de Análisis y Álgebra en el CBC. Él fue la primera persona que me enseñó que la matemática es mucho más que la manipulación de expresiones algebraicas y la mecanización de procedimientos: es una manera de pensar y de mirar al mundo. Un lenguaje para hacer precisas ideas difíciles de manejar. A Jonathan A. Barmak, por terminar de enseñarme qué significa el rigor en la matemática, y acompañarme en tantas materias más avanzadas. A Victoria Paternostro, por enseñarme la belleza del análisis en esa cursada de Análisis Real el segundo cuatrimestre de 2023. Como la lista es interminable, prefiero terminarla ahora, aunque sé que estoy cometiendo una injusticia hacia todos los demás docentes que me inspiraron a lo largo de la carrera.

A mis padres, por creer en mí incondicionalmente, aún en los momentos en los que yo mismo no creía que iba a llegar a ningún lado. Tuve el privilegio de crecer en una casa

donde nunca me faltó nada, y de poder dedicarme a mis estudios sabiendo que nunca me iba a faltar un plato de comida o un lugar para dormir. Aunque muchas veces no lo hayan entendido, muchas gracias por siempre apoyar mi pasión por la matemática, por alegrarse por mí en cada pequeño logro y compartirlos con su gente (aunque la mayoría de veces me haya dado vergüenza la atención), y simplemente por estar ahí en cada paso de mi vida.

A mis hermanos, en especial a Toribio y a Belén, por ser una extensión de ese amor incondicional de nuestros padres. A ustedes les tocó sacarme de mis pozos oscuros más difíciles. Hablarme cuando me sentía un fracaso absoluto, un farsante o simplemente un chico frustrado con la vida. Siempre lograron ver en mí una luz que yo muchas veces no pude ver. Sus palabras de cariño, fe y apoyo fueron uno de los grandes motores para que yo siga apostando por lo que me apasiona, por lo que me mueve, por lo que me enciende. En definitiva, gracias a ustedes yo seguí apostando por mí, y les estoy eternamente agradecido por eso. Espero poder retribuirles todo esto siendo el hermano que se merecen.

A mis amigos más cercanos: Chino y Cata. Ustedes hacen de la vida un lugar hermoso para estar. Su amistad es uno de mis lugares preferidos para descansar del frenesí del mundo. Son personas llenas de luz y tenerlos en mi vida es un privilegio del cual estoy infinitamente agradecido. Los momentos compartidos con ustedes (manejar escuchando Reneé Rapp, escuchar a Chino explicar el tercer juego de mesa de la noche, etc.) me llenan de una alegría que no se puede comparar. No suelo expresarles lo mucho que los quiero, pero eso es simplemente por limitaciones de mi propia persona, y no por falta de cariño. Gracias por seguir eligiendo pasar su tiempo conmigo.

A todos mis compañeros de cursada durante estos años, que hoy en día no tengo problema de llamar amigos. Ian, Bruno, Mateo, Fran, Luca, Lucas, Jorge, Lau, Cami H., Cami M., Juli, Pedro, Tomer, Franco, Vicky, Lucho, Nico, Delfi, Octa, Ana y tantos más que, nuevamente, no me queda otra que cometer una injusticia dejando personas sin nombrar. Ustedes hicieron de la carrera una de las mejores experiencias de mi vida. Discutir a las 2 de la mañana problemas de topología, juntarnos a comer un asado y jugar al truco, viajar juntos a Catamarca o a Pinamar. Compartir estas experiencias con ustedes fue un regalo, y por ello los llevo a todos en el corazón. Todos ustedes serán unos matemáticos increíbles, y estoy orgulloso de en el futuro poder llamarlos colegas.

A todos los amigos que me regaló el teatro musical. Barbi Wendy, Oli, Lula, Eze, Vane, Luz por nombrar a algunos. Ustedes hicieron de los martes a la noche (y de varios momentos más) un espectáculo. Incorporar el arte en mi vida fue algo fundamental para que mi existencia no gire únicamente en torno a la matemática. En varias ocasiones he dicho que sin ustedes, yo probablemente sería un robot que solamente se puede comunicar en el lenguaje de la lógica. Me enseñaron y me siguen enseñando a no tomarme a mí mismo tan seriamente, y que el juego en la vida hace que la vida sea un paseo para disfrutar. Compartir con ustedes el teatro y la vida fue uno de los regalos más lindos que se me dio.

A mis amigos de la secundaria, que si bien no estuvimos coincidiendo mucho en este último tiempo, cada vez que nos vemos es como si no hubiera pasado el tiempo. Ustedes

son otro de mis santuarios en vida, y disfruto mucho de su compañía. Aunque a veces no lo parezca, pasar tiempo con ustedes me deja siempre un poco más contento. Espero que podamos seguir encontrando lugares para compartir las cosas más simples (pero no por ello poco importantes) de vivir.

A la empresa LambdaClass, por otorgarme una beca durante el último cuatrimestre de mi carrera sin pedirme nada a cambio. Fue un lujo contar con un estipendio que me permitió enfocarme en terminar la tesis. Ojalá puedan seguir ayudando a estudiantes como yo en el futuro.

A pesar de que conlleve un riesgo de sonar egocéntrico o soberbio, quiero agradecerle a la versión de mí mismo del 2021, que tomó la decisión de dejar la carrera de ingeniería para empezar a estudiar matemática, algo totalmente impensado para mí en ese entonces. Fue un salto al vacío que terminó siendo una de las mejores decisiones que tomé en mi vida.

Por último, quiero agradecerle a todo aquel que haya leído estos agradecimientos hasta el final. Sea porque están nombrados en ellos, o porque me tienen el suficiente cariño como para querer ver mi tesis de licenciatura, o simplemente porque se encontraron de manera aleatoria este documento y quisieron ver qué decía. Ahora ustedes saben de todas las personas que fueron importantes en mi camino, y eso me llena el alma.

Introducción

Los *espacios de Hilbert con núcleo reproductor* (RKHS por sus siglas en inglés) fueron introducidos formalmente en 1950 por N. Aronszajn en [1] y su uso se ha multiplicado en áreas como la teoría de muestreo, el aprendizaje automático, los procesos estocásticos y la estadística funcional, lo que se suma a sus aplicaciones clásicas al análisis (armónico, funcional y complejo), la teoría de representaciones de grupos y la teoría de operadores integrales. Estos espacios ofrecen un marco unificado para tratar diversos problemas de interpolación, aproximación y análisis de datos, lo que los hace atractivos tanto desde el punto de vista teórico como desde las aplicaciones.

Un RKHS es un espacio de Hilbert $\mathcal{H}$ de funciones (sobre un conjunto T) en el cual los funcionales de evaluación son continuos. Es decir, para cada $t \in T$, la operación $f \mapsto f(t)$ es continua de $\mathcal{H}$ en $\mathbb{K}$ (el cuerpo de escalares reales o complejos). Esta propiedad, a través del teorema de representación de Riesz, permite definir un *núcleo* $K : T \times T \to \mathbb{K}$ que *reproduce* los elementos de $\mathcal{H}$: para cada $t \in T$ y cada $f \in \mathcal{H}$ se tiene

$$f(t) = \langle f, K(\cdot, t)\rangle.$$

Un teorema importante de Moore establece que dar un RKHS sobre un conjunto T es equivalente a dar un *núcleo* sobre ese mismo conjunto. Es decir, una función $K : T\times T \to \mathbb{K}$ tal que para toda elección de puntos distintos $t_1, \ldots, t_n \in T$, la matriz $(K(t_i, t_j))_{1\leq i,j\leq n}$ es positiva. En vista de esta correspondencia, se suele denotar por $\mathcal{H}(K)$ al único RKHS en T cuyo núcleo reproductor es K. Es natural preguntarse, entonces, cómo las distintas operaciones que se le podían efectuar a los núcleos impactaban a los espacios resultantes, y viceversa. Esto fue el contenido del trabajo [1]. El primer objetivo de este trabajo es dar una introducción a estos espacios, junto con los resultados clásicos de la teoría.

La estructura de RKHS, más allá de sus numerosas aplicaciones en el análisis funcional, armónico y complejo, ha resultado muy útil en el estudio de *procesos estocásticos*. Un proceso estocástico es una familia de variables aleatorias $\{X_t\}_{t\in T}$ definidas sobre un espacio de probabilidad común Ω. Equivalentemente, es una aplicación $X : \Omega \times T \to \mathbb{R}$ tal que $X(\cdot, t)$ es una variable aleatoria para todo $t \in T$. Si todas las variables aleatorias tienen segundo momento finito, un núcleo natural asociado está dado por la *función de covarianza* del proceso definida por $K(s, t) = Cov(X_s, X_t)$. Resulta entonces de interés, para comprender mejor el proceso, estudiar el RKHS asociado a este núcleo particular.

Una de las cuestiones centrales en el estudio de los procesos estocásticos es la de la *regula-*

ridad de las trayectorias. Una *trayectoria* de un proceso estocástico es una aplicación de T en $\mathbb{R}$ de la forma $t \mapsto X_t(\omega)$, con $\omega \in \Omega$ fijo. Una pregunta fundamental en esta situación es: dado un RKHS $\mathcal{H}$ de funciones sobre el mismo conjunto T, ¿cuál es la probabilidad de que una trayectoria $X_\bullet(\omega)$ sea un elemento de $\mathcal{H}$? Este es un primer acercamiento a la cuestión inicial: distintos núcleos imponen distintas condiciones de regularidad sobre las funciones en $\mathcal{H}$, por lo que esta pregunta se puede traducir en determinar la probabilidad de que una trayectoria del proceso tenga cierto tipo de regularidad.

M. Driscoll da una primera respuesta a este problema en 1973, en su trabajo pionero [5]: con ciertas hipótesis sobre el proceso X, se puede ver que esta probabilidad es necesariamente 0 ó 1, resultado que se conoce como la ley cero-uno de Driscoll. En [11], M. Lukić y J. Beder logran avances en esta línea, encontrando condiciones necesarias y suficientes (sobre el proceso X y los núcleos reproductores involucrados) para que las trayectorias de X pertenezcan a $\mathcal{H}$ con probabilidad 1. Notablemente, ellos logran debilitar aún más las suposiciones hechas por Driscoll en su trabajo original, logrando probar sus resultados para procesos estocásticos arbitrarios (en su trabajo, Driscoll trabaja únicamente con procesos gaussianos definidos sobre espacios métricos separables y con función de covarianza continua). Otro de los objetivos del presente trabajo es dar una demostración del resultado de Lukić y Beder (ver Teorema 2.4.1).

Recientemente, se ha comenzado a explorar el marco más general de los *espacios de Banach con núcleo reproductor* (RKBS, por sus siglas en inglés) y su relación con el modelado de redes neuronales de ancho arbitrario (es decir, sin ancho prefijado). Un RKBS es, esencialmente, un RKHS, pero donde se relaja la hipótesis de que la norma provenga de un producto interno. En el trabajo [2] de 2023, los autores muestran una familia de RKBSs definidos mediante integrales abstractas, que generalizan a las redes neuronales clásicas. Más aún, en ese contexto abstracto logran probar resultados de optimización convexa íntimamente relacionados con los problemas de aprendizaje automático. Un tercer objetivo del presente trabajo es dar un recorrido de estos resultados.

El trabajo se estructurará de la siguiente manera: en el Capítulo 1 desarrollaremos la teoría de los RKHSs que vertebrará todo el trabajo. Pondremos el foco en los resultados generales y, hacia el final del capítulo, en la Sección 1.4, introduciremos algunos conceptos necesarios para atacar la ley cero-uno de Driscoll. El Capítulo 2 está destinado a demostrar la generalización de la ley cero-uno de Driscoll. Para ello, introduciremos los procesos estocásticos como objetos de estudio, y luego daremos un recorrido por la teoría de probabilidades en espacios de Hilbert, para tener todas las herramientas técnicas necesarias. Al final del capítulo, combinaremos las nociones probabilísticas con los RKHSs para llegar al resultado prometido. En el Capítulo 3 ampliaremos la noción de RKHS para comenzar a estudiar los RKBSs. Daremos una introducción a las redes neuronales clásicas y veremos cómo los RKBSs permiten modelar una extensión de ellas. Concluiremos el capítulo con algunos resultados relacionados con el estudio de problemas de optimización en estos espacios. Por último, en el Apéndice A demostraremos algunos resultados clásicos sobre puntos extremales utilizados en el Capítulo 3.

Notación

A lo largo del trabajo, consideraremos espacios vectoriales sobre los números reales $\mathbb{R}$ o sobre los números complejos $\mathbb{C}$. Denotaremos por $\mathbb{K}$ a un cuerpo que puede ser cualquiera de ambos.

Si X es un espacio vectorial y $\{x_i\}_{i\in I}$ es una familia de vectores de X, denotaremos $[x_i]_{i\in I}$ o $[x_i : i \in I]$ al subespacio generado. Si X es un espacio de Banach, denotaremos X' a su espacio dual (continuo). La misma notación se usará si X es un espacio vectorial topológico. También notaremos $\overline{B}_X := \{x \in X : ||x||_X \leq 1\}$, $B_X := \{x \in X : ||x||_X < 1\}$ y $S_X := \{x \in X : ||x||_X = 1\}$.

Si $\mathcal{H}$ es un espacio de Hilbert y $S \subseteq \mathcal{H}$ es un subespacio, denotamos por $S^\perp$ a su complemento ortogonal. En ocasiones, será necesario especificar dentro de qué espacio se toma el complemento ortogonal, en cuyo caso también notaremos $\mathcal{H} \ominus S := S^\perp$ según convenga.

Cuando T sea un conjunto, para $s, t \in T$ denotamos $\delta_{s,t}$ a la *delta de Kronecker* definida por

$$\delta_{s,t} = \begin{cases} 1 & \text{si } s = t \\ 0 & \text{si } s \neq t \end{cases}.$$

Capítulo 1

Espacios de Hilbert con Núcleo Reproductor

En este capítulo establecemos los cimientos de la teoría de espacios de Hilbert con núcleo reproductor, que vertebra todo el trabajo. La Sección 1.1 está destinada a introducir las definiciones de espacio de Hilbert con núcleo reproductor y de su núcleo asociado, junto con ejemplos de estos conceptos, provenientes principalmente del análisis. En la Sección 1.2 demostramos los primeros resultados de la teoría, culminando en el teorema de Moore que establece una correspondencia biunívoca entre espacios de Hilbert con núcleo reproductor en un conjunto y núcleos definidos en ese mismo conjunto. La referencia principal para estas secciones es [13]. En la Sección 1.3 estudiamos cómo cambian los núcleos de los espacios bajo distintas operaciones conjuntistas y viceversa, apoyándonos principalmente en [1]. Por último, la Sección 1.4 está dedicada a complementar las secciones anteriores con los resultados necesarios para abordar la generalización de ley cero-uno de Driscoll en el Capítulo 2, tomados casi en su totalidad de [11, Section 4].

1.1. Primeras definiciones y ejemplos

Dado un conjunto T, denotamos por $\mathbb{K}^T$ al conjunto de funciones de T en $\mathbb{K}$. En este conjunto, consideramos la estructura de $\mathbb{K}-$espacio vectorial dada por la suma y producto por escalares punto a punto.

Definición 1.1.1. Sea T un conjunto. Diremos que un subespacio vectorial $\mathcal{H} \subseteq \mathbb{K}^T$ con un producto interno $\langle \cdot, \cdot \rangle$ es un **espacio de Hilbert con núcleo reproductor** (abreviado **RKHS** por sus siglas en inglés) en T si cumple que

(i) $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ es un espacio de Hilbert;

(ii) para cada $t \in T$, el funcional de evaluación $E_t : \mathcal{H} \to \mathbb{K}$ dado por $E_t f = f(t)$ resulta acotado.

Si $\mathcal{H}$ es un RKHS en T, por el teorema de representación de Riesz vale que para cada $t \in T$, existe $K_t \in \mathcal{H}$ tal que $E_t = \langle \cdot, K_t \rangle$. Es decir, para toda $f \in \mathcal{H}$ se cumple que

$$f(t) = E_t f = \langle f, K_t \rangle.$$

Definición 1.1.2. La función $K_t \in \mathcal{H}$ se llama **núcleo reproductor para el punto** t. La función $K : T \times T \to \mathbb{K}$ dada por

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

se llama el **núcleo reproductor de** $\mathcal{H}$.

Notemos que $K(s,t) = K_t(s) = \langle K_t, K_s \rangle$ y, por lo tanto,

$$K(s,t) = \langle K_t, K_s \rangle = \overline{\langle K_s, K_t \rangle} = \overline{K(t,s)}$$

en el caso complejo y $K(s,t) = K(t,s)$ en el caso real. También vale que

$$||E_t||^2 = ||K_t||^2 = \langle K_t, K_t \rangle = K(t,t). \tag{1.1}$$

Ejemplo 1.1.3. Sea $T = \{1, \dots, n\}$ e identifiquemos $\mathbb{C}^n$ con el conjunto de funciones de T en $\mathbb{C}$. Con el producto interno canónico, $\mathbb{C}^n$ es un espacio de Hilbert. Además, si $1 \leq i \leq n$, el funcional de evaluación en i no es otra cosa que la proyección a la $i-$ésima coordenada, que es acotada. Entonces $\mathbb{C}^n$ resulta ser un RKHS en T. Si llamamos K al núcleo reproductor, para $x \in \mathbb{C}^n$ tenemos que $\langle x, K_i \rangle = x_i = \langle x, e_i \rangle$, con e_i el $i-$ésimo vector canónico. Por lo tanto, $K_i = e_i$ y luego

$$K(i,j) = \langle K_j, K_i \rangle = \langle e_j, e_i \rangle = \delta_{ij}.$$

En la Proposición 1.3.5 se caracteriza una familia amplia de RKHSs definidos sobre conjuntos finitos.

Ejemplo 1.1.4 (un espacio de funciones que no es un RKHS)**.** Consideremos en el espacio $C([0,1])$ de funciones continuas $f : [0,1] \to \mathbb{R}$ el producto interno dado por

$$\langle f, g \rangle = \int_0^1 f(x) g(x) dx.$$

Si bien este espacio no es completo, los funcionales de evaluación están bien definidos. En caso de ser acotados, podrían extenderse a la completación de $C([0,1])$, que no es otra cosa que el espacio $L^2([0,1])$, lo cual resultaría en un RKHS en el intervalo $[0,1]$. Sin embargo, esto no es así. En efecto, para $0 < t < 1$ fijo, consideremos para cada $n \in \mathbb{N}$ la función

$$f_n(x) = \begin{cases} \left(\frac{x}{t}\right)^n & \text{si } x \leq t \\ \left(\frac{1-x}{1-t}\right)^n & \text{si } x \geq t \end{cases}.$$

Se puede ver que $f_n(t) = 1$ para todo $n \in \mathbb{N}$, pero que $f_n \xrightarrow{||\cdot||} 0$, por lo que el funcional de evaluación E_t no es acotado. Similarmente, se puede ver que los funcionales E_0 y E_1 tampoco son acotados.

Tiene sentido que este ejemplo falle: como el espacio $L^2([0,1])$ es un conjunto de clases de equivalencia en casi todo punto de funciones, los funcionales de evaluación no están bien definidos pues los conjuntos unipuntuales tienen medida 0, así que no hay forma de darle una estructura de RKHS en el intervalo $[0,1]$. Citando una frase recurrente de D. G. Carando: "Gracias al teorema de aplicación abierta, en líneas generales, un operador entre espacios de Banach es acotado si y sólo si está bien definido." Recordemos que $C([0,1])$ no era completo con el producto interno dado.

En [14], I. Steinwart prueba que si T es un espacio métrico compacto no contable, entonces no existe ningún RKHS en T que contenga a $C(T)$.

Ejemplo 1.1.5 (espacio de Sobolev en el intervalo)**.** Sea $\mathcal{H}$ el espacio de funciones $f : [0,1] \to \mathbb{R}$ absolutamente continuas tales que $f(0) = f(1) = 0$ y $f' \in L^2([0,1])$. El producto interno en $\mathcal{H}$ está dado por

$$\langle f, g \rangle = \int_0^1 f'(x)g'(x)dx.$$

Dados $t \in [0,1]$ y $f \in \mathcal{H}$, por la desigualdad de Hölder tenemos que

$$\begin{aligned} |f(t)| &= \left| \int_0^t f'(x)dx \right| \le \int_0^t |f'(x)|dx = \int_0^1 \left|f'(x)\chi_{[0,t]}(x)\right|dx \\ &\le \left(\int_0^1 |f'(x)|^2 dx \right)^{1/2} \left(\int_0^1 |\chi_{[0,t]}(x)|^2 dx \right)^{1/2} \\ &= \sqrt{t}||f||_{\mathcal{H}}. \end{aligned}$$

Por lo tanto, el funcional de evaluación en t es acotado con $||E_t|| \le \sqrt{t}$.

Para ver la completitud, notemos que si $(f_n)_{n\in\mathbb{N}}$ es una sucesión de Cauchy en $\mathcal{H}$, entonces por la desigualdad anterior, debe ser puntualmente de Cauchy. Luego converge puntualmente, digamos a una función f. Por otra parte, la sucesión $(f_n')_{n\in\mathbb{N}}$ es de Cauchy en $L^2([0,1])$, así que converge en $||\cdot||_{L^2}$ a una cierta función $g \in L^2([0,1])$. Como $||\cdot||_{L^1} \le ||\cdot||_{L^2}$, tenemos que para $t \in [0,1]$:

$$f(t) = \lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} \int_0^t f_n'(x)dx = \int_0^t g(x)dx,$$

de donde se ve que f es absolutamente continua con $f' = g \in L^2([0,1])$. De la continuidad de los funcionales E_0 y E_1 se desprende que $f(0) = f(1) = 0$ por lo que $f \in \mathcal{H}$ y $f_n \xrightarrow{\mathcal{H}} f$.

Calculemos el núcleo reproductor de $\mathcal{H}$ en $t \in [0,1]$. Si $f \in \mathcal{H}$, entonces integrando formalmente por partes,

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

Si denotamos δ_t a la delta de Dirac, entonces podemos pensar que $f(t) = \int_0^1 f(x)\delta_t(x)dx$. Por lo tanto, K_t debe satisfacer (al menos formalmente) el problema de contorno

$$\begin{cases} -K_t'' = \delta_t \\ K_t(0) = K_t(1) = 0 \end{cases}.$$

Integrando dos veces y sustituyendo las condiciones de contorno, se llega a la expresión

$$K_t(x) = \begin{cases} (1-t)x & \text{si } x \leq t \\ (1-x)t & \text{si } x \geq t \end{cases}.$$

La función K_t así definida es efectivamente un elemento de $\mathcal{H}$, y para $f \in \mathcal{H}$,

$$\begin{aligned} \langle f, K_t \rangle &= \int_0^1 f'(x) K_t'(x) dx = \int_0^t f'(x)(1-t) dx + \int_t^1 f'(x)(-t) dx \\ &= (1-t)f(t) + t f(t) \\ &= f(t). \end{aligned}$$

Como se ve de este ejemplo, en general no es fácil caracterizar el núcleo reproductor de un RKHS. Sin embargo, tener una tal caracterización proporciona ventajas enormes. Una de ellas es la posibilidad de calcular de forma exacta la norma de los funcionales de evaluación. Para probar que $\mathcal{H}$ era un RKHS, dimos la estimación $||E_t|| \leq \sqrt{t}$. Conociendo explícitamente el núcleo, podemos decir aún más: gracias a (1.1) tenemos $||E_t|| = \sqrt{K(t,t)} = \sqrt{t(1-t)}$. Sin la ayuda de esta teoría, no era para nada evidente que la estimación anterior se podía mejorar. Pero por (1.1), el cálculo de $||E_t||$ se redujo a una simple sustitución.

1.2. Resultados generales

A continuación se demuestran algunos resultados básicos (pero fundamentales) de la teoría de RKHSs.

Proposición 1.2.1. *Sea $\mathcal{H}$ un RKHS en un conjunto T con núcleo K. Entonces el subespacio generado por las funciones $K_t = K(\cdot, t)$ es denso en $\mathcal{H}$.*

Demostración. Supongamos que una función $f \in \mathcal{H}$ es ortogonal a todas las funciones K_t. Luego, para cada $t \in T$ se tiene que

$$0 = \langle f, K_t \rangle = f(t),$$

de donde concluimos que $f = 0$, lo cual termina la prueba. □

La siguiente proposición establece que un RKHS queda totalmente determinado por su núcleo reproductor. Este resultado es la mitad de la correspondencia mencionada al principio del capítulo.

Proposición 1.2.2. *Sean $\mathcal{H}_1, \mathcal{H}_2$ dos RKHSs en un conjunto T con núcleos K_1, K_2 respectivamente. Llamemos $||\cdot||_i$ a la norma del espacio $\mathcal{H}_i$. Si $K_1(s,t) = K_2(s,t)$ para cualesquiera $s, t \in T$, entonces $\mathcal{H}_1 = \mathcal{H}_2$ y $||f||_1 = ||f||_2$ para toda $f \in \mathcal{H}_1 = \mathcal{H}_2$.*

Demostración. Llamemos $K = K_1 = K_2$, $K_t = K(\cdot, t)$ y $\mathcal{W} = [K_t]_{t \in T} \subseteq \mathcal{H}_1 \cap \mathcal{H}_2$. Por la proposición anterior, sabemos que $\mathcal{W}$ es denso en $\mathcal{H}_1$ y en $\mathcal{H}_2$ con sus respectivas normas.

Si $f = \sum_i \alpha_i K_{t_i} \in \mathcal{W}$, podemos decir que

$$||f||_1^2 = \sum_{i,j} \alpha_i \bar{\alpha}_j \langle K_{t_i}, K_{t_j} \rangle_{\mathcal{H}_1} = \sum_{i,j} \alpha_i \bar{\alpha}_j K(t_j, t_i) = \sum_{i,j} \alpha_i \bar{\alpha}_j \langle K_{t_i}, K_{t_j} \rangle_{\mathcal{H}_2} = ||f||_2^2,$$

de donde $||f||_1 = ||f||_2$.

Veamos ahora la igualdad de conjuntos: tomemos $f \in \mathcal{H}_1$ y veamos que $f \in \mathcal{H}_2$.

Como $\mathcal{W}$ es denso en $\mathcal{H}_1$, existe una sucesión $(f_n)_{n \in \mathbb{N}} \subseteq \mathcal{W}$ tal que $f_n \xrightarrow{||\cdot||_1} f$. Como $(f_n)_{n \in \mathbb{N}}$ es de Cauchy en $||\cdot||_1$ y las normas de ambos espacios coinciden en $\mathcal{W}$, $(f_n)_{n \in \mathbb{N}}$ es de Cauchy en $||\cdot||_2$, por lo que existe $g \in \mathcal{H}_2$ tal que $f_n \xrightarrow{||\cdot||_2} g$. Ahora bien, dado $t \in T$, el funcional de evaluación E_t es acotado en $\mathcal{H}_1$. Luego la convergencia en $||\cdot||_1$ nos permite afirmar que $f_n(t) \to f(t)$. Por el mismo motivo, podemos también decir que $f_n(t) \to g(t)$ y en consecuencia $f(t) = g(t)$. Como $t \in T$ era arbitrario, concluimos que $f = g \in \mathcal{H}_2$. Esto termina de probar que $\mathcal{H}_1 \subseteq \mathcal{H}_2$ y simétricamente se prueba la otra inclusión.

Por último, para $f \in \mathcal{H}_1 = \mathcal{H}_2$ tomemos $(f_n)_{n \in \mathbb{N}} \subseteq \mathcal{W}$ tal que que $f_n \xrightarrow{||\cdot||_1} f$. Entonces vale que $||f_n||_1 \to ||f||_1$. Ahora, reproduciendo el argumento anterior se ve que $f_n \xrightarrow{||\cdot||_2} f$ y luego también vale que $||f_n||_2 \to ||f||_2$. Como $||f_n||_1 = ||f_n||_2$, concluimos que $||f||_1 = ||f||_2$. □

Antes de seguir nuestra discusión sobre RKHSs, necesitamos introducir un poco de terminología.

Sea $\{h_s : s \in S\}$ una familia de vectores en un espacio normado $\mathcal{H}$ indexada por un conjunto S arbitrario. Decimos que

$$h = \sum_{s \in S} h_s$$

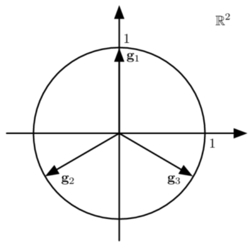

Figura 1.1: El frame de Mercedes-Benz.

si para todo $\varepsilon > 0$, existe $F_0 \subseteq S$ finito tal que para todo $F \subseteq S$ finito que contiene a F_0 vale que $||h - \sum_{s\in F} h_s|| < \varepsilon$. Esto se puede interpretar de la siguiente forma: $\Lambda = \{F \subseteq S : F \text{ finito}\}$ es un conjunto dirigido cuando se lo ordena por inclusión. Si llamamos $x_F = \sum_{s\in F} h_s$, decir que $h = \sum_{s\in S} h_s$ es equivalente a decir que la red $(x_F)_{F\in\Lambda}$ converge a h.

Definición 1.2.3. Sea $\mathcal{H}$ un espacio de Hilbert con producto interno $\langle\cdot,\cdot\rangle$. Una familia de vectores $\{f_s : s \in S\} \subseteq \mathcal{H}$ se dice un **frame de Parseval** si para cada $f \in \mathcal{H}$ vale que

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

Notemos que, en general, un frame de Parseval no tiene por qué ser linealmente independiente. Por ejemplo, nada de la definición impide que el 0 sea un elemento de un frame de Parseval. El siguiente ejemplo muestra un frame de Parseval que es "naturalmente" linealmente dependiente.

Ejemplo 1.2.4. Identificando $\mathbb{C} = \mathbb{R}^2$, consideramos en $\mathbb{C}$ el producto interno canónico dado por $\langle x_1 + ix_2, y_1 + iy_2\rangle = x_1y_1 + x_2y_2$. Sea $\xi = e^{2\pi i/3}$ y consideremos el conjunto $\mathcal{F} = \{1, \xi, \xi^2\}$. Si $z = z_1 + iz_2 \in \mathbb{C}$, entonces

$$\begin{aligned}
|\langle z,1\rangle|^2 + |\langle z,\xi\rangle|^2 + |\langle z,\xi^2\rangle|^2 &= z_1^2 + \left(z_1\left(-\frac{1}{2}\right) + z_2\frac{\sqrt{3}}{2}\right)^2 + \left(z_1\left(-\frac{1}{2}\right) - z_2\frac{\sqrt{3}}{2}\right)^2 \\
&= z_1^2 + \frac{1}{4}z_1^2 + \frac{3}{4}z_2^2 - \frac{\sqrt{3}}{2}z_1z_2 + \frac{1}{4}z_1^2 + \frac{3}{4}z_2^2 + \frac{\sqrt{3}}{2}z_1z_2 \\
&= \frac{3}{2}z_1^2 + \frac{3}{2}z_2^2 \\
&= \frac{3}{2}|z|^2.
\end{aligned}$$

Si llamamos $c = \sqrt{2/3}$, por el cálculo anterior vale que $\mathcal{F}' = \{c, c\xi, c\xi^2\}$ es un frame de Parseval para $\mathbb{C}$, que debe ser linealmente dependiente por cuestiones dimensionales.

Como los frames de Parseval se preservan por isometrías, $i\mathcal{F}' = \{ic, ic\xi, ic\xi^2\}$ también es un frame de Parseval. A este frame se lo denomina comúnmente *frame de Mercedes-Benz* por motivos razonablemente obvios (ver Figura 1.1).

Los frames de Parseval (y los frames en general) son objetos centrales en la teoría de operadores y en la teoría de muestreo. El artículo [9] presenta las ideas generales detrás de los frames, junto con un poco de su historia y las preguntas actuales alrededor de ellos. En este trabajo nos limitaremos a exponer algunas generalidades sobre ellos, para luego relacionarlos con los RKHSs.

Proposición 1.2.5. *Sea $\mathcal{H}$ un espacio de Hilbert, $\{f_s\}_{s\in S} \subseteq \mathcal{H}$. Las siguientes afirmaciones son equivalentes.*

(1) $\{f_s\}_{s\in S}$ es un frame de Parseval.

(2) La función $V : \mathcal{H} \to \ell^2(S)$ dada por $Vh = (\langle h, f_s\rangle)_{s\in S}$ es una isometría bien definida.

(3) Para todo $h \in \mathcal{H}$ se tiene que $h = \sum_{s\in S}\langle h, f_s\rangle f_s$.

(4) Para $h_1, h_2 \in \mathcal{H}$ tenemos que $\langle h_1, h_2\rangle = \sum_{s\in S}\langle h_1, f_s\rangle\langle f_s, h_2\rangle$.

Demostración.

(1) $\implies$ *(2)* Como $\{f_s\}_{s\in S}$ es un frame de Parseval, sabemos que para $h \in \mathcal{H}$ vale:

$$||h||^2_{\mathcal{H}} = \sum_{s\in S} |\langle h, f_s\rangle|^2 = ||(\langle h, f_s\rangle)_{s\in S}||^2_{\ell^2(S)} = ||Vh||^2_{\ell^2(S)},$$

por lo que V resulta estar bien definida y es una isometría.

(2) $\implies$ *(3)* Si denotamos $\{e_s\}_{s\in S} \subseteq \ell^2(S)$ a la base canónica, podemos observar que para $h \in \mathcal{H}$ se tiene

$$\langle h, V^*e_s\rangle = \langle Vh, e_s\rangle = \langle h, f_s\rangle,$$

por lo que $V^*e_s = f_s$. Como V es una isometría, $V^*V = I_{\mathcal{H}}$. Entonces,

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

(3) $\implies$ *(4)* Escribiendo $h_1 = \sum_{s\in S}\langle h_1, f_s\rangle f_s$ tenemos que

$$\langle h_1, h_2\rangle = \left\langle \sum_{s\in S}\langle h_1, f_s\rangle f_s, h_2 \right\rangle = \sum_{s\in S}\langle h_1, f_s\rangle\langle f_s, h_2\rangle.$$

(4) $\implies$ *(1)* Para $h \in \mathcal{H}$,

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

lo cual termina la prueba. □

Observación 1.2.6. Sea $\mathcal{H}$ es un espacio de Hilbert con base ortonormal $\{e_s\}_{s\in S}$. Si $\mathcal{H}_0 \subseteq \mathcal{H}$ es un subespacio cerrado y $P : \mathcal{H} \to \mathcal{H}_0$ es la proyección ortogonal sobre $\mathcal{H}_0$, entonces $\{Pe_s\}_{s\in S}$ es un frame de Parseval para $\mathcal{H}_0$. En efecto, si $h \in \mathcal{H}_0$, entonces

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

La siguiente proposición dice que de hecho vale la vuelta.

Proposición 1.2.7 (Larson)**.** *Sea $\{f_s\}_{s\in S}$ un frame de Parseval para un espacio de Hilbert $\mathcal{H}$. Entones existe un espacio de Hilbert $\mathcal{K}$ que contiene a $\mathcal{H}$ y una base ortonormal $\{e_s\}_{s\in S}$ de $\mathcal{K}$ tal que para todo $s \in S$ se tiene $f_s = Pe_s$, donde $P : \mathcal{K} \to \mathcal{H}$ es la proyección ortogonal.*

Demostración. Sea $\mathcal{K} = \ell^2(S)$ y $V : \mathcal{H} \to \ell^2(S)$ la isometría de la Proposición 1.2.5. Vía esta isometría, identificamos a $\mathcal{H}$ con el subespacio $V(\mathcal{H})$ de $\ell^2(S)$.

Con esta identificación, tenemos que $V^* = P$ pues para $k \in \mathcal{K}, h \in \mathcal{H}$ vale

$$\langle Vh, k\rangle = \langle h, k\rangle = \langle Ph, k\rangle = \langle h, Pk\rangle,$$

y por lo tanto, al igual que en la Proposición 1.2.5, $Pe_s = V^*e_s = f_s$ para todo $s \in S$. $\square$

El siguiente teorema da una condición necesaria y suficiente para que una familia de vectores en un RKHS sea un frame de Parseval.

Teorema 1.2.8 (Papadakis)**.** *Sea $\mathcal{H}$ un RKHS en un conjunto T con núcleo K y sea $\{f_s\}_{s\in S} \subseteq \mathcal{H}$. Luego, $\{f_s\}_{s\in S}$ es un frame de Parseval si y sólo si para todos $t, t' \in T$ vale que*

$$K(t, t') = \sum_{s\in S} f_s(t)\overline{f_s(t')}.$$

Demostración.

($\implies$) Como $\{f_s\}_{s\in S}$ es un frame de Parseval, por la Proposición 1.2.5 tenemos que

$$K(t, t') = \langle K_{t'}, K_t\rangle = \sum_{s\in S}\langle K_{t'}, f_s\rangle\langle f_s, K_t\rangle = \sum_{s\in S}\langle f_s, K_t\rangle\overline{\langle f_s, K_{t'}\rangle} = \sum_{s\in S} f_s(t)\overline{f_s(t')}.$$

($\impliedby$) Sea $h = \sum_j \alpha_j K_{t_j}$ una combinación lineal finita de las funciones K_{t_j}. Entonces,

$$\begin{aligned} ||h||^2 &= \langle h, h\rangle = \sum_{i,j}\alpha_i\bar{\alpha}_j\langle K_{t_i}, K_{t_j}\rangle = \sum_{i,j}\alpha_i\bar{\alpha}_j K(t_j, t_i) = \sum_{i,j}\alpha_i\bar{\alpha}_j\sum_{s\in S} f_s(t_j)\overline{f_s(t_i)} \\ &= \sum_{i,j}\alpha_i\bar{\alpha}_j\sum_{s\in S}\langle f_s, K_{t_j}\rangle\overline{\langle f_s, K_{t_i}\rangle} = \sum_{s\in S}\left\langle f_s, \sum_j \alpha_j K_{t_j}\right\rangle\left\langle \sum_i \alpha_i K_{t_i}, f_s\right\rangle \\ &= \sum_{s\in S}|\langle h, f_s\rangle|^2. \end{aligned}$$

Si llamamos $\mathcal{L}$ al subespacio generado por las funciones K_t, el cálculo anterior nos permite decir que el operador $V : \mathcal{L} \to \ell^2(S)$ dado por $Vh = (\langle h, f_s\rangle)_{s\in S}$ es una isometría bien definida. Como $\mathcal{L}$ es denso en $\mathcal{H}$, V se extiende a una isometría $V : \mathcal{H} \to \ell^2(S)$, probando así que $\{f_s\}_{s\in S}$ es un frame de Parseval. $\square$

A continuación, daremos las definiciones necesarias para el enunciado del teorema de Moore.

Recordemos que para una matriz $A \in \mathbb{C}^{n\times n}$, las siguientes condiciones son equivalentes:

(i) $A = A^*$ y todos los autovalores de A son no negativos;

(ii) para todo $\alpha \in \mathbb{C}^n$ vale que $\alpha^* A\alpha = \sum_{i,j} \alpha_i \bar{\alpha}_j A_{ij} \geq 0$.

Si además $A \in \mathbb{R}^{n\times n}$, las condiciones anteriores son equivalentes a que A sea simétrica y que para todo $\alpha \in \mathbb{R}^n$ valga que $\alpha^T A\alpha \geq 0$. Diremos que A es una **matriz positiva** y notaremos $A \geq 0$ cuando A esté en las condiciones recién mencionadas. Notar que esto es consistente con la noción de operador positivo en un espacio de Hilbert.

Definición 1.2.9. Sea T un conjunto y $K : T \times T \to \mathbb{K}$ una función. Diremos que K es un **núcleo** en T si para todo $n \in \mathbb{N}$ y toda elección de n puntos distintos $t_1, \ldots, t_n \in T$ vale que la matriz $(K(t_i, t_j))_{i,j}$ es positiva.

Proposición 1.2.10. *Sea $\mathcal{H}$ un RKHS en un conjunto T con núcleo reproductor K. Entonces K es un núcleo en T.*

Demostración. Dados $t_1, \ldots, t_n \in T$, tenemos que $K(t_i, t_j) = \langle K_{t_j}, K_{t_i}\rangle$. En particular, la matriz $(K(t_i, t_j))_{i,j}$ es hermitiana (o simétrica en el caso real). Si $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$, entonces

$$\sum_{i,j} \alpha_j \bar{\alpha}_i K(t_i, t_j) = \sum_{i,j} \alpha_j \bar{\alpha}_i \langle K_{t_j}, K_{t_i}\rangle = \left\langle \sum_j \alpha_j K_{t_j}, \sum_i \alpha_i K_{t_i} \right\rangle \geq 0,$$

lo cual termina la prueba. □

El teorema de Moore es simplemente la recíproca de esta proposición.

Teorema 1.2.11 (Moore)**.** *Sea T un conjunto y $K : T\times T \to \mathbb{K}$ un núcleo en T. Entonces existe un RKHS en T cuyo núcleo reproductor es K.*

Demostración. Para cada $t \in T$, sea $K_t : T \to \mathbb{K}$ la función dada por $K_t(s) = K(s, t)$. Notemos que si K es el núcleo de un RKHS $\mathcal{H}$ en T, éstas funciones deben generar un subespacio denso en $\mathcal{H}$. Luego, la idea es definir un producto interno en este subespacio y luego completarlo para obtener un espacio de Hilbert.

Sea $\mathcal{W} = [K_t]_{t\in T} \subseteq \mathbb{K}^T$ y sea $B : \mathcal{W} \times \mathcal{W} \to \mathbb{K}$ definida por

$$B\left(\sum_i \alpha_i K_{t_i}, \sum_j \beta_j K_{s_j} \right) := \sum_{i,j} \alpha_i \bar{\beta}_j K(s_j, t_i),$$

donde α_i, β_j son escalares. Como una función de $\mathcal{W}$ puede ser expresada de muchas formas distintas como combinación lineal de las funciones K_t, debemos ver que la definición de B no depende de esa escritura. Para ello, basta ver que si $f = \sum_i \alpha_i K_{t_i}$ es idénticamente nula como función en T, entonces $B(f,g) = B(g,f) = 0$ para toda $g \in \mathcal{W}$. Como las funciones K_t generan al subespacio $\mathcal{W}$, basta ver la igualdad anterior para $g = K_t$. En este caso, tenemos que

$$B(f, K_t) = \sum_i \alpha_i K(t, t_i) = f(t) = 0,$$

pues f era la función nula. Simétricamente se ve la otra igualdad.

Una vez probada la buena definición, es inmediato ver que B resulta sesquilineal y hermitiana. Además, para $f = \sum_i \alpha_i K_{t_i}$ tenemos que

$$B(f,f) = \sum_{i,j} \alpha_i \bar{\alpha}_j K(x_j, x_i) \geq 0$$

por definición de núcleo. Esto nos dice que B es un producto semiinterno en $\mathcal{W}$. Por último, si $B(f,f) = 0$, entonces por la desigualdad de Cauchy-Schwarz para productos semiinternos vale que $B(f,g) = 0$ para toda $g \in W$. En particular, tomando $g = K_t$ tenemos que $0 = B(f, K_t) = f(t)$, probando así que $f = 0$. Podemos afirmar entonces que B es un producto interno en $\mathcal{W}$.

Ahora que $\mathcal{W}$ es un espacio vectorial con producto interno, podemos considerar $\mathcal{H}$ una completación de $\mathcal{W}$. Queremos identificar $\mathcal{H}$ con un subespacio adecuado de $\mathbb{K}^T$. Para ello, definimos para cada $h \in \mathcal{H}$ la función $\hat{h} : T \to \mathbb{K}$ dada por

$$\hat{h}(t) = \langle h, K_t \rangle_{\mathcal{H}}.$$

Esto define una transformación lineal $L : \mathcal{H} \to \mathbb{K}^T$ dada por $Lh = \hat{h}$. Además, si $Lh = 0$, entonces $\langle h, K_t \rangle_{\mathcal{H}} = 0$ para todo $t \in T$. Como las funciones K_t generan un subespacio denso en $\mathcal{H}$, esto implica que $h = 0$. Es decir, L es inyectiva.

Si $\hat{\mathcal{H}} = \operatorname{Im}(L)$, entonces L da un isomorfismo lineal entre $\mathcal{H}$ y $\hat{\mathcal{H}}$. Definiendo en $\hat{\mathcal{H}}$ el producto interno $\langle \hat{h}_1, \hat{h}_2 \rangle_{\hat{\mathcal{H}}} = \langle h_1, h_2 \rangle_{\mathcal{H}}$, L resulta una isometría entre $\mathcal{H}$ y $\hat{\mathcal{H}}$. Como $\mathcal{H}$ es un espacio de Hilbert, $\hat{\mathcal{H}}$ resulta un espacio de Hilbert con ese producto interno. Además, para $t \in T$ tenemos que

$$\hat{K}_t(s) = \langle K_t, K_s \rangle_{\mathcal{H}} = B(K_t, K_s) = K_t(s)$$

por lo que $\hat{K}_t = K_t$. Finalmente, si $\hat{h} \in \hat{\mathcal{H}}$ y $t \in T$, tenemos que

$$\langle \hat{h}, \hat{K}_t \rangle_{\hat{\mathcal{H}}} = \langle h, K_t \rangle_{\mathcal{H}} = \hat{h}(t)$$

por la propia definición de $\hat{h}$. Como $\hat{K}_t = K_t$, esto prueba que $\hat{\mathcal{H}}$ es un RKHS en T con núcleo reproductor K. □

Juntando el teorema de Moore (Teorema 1.2.11) con la Proposición 1.2.2, llegamos finalmente a la correspondencia prometida.

Si T es un conjunto, para cada núcleo K definido en T existe un único RKHS en T cuyo núcleo reproductor es K, al cual denotamos $\mathcal{H}(K)$.

En ocasiones, es necesario especificar el conjunto sobre el cual el RKHS está definido. En estos casos notaremos $\mathcal{H}(K,T)$ al RKHS definido sobre el conjunto T con núcleo K. Cuando haya riesgo de ambigüedad, notaremos $||\cdot||_K$ y $\langle\cdot,\cdot\rangle_K$ a la norma y producto interno del RKHS $\mathcal{H}(K)$.

Esta correspondencia entre RKHSs y núcleos es el corazón de la teoría. El resto de este capítulo y el siguiente se dedican a explotarla.

A continuación presentamos un ejemplo de cómo hallar el RKHS correspondiente a un núcleo dado.

Ejemplo 1.2.12. Sea $T = \mathbb{N}$ y consideramos en T el núcleo $K(m,n) = n^{-p}\delta_{m,n}$ con $p \geq 0$. Buscamos caracterizar $\mathcal{H}(K)$.

Si llamamos e_n al $n-$ésimo vector canónico de $\mathbb{K}^{\mathbb{N}}$, notemos que $K_n = n^{-p}e_n$. Ahora bien, dados $m,n \in \mathbb{N}$,

$$\langle K_n, K_m\rangle_K = K(m,n) = n^{-p}\delta_{m,n},$$

por lo que $\{n^{p/2}K_n\}_{n\in\mathbb{N}}$ es una sucesión ortonormal en $\mathcal{H}(K)$. Como esta sucesión genera un subespacio denso, entonces resulta una base ortonormal de $\mathcal{H}(K)$. Luego, para $a \in \mathcal{H}(K)$, por la identidad de Parseval,

$$\sum_{n=1}^{\infty} n^p|a_n|^2 = \sum_{n=1}^{\infty} |\langle a, n^{p/2}K_n\rangle_K|^2 = ||a||_K^2 < +\infty.$$

Recíprocamente, si $a \in \mathbb{K}^{\mathbb{N}}$ es una sucesión tal que $\sum_{n=1}^{\infty} n^p|a_n|^2 < +\infty$, tenemos que la serie $\sum_{n=1}^{\infty} a_n n^p K_n = \sum_{n=1}^{\infty} a_n e_n$ converge en $||\cdot||_K$ a un elemento $b \in \mathcal{H}(K)$. Como la convergencia en $||\cdot||_K$ implica convergencia puntual, obtenemos que para cada $n \in \mathbb{N}$, $a_n = b_n$ y luego $a = b \in \mathcal{H}(K)$. Juntando todo, llegamos a que

$$\mathcal{H}(K) = \left\{a \in \mathbb{K}^{\mathbb{N}} : \sum_{n=1}^{\infty} n^p|a_n|^2 < +\infty\right\}.$$

Como conocemos explícitamente una base ortonormal de $\mathcal{H}(K)$, no es difícil ver que el producto interno está dado por

$$\langle a, b\rangle_K = \sum_{n=1}^{\infty} n^p a_n \bar{b}_n.$$

1.3. Relación entre un núcleo y su RKHS asociado

Si bien es cierto que todo núcleo determina un único RKHS, en general es difícil caracterizar tanto la clase de funciones que pertenece a dicho RKHS, como el producto interno del espacio. A este proceso de, dado un núcleo K, hallar una descripción concreta de $\mathcal{H}(K)$, se lo conoce en la literatura como el *problema de la reconstrucción*. Los siguientes dos resultados avanzan en la dirección de la primera parte de este problema: caracterizar la clase de funciones que pertenece al RKHS definido por un núcleo.

Proposición 1.3.1. *Sean T un espacio topológico y K un núcleo en T. Consideramos en $T \times T$ la topología producto. Si $K : T \times T \to \mathbb{K}$ es continua, entonces la aplicación $K_\bullet : T \to \mathcal{H}(K)$ dada por $t \mapsto K_t$ es continua, y todas las funciones de $\mathcal{H}(K)$ también lo son.*

Demostración. Dados $s \in T$ y $\varepsilon > 0$, por la continuidad de K y la definición de topología producto, existe $U \subseteq T$ entorno abierto de s tal que para todo $(t, t') \in U \times U$ vale que $|K(t, t') - K(s, s)| < \varepsilon^2/3$. Luego, si $t \in U$,

$$\begin{aligned} ||K_t - K_s||^2 &= |\langle K_t - K_s, K_t - K_s\rangle| = |\langle K_t, K_t\rangle - \langle K_t, K_s\rangle - \langle K_s, K_t\rangle + \langle K_s, K_s\rangle| \\ &= |K(t,t) - K(s,t) - K(t,s) + K(s,s)| \\ &\leq |K(t,t) - K(s,s)| + |K(s,s) - K(s,t)| + |K(s,s) - K(t,s)| \\ &< \varepsilon^2. \end{aligned}$$

Es decir, $K_\bullet(U) \subseteq B_\varepsilon(K_s)$, lo cual termina de probar que $K_\bullet$ es continua.

Si $f \in \mathcal{H}(K)$ y $s, t \in T$, entonces

$$|f(s) - f(t)| = |\langle f, K_s - K_t\rangle| \leq ||f||\ ||K_s - K_t||, \tag{1.2}$$

y la continuidad de f es consecuencia de la continuidad $K_\bullet$. □

Notemos que de (1.2) se desprende que las funciones de $\mathcal{H}(K)$ heredan varias de las propiedades analíticas del núcleo K, como por ejemplo la continuidad uniforme, o la Lipschitz continuidad.

El siguiente resultado, debido a R. Fortet, da una caracterización completa de la clase de funciones que conforman un RKHS en términos de su núcleo reproductor. Fue publicado sin demostración en [7, Théorème (1,1)]. La prueba brindada es original de este trabajo (aunque probablemente no sea muy distinta de lo que tenía en mente Fortet).

Proposición 1.3.2 (Fortet)**.** *Sean T un conjunto, K un núcleo no nulo en T y $f : T \to \mathbb{K}$ una función arbitraria. Entonces $f \in \mathcal{H}(K)$ si y sólo si*

$$\sup \frac{\left|\sum_i \alpha_i \overline{f(t_i)}\right|^2}{\sum_{i,j} \alpha_i \overline{\alpha}_j K(t_j, t_i)} < +\infty, \tag{1.3}$$

donde el supremo se toma sobre todas las posibles elecciones de finitos $\alpha_i \in \mathbb{K}$ y $t_i \in T$ tales que el denominador de (1.3) es no nulo. Más aún, en este caso se tiene que (1.3) coincide con $||f||_K^2$.

Demostración. Llamemos $\mathcal{W} = [K_t]_{t\in T} \subseteq \mathcal{H}(K)$.

($\Longrightarrow$) Consideremos el funcional $\varphi : \mathcal{H}(K) \to \mathbb{K}$ dado por $\varphi(g) = \langle g, f\rangle_K$. Sabemos que

$$||f||_K^2 = ||\varphi||^2 = \sup_{g\in\mathcal{H}(K)\setminus\{0\}} \frac{|\varphi(g)|^2}{||g||_K^2} = \sup_{g\in\mathcal{H}(K)\setminus\{0\}} \frac{|\langle g, f\rangle_K|^2}{\langle g, g\rangle_K} = \sup_{g\in\mathcal{W}\setminus\{0\}} \frac{|\langle g, f\rangle_K|^2}{\langle g, g\rangle_K},$$

donde en la última igualdad usamos que $\mathcal{W} \subseteq \mathcal{H}(K)$ es denso. Si $g = \sum_i \alpha_i K_{t_i} \in \mathcal{W}\setminus\{0\}$, entonces

$$\langle g, f\rangle_K = \overline{\left\langle f, \sum_i \alpha_i K_{t_i}\right\rangle_K} = \sum_i \alpha_i \overline{\langle f, K_{t_i}\rangle_K} = \sum_i \alpha_i \overline{f(t_i)},$$

y por otra parte

$$\langle g, g\rangle_K = \left\langle \sum_i \alpha_i K_{t_i}, \sum_j \alpha_j K_{t_j}\right\rangle_K = \sum_{i,j} \alpha_i\overline{\alpha}_j \langle K_{t_i}, K_{t_j}\rangle_K = \sum_{i,j} \alpha_i\overline{\alpha}_j K(t_j, t_i).$$

Sustituyendo esto en lo anterior, obtenemos que

$$\sup \frac{\left|\sum_i \alpha_i \overline{f(t_i)}\right|^2}{\sum_{i,j} \alpha_i\overline{\alpha}_j K(t_j, t_i)} = ||f||_K^2 < +\infty.$$

($\Longleftarrow$) Definamos $\varphi : \mathcal{W} \to \mathbb{K}$ por $\varphi(K_t) = \overline{f(t)}$ y extendido por linealidad. Como $\{K_t\}_{t\in T}$ no tiene por qué ser linealmente independiente, debemos ver que φ está bien definido. Es decir, debemos ver que si $\sum_i \alpha_i K_{t_i}$ es la función nula, entonces $\sum_i \alpha_i \overline{f(t_i)} = 0$.

Argumentando por el absurdo, sea g la función nula y supongamos que podemos escribir a g como una combinación lineal no trivial $g = \sum_{i=1}^n \alpha_i K_{t_i}$, de manera tal que $\sum_{i=1}^n \alpha_i \overline{f(t_i)} \neq 0$. Como K es no nulo, existe $t_0 \in T$ tal que $K_{t_0} \neq 0$. Luego si $\varepsilon > 0$, vale que $g_\varepsilon = g + \varepsilon K_{t_0} = \varepsilon K_{t_0} \neq 0$. Consecuentemente, tenemos que

$$0 < \langle g_\varepsilon, g_\varepsilon\rangle_K = \varepsilon^2 \langle K_{t_0}, K_{t_0}\rangle_K = \varepsilon^2 K(t_0, t_0).$$

Por otra parte, llamando $\alpha_0 = \varepsilon$,

$$\langle g_\varepsilon, g_\varepsilon\rangle_K = \langle g + \varepsilon K_{t_0}, g + \varepsilon K_{t_0}\rangle_K = \left\langle \sum_{i=0}^n \alpha_i K_{t_i}, \sum_{j=0}^n \alpha_j K_{t_j}\right\rangle_K = \sum_{i,j=0}^n \alpha_i\overline{\alpha}_j K(t_j, t_i).$$

Juntando todo, esto quiere decir que

$$\sum_{i,j=0}^n \alpha_i\overline{\alpha}_j K(t_j, t_i) = \varepsilon^2 K(t_0, t_0) > 0.$$

Luego, podemos argumentar que

$$\frac{\left|\sum_{i=0}^n \alpha_i \overline{f(t_i)}\right|^2}{\sum_{i,j=0}^n \alpha_i\overline{\alpha}_j K(t_j, t_i)} = \frac{\left|\varepsilon\overline{f(t_0)} + \sum_{i=1}^n \alpha_i \overline{f(t_i)}\right|^2}{\varepsilon^2 K(t_0, t_0)} \xrightarrow{\varepsilon\to 0^+} +\infty$$

pues $\sum_{i=1}^{n} \alpha_i \overline{f(t_i)} \neq 0$, lo cual contradice la hipótesis (1.3). Esto termina de probar que φ está bien definido.

Tenemos entonces que $\varphi : \mathcal{W} \to \mathbb{K}$ es un funcional lineal bien definido. Más aún, φ es acotado por (1.3) y se extiende por densidad a un funcional acotado $\varphi : \mathcal{H}(K) \to \mathbb{K}$ con la misma norma. Por el teorema de representación de Riesz, existe $\hat{f} \in \mathcal{H}(K)$ tal que $\varphi = \langle \cdot, \hat{f} \rangle_K$. Finalmente, para $t \in T$ se tiene que

$$\hat{f}(t) = \langle \hat{f}, K_t \rangle = \overline{\langle K_t, \hat{f} \rangle} = \overline{\varphi(K_t)} = f(t),$$

de donde $f = \hat{f} \in \mathcal{H}(K)$. $\square$

El resto de resultados de esta sección están dirigidos a construir nuevos RKHSs a partir de espacios preexistentes, y a analizar cómo es la relación entre ellos.

1.3.1. Núcleos de subespacios cerrados

Proposición 1.3.3. *Sea $\mathcal{H}$ un RKHS en un conjunto T con núcleo reproductor K y sea $S \subseteq \mathcal{H}(K)$ un subespacio cerrado. Denotemos por $P : \mathcal{H} \to S$ a la proyección ortogonal sobre S. Entonces S es un RKHS en T con núcleo reproductor K^S, que satisface $K_t^S = PK_t$ para todo $t \in T$.*

Demostración. Como S es cerrado, es un espacio de Hilbert con el producto interno inducido por $\mathcal{H}$. Además, como los funcionales de evaluación de S son las restricciones a S de los funcionales de evaluación de $\mathcal{H}$, éstos resultan acotados. Por lo tanto, S es un RKHS en T.

Si $f \in S$ y $t \in T$, por ser P un operador autoadjunto vale que

$$f(t) = \langle f, K_t \rangle_{\mathcal{H}} = \langle Pf, K_t \rangle_{\mathcal{H}} = \langle f, PK_t \rangle_S,$$

y luego $K_t^S = PK_t$ como queríamos ver. $\square$

Corolario 1.3.4. *Sean $\mathcal{H}$ un RKHS en un conjunto T con núcleo K, $F \subseteq \mathcal{H}$ un subespacio cerrado y $H = F^{\perp}$. Llamemos K^F y K^H a los respectivos núcleos reproductores. Entonces, $K = K^F + K^H$.*

Demostración. Si P_F y P_H denotan a las proyecciones ortogonales sobre F y H respectivamente, por la Proposición 1.3.3 tenemos que, para $t \in T$,

$$K_t = P_F K_t + P_H K_t = K_t^F + K_t^H,$$

y por lo tanto $K = K^F + K^H$. $\square$

1.3.2. RKHS en un conjunto finito

Proposición 1.3.5. *Sea T un conjunto finito y sea K un núcleo en T. Identificamos a K con la matriz $(K(s,t))_{s,t\in T}$. Supongamos además que la matriz K es inversible. Entonces, $\mathcal{H}(K) = \mathbb{K}^T$ con el producto interno dado por*

$$\langle f, g\rangle_K = \sum_{s,t\in T} \overline{g(t)} K^{-1}(t,s) f(s).$$

Demostración. El subespacio $[K_t]_{t\in T}$ es denso en $\mathcal{H}(K)$ pero, como es de dimensión finita, también es cerrado en $\mathcal{H}(K)$. Luego, debe ser que $\mathcal{H}(K) = [K_t]_{t\in T} = \mathbb{K}^T$, ya que como la matriz K es inversible, las funciones K_t son linealmente independientes.

Si $g = \sum_t \alpha_t K_t$ entonces $g(s) = \langle g, K_s\rangle_K = \sum_t \alpha_t K(s,t)$. Matricialmente, esto se puede escribir como

$$(K(s,t))_{s,t\in T} \cdot (\alpha_t)_{t\in T} = (g(s))_{s\in S},$$

y luego $\alpha_t = \sum_s g(s) K^{-1}(t,s)$, de donde obtenemos que $g = \sum_{s,t\in T} K^{-1}(t,s) g(s) K_t$.

Finalmente, si $f, g \in \mathcal{H}(K)$ entonces

$$\begin{aligned}
\langle f, g\rangle_K &= \left\langle \sum_{s',t'\in T} K^{-1}(t',s') f(s') K_{t'}, \sum_{s,t\in T} K^{-1}(t,s) g(s) K_t \right\rangle \\
&= \sum_{s,s',t,t'} f(s') \overline{g(s)} K^{-1}(t',s') K(t,t') \overline{K^{-1}(t,s)} \\
&= \sum_{s,s',t} f(s') \overline{g(s)} K^{-1}(s,t) \underbrace{\sum_{t'} K(t,t') K^{-1}(t',s')}_{\delta_{t,s'}} \\
&= \sum_{s,t\in T} f(t) \overline{g(s)} K^{-1}(s,t),
\end{aligned}$$

como queríamos probar. $\square$

1.3.3. Restricción de un núcleo

Sean T un conjunto, K un núcleo definido en T y $T_0 \subseteq T$. La restricción $K|_{T_0\times T_0}$ de K a T_0 define un núcleo en T_0. En lo que sigue, buscamos entender el RKHS asociado a la restricción de K en términos de $\mathcal{H}(K)$.

Para no sobrecargar la notación, notaremos

$$\mathcal{H}(K, T_0) := \mathcal{H}(K|_{T_0\times T_0}).$$

Esto es consistente con la notación $\mathcal{H}(K,T) := \mathcal{H}(K)$ mencionada inmediatamente después de la demostración del teorema de Moore.

Proposición 1.3.6. *Sean T un conjunto, K un núcleo en T y $T_0 \subseteq T$. Si llamamos $\mathcal{H} = \mathcal{H}(K,T)$ y $\mathcal{H}_0 = \mathcal{H}(K,T_0)$, entonces $\mathcal{H}_0 = \{f|_{T_0} : f \in \mathcal{H}\}$ y la norma de $\mathcal{H}_0$ está dada por*

$$||g||_{\mathcal{H}_0} = \text{mín}\{||f||_{\mathcal{H}} : f \in \mathcal{H}, f|_{T_0} = g\}.$$

En particular, la aplicación $\{f \in \mathcal{H} : f|_{T_0} = 0\}^{\perp} \to \mathcal{H}_0$ dada por $f \mapsto f|_{T_0}$ es un isomorfismo isométrico.

Demostración. Sea

$$F := \{f \in \mathcal{H} : f|_{T_0} = 0\} = \bigcap_{t \in T_0} \ker E_t,$$

que resulta un subespacio cerrado de $\mathcal{H}$. Consideremos $H = F^{\perp}$. Tanto F como H son RKHSs en T pues son subespacios cerrados de $\mathcal{H}$. Si llamamos K^F y K^H a sus respectivos núcleos reproductores, por el Corolario 1.3.4 se tiene que $K = K^F + K^H$. En particular, notemos que si $t \in T_0$ entonces $K^F_t(s) = \overline{K^F_s(t)} = 0$ pues $K^F_s \in F$. Dicho de otro modo, $K^F_t = 0$ y luego $K_t = K^H_t$.

Sea $\mathcal{H}_1 := \{f|_{T_0} : f \in \mathcal{H}\}$ y definamos $L : \mathcal{H} \to \mathcal{H}_1$ por $Lf = f|_{T_0}$. Es claro que $\ker L = F$ y que $\text{Im}(L) = \mathcal{H}_1$. Si consideramos en $\mathcal{H}_1$ la norma cociente

$$||g||_{\mathcal{H}_1} = \inf_{Lf=g} ||f||_{\mathcal{H}},$$

entonces L se restringe a un isomorfismo isométrico $L : H \to \mathcal{H}_1$. En particular, tenemos que $(\mathcal{H}_1, ||\cdot||_{\mathcal{H}_1})$ es un espacio de Hilbert y el ínfimo que define a la norma cociente se realiza en H.

Por último, si $t \in T_0$, ya vimos que $K_t = K^H_t \in H$. Si $f \in H$, vale que

$$f|_{T_0}(t) = f(t) = \langle f, K_t\rangle_{\mathcal{H}} = \langle Lf, LK_t\rangle_{\mathcal{H}_1} = \langle f|_{T_0}, K_t|_{T_0}\rangle_{\mathcal{H}_1},$$

lo cual prueba que $\mathcal{H}_1$ es un RKHS cuyo núcleo reproductor es la restricción de K a T_0. Por la Proposición 1.2.2, vale que $\mathcal{H}_0 = \mathcal{H}_1$ y $||\cdot||_{\mathcal{H}_0} = ||\cdot||_{\mathcal{H}_1}$, como queríamos ver. $\square$

Observación 1.3.7. Manteniendo la notación de la Proposición 1.3.6, en su demostración se vió que para $t \in T_0$ vale que $K_t = K^H_t$.

Observación 1.3.8. Sean T un conjunto, $S \subseteq T$ y K un núcleo en T. De la Proposición 1.3.6 se desprende que para cada $f \in \mathcal{H}(K,S)$, existe $F \in \mathcal{H}(K,T)$ tal que $F|_S = f$ y $||F||_{\mathcal{H}(K,T)} = ||f||_{\mathcal{H}(K,S)}$.

Definición 1.3.9. Sean T un conjunto, K un núcleo en T y $T_0 \subseteq T$. Decimos que T_0 es un **conjunto determinante** para $\mathcal{H}(K)$ si $\{f \in \mathcal{H}(K) : f|_{T_0} = 0\} = \{0\}$.

En vistas de la Proposición 1.3.6, tenemos lo siguiente.

Proposición 1.3.10. *Sea T un conjunto, K un núcleo en T y $T_0 \subseteq T$. Llamemos $\mathcal{H}$ al RKHS en T definido por K y $\mathcal{H}_0$ al RKHS en T_0 definido por la restricción de K. Son equivalentes*

(1) T_0 *es un conjunto determinante para* $\mathcal{H}$*;*

(2) la aplicación $\mathcal{H} \to \mathcal{H}_0$ *dada por* $f \mapsto f|_{T_0}$ *es un isomorfismo isométrico;*

(3) toda $f \in \mathcal{H}_0$ *se extiende de forma única a una* $\hat{f} \in \mathcal{H}$ *y vale que* $||\hat{f}||_{\mathcal{H}} = ||f||_{\mathcal{H}_0}$*;*

(4) el subespacio $[K_t]_{t \in T_0}$ *es denso en* $\mathcal{H}$*;*

(5) la familia de funcionales acotados $\{\langle \cdot, K_t \rangle_{\mathcal{H}}\}_{t \in T_0}$ *separa puntos de* $\mathcal{H}$*.*

Demostración.

(1) $\implies$ *(2)* Esta implicación es consecuencia de la Proposición 1.3.6, y de que, como T_0 es determinante, para cada $g \in \mathcal{H}_0$ existe una única $f \in \mathcal{H}$ tal que $f|_{T_0} = g$.

(2) $\implies$ *(3)* Es inmediato, reinterpretando qué significa que la acción de restringir a T_0 sea un isomorfismo isométrico.

(3) $\implies$ *(4)* Basta ver que si $f \in \mathcal{H}$ es ortogonal a todas las funciones K_t con $t \in T_0$, entonces $f = 0$. Pero en este caso, se tiene $0 = \langle f, K_t \rangle = f(t)$ para todo $t \in T_0$. Dicho de otra forma, $f|_{T_0} = 0$. Por (3), $f|_{T_0}$ admite una única posible extensión a un elemento de $\mathcal{H}$ y como la función nula es una posible extensión, debe ser $f = 0$.

(4) $\implies$ *(5)* Si $f \in \mathcal{H}$ es ortogonal a todas las funciones K_t con $t \in T_0$, como éstas generan un subespacio denso en $\mathcal{H}$, tenemos necesariamente que $f = 0$, lo cual prueba (5).

(5) $\implies$ *(1)* Si $f|_{T_0} = 0$ entonces f y la función nula no se pueden separar por ningún funcional $\langle \cdot, K_t \rangle_{\mathcal{H}}$ con $t \in T_0$, por lo que deben ser iguales. □

1.3.4. Suma de núcleos

Proposición 1.3.11. *Sea* T *un conjunto,* K_1, K_2 *dos núcleos en* T *y* $K = K_1 + K_2$*. Llamemos* $\mathcal{H}_1, \mathcal{H}_2$ *y* $\mathcal{H}$ *a los RKHSs inducidos por* K_1, K_2 *y* K*, respectivamente. Entonces* $\mathcal{H} = \{f_1 + f_2 : f_i \in \mathcal{H}_i\}$ *y la norma en* $\mathcal{H}$ *está dada por*

$$||f||^2_{\mathcal{H}} = \inf_{f = f_1 + f_2} ||f_1||^2_{\mathcal{H}_1} + ||f_2||^2_{\mathcal{H}_2},$$

donde el ínfimo se toma sobre todas las descomposiciones $f = f_1 + f_2$ *con* $f_i \in \mathcal{H}_i$*.*

Demostración. Consideremos el espacio $\mathcal{K} = \mathcal{H}_1 \times \mathcal{H}_2$, que es un espacio de Hilbert con el producto interno dado por

$$\langle (f_1, f_2), (g_1, g_2) \rangle = \langle f_1, g_1 \rangle_1 + \langle f_2, g_2 \rangle_2,$$

y norma dada por

$$||(f_1, f_2)||^2 = ||f_1||^2_1 + ||f_2||^2_2.$$

Sea $\mathcal{K}_0 := \{(f, -f) : f \in \mathcal{H}_1 \cap \mathcal{H}_2\} \subseteq \mathcal{K}$. Veamos que $\mathcal{K}_0$ es un subespacio cerrado de $\mathcal{K}$. Si una sucesión $((f_n, -f_n))_{n\in\mathbb{N}}$ en $\mathcal{K}_0$ converge a $(f, g) \in \mathcal{K}$, entonces $f_n \xrightarrow{||\cdot||_1} f$ y $-f_n \xrightarrow{||\cdot||_2} g$. Como en cada $\mathcal{H}_i$ los funcionales de evaluación son acotados, esto implica que para cada $t \in T$ se tiene que $f_n(t) \to f(t)$ y $-f_n(t) \to g(t)$, de donde concluimos que $g = -f \in \mathcal{H}_1 \cap \mathcal{H}_2$ y en consecuencia $(f, g) = (f, -f) \in \mathcal{K}_0$.

Sea $L : \mathcal{K} \to \mathbb{K}^T$ la transformación lineal definida por $L(f_1, f_2) = f_1 + f_2$. Es inmediato de la definición que $\ker(L) = \mathcal{K}_0$. Si llamamos $S = \mathrm{Im}(L)$, entonces L induce un isomorfismo $\bar{L} : \mathcal{K}/\mathcal{K}_0 \to S$. Definiendo para $f \in S$ la norma $||f|| := ||\bar{L}^{-1}f||$, $\bar{L}$ resulta ser un isomorfismo isométrico y, al ser $\mathcal{K}/\mathcal{K}_0$ un espacio de Hilbert (de hecho isométricamente isomorfo a $\mathcal{K}_0^\perp$), tenemos entonces que S equipado con la norma $||\cdot||$ es un espacio de Hilbert. Afirmamos que $S = \mathcal{H}(K)$. En vistas de la Proposición 1.2.2, para verlo, nos basta con probar que el núcleo reproductor de S es K. Para ello, buscamos entender mejor el producto interno de S.

Por teoría general de espacios de Hilbert, sabemos que la proyección al cociente $\mathcal{K}_0^\perp \to \mathcal{K}/\mathcal{K}_0$, $h \mapsto \bar{h}$ es un isomorfismo isométrico. Entonces, para $f = f_1 + f_2$, $g = g_1 + g_2$ con $f_i, g_i \in \mathcal{H}_i$ y $(f_1, f_2), (g_1, g_2) \in \mathcal{K}_0^\perp$, tenemos que

$$\langle f, g\rangle = \langle \bar{L}^{-1}f, \bar{L}^{-1}g\rangle = \langle \overline{(f_1, f_2)}, \overline{(g_1, g_2)}\rangle = \langle (f_1, f_2), (g_1, g_2)\rangle = \langle f_1, g_1\rangle_1 + \langle f_2, g_2\rangle_2.$$

Para $t \in T$, sabemos que $K_t^i := K_i(\cdot, t) \in \mathcal{H}_i$. Además, para $f \in \mathcal{H}_1 \cap \mathcal{H}_2$, podemos decir que

$$\langle (f, -f), (K_t^1, K_t^2)\rangle = \langle f, K_t^1\rangle_1 + \langle -f, K_t^2\rangle_2 = f(x) - f(x) = 0,$$

de donde $(K_t^1, K_t^2) \in \mathcal{K}_0^\perp$. Luego, dada $g = g_1 + g_2$ con $(g_1, g_2) \in \mathcal{K}_0^\perp$, tenemos que

$$\langle g, K(\cdot, t)\rangle = \langle g_1 + g_2, K_t^1 + K_t^2\rangle = \langle g_1, K_t^1\rangle_1 + \langle g_2, K_t^2\rangle_2 = g_1(t) + g_2(t) = g(t).$$

Esto prueba que el núcleo reproductor de S es K y, consecuentemente, $\mathcal{H}(K) = S$.

Si recordamos la definición de la norma del cociente $\mathcal{K}/\mathcal{K}_0$ y tenemos en cuenta que $\bar{L}^{-1}f = \overline{(f_1, f_2)}$ si y sólo si $f = f_1 + f_2$, para $f \in \mathcal{H}(K)$ vale que

$$||f||^2 = \inf_{f=f_1+f_2} ||(f_1, f_2)||^2 = \inf_{f=f_1+f_2} ||f_1||_1^2 + ||f_2||_2^2,$$

donde el ínfimo se toma sobre todas las descomposiciones $f = f_1 + f_2$ con $f_i \in \mathcal{H}_i$. □

1.4. Dominancia de núcleos

En esta última sección desarrollamos los conceptos generales de RKHSs necesarios para abordar la generalización de la ley cero-uno de Driscoll en el Capítulo 2. Introducimos las nociones de dominancia y dominancia nuclear de núcleos de RKHSs, y demostramos una fórmula para la traza de un operador de dominancia en términos de trazas de ciertas matrices positivas. Éstos resultados son tomados de [11, Section 4]. A lo largo de esta sección, todos los espacios de Hilbert son reales.

Definición 1.4.1. Sea T un conjunto y sean K, R dos núcleos en T. Decimos que R **domina** a K y notamos $K \leq R$ si $\mathcal{H}(K) \subseteq \mathcal{H}(R)$.

Proposición 1.4.2. *Sea T un conjunto y sean K, R dos núcleos en T. Supongamos que $K \leq R$. Entonces el operador de inclusión $I : \mathcal{H}(K) \to \mathcal{H}(R)$ es acotado.*

Demostración. Sea $\mathcal{H} = \mathcal{H}(K + R)$. Por la Proposición 1.3.11 y teniendo en cuenta que $\mathcal{H}(K) \subseteq \mathcal{H}(R)$, vale la igualdad de conjuntos $\mathcal{H} = \mathcal{H}(K) + \mathcal{H}(R) = \mathcal{H}(R)$. De la misma proposición se desprende que $||f||_{\mathcal{H}} \leq ||f||_R$ para toda $f \in \mathcal{H}(R)$, por lo que la identidad $\mathcal{H}(R) \to \mathcal{H}$ es acotada. Como $\mathcal{H} = \mathcal{H}(R)$ es completo respecto de ambas normas, por el teorema de aplicación abierta tenemos que la identidad $\mathcal{H} \to \mathcal{H}(R)$ también es acotada.

Nuevamente, por la Proposición 1.3.11, sabemos que para cada $f \in \mathcal{H}(K)$ se tiene que $||f||_{\mathcal{H}} \leq ||f||_K$ y, en consecuencia, la inclusión $\mathcal{H}(K) \to \mathcal{H}$ es acotada. Componiendo con la identidad $\mathcal{H} \to \mathcal{H}(R)$, concluimos que la inclusión $\mathcal{H}(K) \to \mathcal{H}(R)$ es acotada como queríamos probar. □

Observación 1.4.3. Sean K, R son dos núcleos en un conjunto T tales que $K \leq R$, y llamemos $I : \mathcal{H}(K) \to \mathcal{H}(R)$ al operador de inclusión. Si $f \in \mathcal{H}(K)$, entonces $If \in \mathcal{H}(R)$ coincide con f como función de T en $\mathbb{R}$. Por este motivo, es usual omitir el operador de inclusión para no sobrecargar la notación, y muchas veces nos referiremos a elementos de $\mathcal{H}(K)$ como elementos de $\mathcal{H}(R)$. Este abuso de notación no es muy dañino, pues la definición de dominancia es justamente que, a nivel de conjuntos, los elementos de $\mathcal{H}(K)$ sean a su vez elementos de $\mathcal{H}(R)$.

Antes de dar la siguiente definición, recordemos la noción de operador nuclear.

Definición 1.4.4. Sean $\mathcal{H}$ un espacio de Hilbert y $L : \mathcal{H} \to \mathcal{H}$ un operador compacto. Luego el operador L^*L es compacto y positivo. Llamemos $(s_n^2)_{n\in\mathbb{N}}$ a la sucesión de elementos no nulos del espectro de L^*L contados con multiplicidad, con $s_n > 0$ y $s_{n+1} \leq s_n$ para todo $n \in \mathbb{N}$. (Observemos que $(s_n)_{n\in\mathbb{N}}$ no es otra cosa que la sucesión de valores singulares de L). Decimos que L es un operador **nuclear** si

$$\sum_{n\in\mathbb{N}} s_n < +\infty. \tag{1.4}$$

Se puede probar que si $L : \mathcal{H} \to \mathcal{H}$ es un operador positivo, entonces L es nuclear si y sólo si

$$\sum_{s\in S} \langle Le_s, e_s\rangle < +\infty \tag{1.5}$$

para alguna (o equivalentemente, toda) base ortonormal $\{e_s\}_{s\in S}$ de $\mathcal{H}$. Más aún, el valor de la serie (1.5) no depende de la base ortonormal elegida, y de hecho coincide con el valor de la serie (1.4). La **traza** de L se define como ese valor y la denotamos $\mathrm{Tr}(L)$. Una exposición más completa de esta clase de operadores puede hallarse en [12, Chapter 16].

Definición 1.4.5. Sea T un conjunto y sean K, R dos núcleos en T tales que $K \leq R$. Denotando $I : \mathcal{H}(K) \to \mathcal{H}(R)$ al operador de inclusión, llamamos **operador de dominancia** al operador $L = II^* : \mathcal{H}(R) \to \mathcal{H}(R)$. Decimos que la dominancia $K \leq R$ es **nuclear**, y notamos $K \ll R$, si el operador de dominancia es nuclear.

Observación 1.4.6. Si $K \leq R$, el operador de dominancia $L : \mathcal{H}(R) \to \mathcal{H}(R)$ queda caracterizado por la propiedad

$$LR_t = K_t \quad \forall t \in T. \tag{1.6}$$

En efecto, si $I : \mathcal{H}(K) \to \mathcal{H}(R)$ denota a la inclusión, por definición vale que $L = II^*$. En particular, $\mathrm{Im}(L) \subseteq \mathcal{H}(K)$. Entonces, para $s, t \in T$ vale:

$$LR_t(s) = \langle II^* R_t, K_s \rangle_K = \langle I^* R_t, K_s \rangle_K = \langle R_t, IK_s \rangle_R = \langle R_t, K_s \rangle_R = K_s(t) = K_t(s).$$

Además, como $[R_t]_{t \in T}$ es denso en $\mathcal{H}(R)$, L es el único operador acotado que verifica (1.6).

Observación 1.4.7. Sean K, R dos núcleos en un conjunto T tales que $K \leq R$. Luego, dados $g \in \mathcal{H}(K)$ y $t \in T$,

$$\langle LR_t, g \rangle_K = \langle K_t, g \rangle_K = g(t) = \langle R_t, g \rangle_R.$$

Por lo tanto, de la densidad de $[R_t]_{t \in T}$ en $\mathcal{H}(R)$ se desprende que para toda $f \in \mathcal{H}(R)$,

$$\langle Lf, g \rangle_K = \langle f, g \rangle_R. \tag{1.7}$$

Ejemplo 1.4.8. Sea $T = \mathbb{N}$ y consideremos en T los núcleos $R(m, n) = \delta_{m,n}$ y $K(m, n) = n^{-p}\delta_{m,n}$ con $p \geq 0$. Por el Ejemplo 1.2.12, tenemos que $\mathcal{H}(R) = \ell^2$ con el producto interno usual, y que

$$\mathcal{H}(K) = \left\{ a \in \mathbb{R}^{\mathbb{N}} : \sum_{n=1}^{\infty} n^p |a_n|^2 < \infty \right\}, \quad \langle a, b \rangle_K = \sum_{n=1}^{\infty} n^p a_n b_n.$$

Es claro que $K \leq R$. Llamemos $L : \mathcal{H}(R) \to \mathcal{H}(R)$ al operador de dominancia. Teniendo en cuenta la Observación 1.4.7, para $a \in \mathcal{H}(R)$,

$$(La)_n = \langle La, K_n \rangle_K = \langle a, K_n \rangle_R = \langle a, n^{-p} R_n \rangle_R = n^{-p} a_n.$$

Es decir, $La = (n^{-p} a_n)_{n \in \mathbb{N}}$. Notar que en particular, como R_n no es más que el $n-$ésimo vector canónico e_n, $LR_n = n^{-p} R_n$. Por último, al ser $\{R_n\}_{n \in \mathbb{N}} = \{e_n\}_{n \in \mathbb{N}}$ una base ortonormal de $\mathcal{H}(R) = \ell^2$,

$$\mathrm{Tr}(L) = \sum_{n=1}^{\infty} \langle LR_n, R_n \rangle_R = \sum_{n=1}^{\infty} \langle n^{-p} e_n, e_n \rangle_{\ell^2} = \sum_{n=1}^{\infty} \frac{1}{n^p},$$

de donde $K \ll R$ si y sólo si $p > 1$.

Con esencialmente el mismo argumento, definiendo $K(m, n) = c_n \delta_{m,n}$ con $c = (c_n)_{n \in \mathbb{N}}$ una sucesión de números reales positivos, se puede ver que $K \leq R$ si y sólo si $c \in \ell^\infty$, y que $K \ll R$ si y sólo si $c \in \ell^1$.

De las Observaciones 1.4.6 y 1.4.7, se ve que hay ciertas similitudes entre los operadores de dominancia y los operadores de proyección ortogonal. Por ejemplo, ambos operadores se definen esencialmente como el operador adjunto de una inclusión. Sin embargo, la diferencia crucial entre ellos es que en el caso de la proyección ortogonal, se considera en el subespacio la norma inducida, mientras que esto no es así en el caso del operador de dominancia.

Aunque la analogía no sea perfecta, permite generar algo de intuición en relación a la dominancia nuclear. Concretamente, para proyectores ortogonales tenemos que un operador de proyección ortogonal es nuclear si y sólo si su imagen tiene dimensión finita. Si bien esto no es necesariamente cierto para los operadores de dominancia (como se vió en el Ejemplo 1.4.8), se podría extrapolar por analogía que un núcleo R domina nuclearmente a un núcleo K cuando $\mathcal{H}(K)$ "es pequeño en relación a $\mathcal{H}(R)$", lo cual justifica la notación $K \ll R$.

Hasta ahora tenemos que toda dominancia $K \leq R$ produce un operador positivo $L : \mathcal{H}(R) \to \mathcal{H}(R)$. La siguiente proposición dice que todos los operadores positivos en un RKHS son de esta forma.

Proposición 1.4.9. *Sean T un conjunto, R un núcleo en T y $L : \mathcal{H}(R) \to \mathcal{H}(R)$ un operador positivo. Entonces $K(s,t) = \langle LR_s, R_t\rangle$ define un núcleo en T tal que $K \leq R$ y L resulta ser el operador de dominancia.*

Demostración. Dados $t_1, \ldots, t_n \in T$ y $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, tenemos que

$$\sum_{i,j} \alpha_i \alpha_j K(t_i, t_j) = \sum_{i,j} \alpha_i \alpha_j \langle LR_{t_i}, R_{t_j}\rangle = \left\langle L\left(\sum_i \alpha_i R_{t_i}\right), \sum_j \alpha_i R_{t_j}\right\rangle \geq 0$$

por ser $L \geq 0$. Además, al ser L autoadjunto vale que $K(s,t) = \langle LR_s, R_t\rangle = \langle R_s, LR_t\rangle = \langle LR_t, R_s\rangle = K(t,s)$, lo cual termina de probar que K es un núcleo.

Para ver que $K \leq R$, primero notemos que si $m > 0$, entonces de la Proposición 1.3.2 se desprende que los conjuntos $\mathcal{H}(mR)$ y $\mathcal{H}(R)$ son iguales (a pesar de que los productos internos de cada espacio no lo son). Ahora bien, si $s, t \in T$,

$$(mR)(s,t) - K(s,t) = m\langle R_s, R_t\rangle - \langle LR_s, R_t\rangle = \langle (mI_{\mathcal{H}(R)} - L)R_s, R_t\rangle.$$

Para $m > ||L||$, el operador $mI_{\mathcal{H}(R)} - L$ es positivo. En efecto, si $h \in \mathcal{H}(R)$,

$$\langle (mI_{\mathcal{H}(R)} - L)h, h\rangle = m||h||^2 - \langle Lh, h\rangle \geq m||h||^2 - ||L||\, ||h||^2 \geq 0.$$

Consecuentemente, $mR - K$ es un núcleo en T. Por la Proposición 1.3.11, se tiene la igualdad de conjuntos

$$\mathcal{H}(R) = \mathcal{H}(mR) = \mathcal{H}(mR - K) + \mathcal{H}(K)$$

y, en particular, $\mathcal{H}(K) \subseteq \mathcal{H}(R)$, lo cual prueba que $K \leq R$.

Por último, si $s, t \in T$ entonces

$$LR_s(t) = \langle LR_s, R_t \rangle = K(s,t) = K_s(t).$$

Es decir, $LR_s = K_s$. Por la Observación 1.4.6, concluimos que L es el operador de dominancia. $\square$

En el Capítulo 2 será de gran importancia poder reducirnos a situaciones en las que los RKHSs involucrados sean separables. Por ejemplo, la fórmula de la traza (Proposición 1.4.20) es válida en este contexto. En lo que sigue, buscaremos entender mejor cuándo el RKHS asociado a un núcleo dado es separable, qué propiedades adicionales cumple y con qué argumentos se pueden efectuar dichas reducciones.

El siguiente lema muestra cómo "intercalar" un RKHS entre dos RKHSs en dominancia nuclear.

Lema 1.4.10. *Sean T un conjunto y K, R dos núcleos en T tales que $K \ll R$. Entonces existe un núcleo R^0 en T tal que $K \ll R^0 \leq R$ y $\mathcal{H}(R^0)$ es separable.*

Demostración. Llamemos L al operador de dominancia de R sobre K. Como L es nuclear, es en particular compacto y luego $\mathrm{Im}(L)$ es un subespacio separable de $\mathcal{H}(R)$. Sea $\mathcal{H}_0 := \overline{\mathrm{Im}(L)}$, donde la clausura se toma respecto de la norma $||\cdot||_R$. Tenemos que $\mathcal{H}_0$ es un subespacio cerrado de un RKHS y es por ende un RKHS en sí mismo, digamos con núcleo R^0. Además es separable por ser la clausura de un subespacio separable. Veamos que $R^0 \geq K$.

Si $f \in \mathcal{H}(R) \ominus \mathcal{H}_0$, entonces para cada $t \in T$ vale que $\langle f, LR_t \rangle_R = 0$. Por la Observación 1.4.6 sabemos que $LR_t = K_t$, de donde concluimos que $\langle f, g \rangle_R = 0$ para toda $g \in [K_t]_{t \in T}$. Por último, si $g \in \mathcal{H}(K)$, existe una sucesión $(g_n)_{n \in \mathbb{N}} \subseteq [K_t]_{t \in T}$ tal que $g_n \xrightarrow{||\cdot||_K} g$. Como el operador de inclusión $\mathcal{H}(K) \to \mathcal{H}(R)$ es acotado, esto implica que $g_n \xrightarrow{||\cdot||_R} g$ y como $\langle f, g_n \rangle_R = 0$ para todo $n \in \mathbb{N}$, concluimos que $\langle f, g \rangle_R = 0$. Esto prueba que

$$\mathcal{H}(R) \ominus \mathcal{H}_0 \subseteq \mathcal{H}(R) \ominus \mathcal{H}(K),$$

y consecuentemente $\overline{\mathcal{H}(K)} \subseteq \mathcal{H}_0$ (notar que $\mathcal{H}(K)$ no tiene por qué cer cerrado en $\mathcal{H}(R)$ respecto de la norma $||\cdot||_R$). En particular, $\mathcal{H}(K) \subseteq \mathcal{H}_0$ como queríamos ver.

Por último, si $P : \mathcal{H}(R) \to \mathcal{H}_0$ denota a la proyección ortogonal, entonces sabemos por la Proposición 1.3.3 que $PR_t = R_t^0$. Si $f \in \mathcal{H}(K)$, tenemos que

$$\langle f, LPR_t \rangle_K = \langle f, PR_t \rangle_R = \langle f, R_t^0 \rangle_R = f(t).$$

Es decir, $LR_t^0 = K_t$. Por lo tanto, la restricción de L a $\mathcal{H}_1$ es el operador de dominancia de R^0 sobre K y, como L es nuclear y positivo, su restricción también lo es, lo cual termina de probar que $R^0 \gg K$. $\square$

Definición 1.4.11. Sea T un conjunto y K un núcleo en T. La **pseudo-métrica asociada a** K es la pseudo-métrica d_K definida por $d_K(s,t) := ||K_s - K_t||_K$.

Observación 1.4.12. La pseudo-métrica d_K resulta ser una métrica si y sólo si $\mathcal{H}(K)$ separa puntos de T. Es decir, si dados $s, t \in T$ distintos existe $f \in \mathcal{H}(K)$ tal que $f(s) \neq f(t)$. En particular, esto ocurre si $\{K_t\}_{t\in T}$ es linealmente independiente.

Lema 1.4.13. *Sea T un conjunto y K un núcleo en T tal que d_K es una métrica en T. Entonces,*

(1) todas las funciones $f \in \mathcal{H}(K)$ son $||f||_K-$Lipschitz para d_K;

(2) el espacio métrico (T, d_K) es separable si y sólo si $\mathcal{H}(K)$ es separable.

Demostración. (1) Dada $f \in \mathcal{H}(K)$, tenemos que $|f(s) - f(t)| = |\langle f, K_s - K_t\rangle| \leq ||f||_K d_K(s,t)$, lo cual prueba *(1)*.

(2) Supongamos que (T, d_K) es un espacio métrico separable y sea $S \subseteq T$ un subconjunto denso contable. Sea A el $\mathbb{Q}-$subespacio generado por $\{K_t\}_{t\in S}$. Luego A es denso en $[K_t]_{t\in T}$, que a su vez es denso en $\mathcal{H}(K)$, con lo que A resulta ser un denso contable en $\mathcal{H}(K)$.

Recíprocamente, supongamos que $\mathcal{H}(K)$ es separable. La aplicación $T \to \mathcal{H}(K)$ dada por $t \mapsto K_t$ es una isometría por definición de d_K. Luego T se identifica isométricamente con un subconjunto de $\mathcal{H}(K)$. Como todo subconjunto de un espacio métrico separable es separable, esto prueba que T es separable. □

Recordemos que si X es un espacio vectorial, una base de Hamel de X es un sistema de generadores de X linealmente independiente (es decir, una base algebraica).

Definición 1.4.14. Si T es un conjunto y K es un núcleo en T, un subconjunto $T_0 \subseteq T$ se dice un subconjunto $K-$**Hamel** (o simplemente Hamel si se sobreentiende el núcleo) de T si $\{K_t\}_{t\in T_0}$ es una base de Hamel de $[K_t]_{t\in T}$.

El siguiente lema es consecuencia inmediata de la Observación 1.4.12 y de la definición de subconjunto Hamel.

Lema 1.4.15. *Si K es un núcleo en T y T_0 es un subconjunto Hamel de T, entonces (T_0, d_K) es un espacio métrico y $[K_t]_{t\in T_0}$ es denso en $\mathcal{H}(K)$.*

La siguiente proposición es uno de los motivos por los cuales es conveniente trabajar con RKHSs separables. Si $\mathcal{H}(K)$ es un RKHS separable definido sobre un conjunto T, este resultado nos permitirá, esencialmente, reducir ciertos argumentos al caso donde el conjunto T es contable, vía la Proposición 1.3.10.

Proposición 1.4.16. *Sean T un conjunto, K un núcleo en T. Supongamos que $\mathcal{H}(K,T)$ es separable y sea $T_0 \subseteq T$ un subconjunto Hamel. Entonces (T_0, d_K) es un espacio métrico separable. Si $S_0 \subseteq T_0$ es un denso contable, entonces S_0 es un conjunto determinante para $\mathcal{H}(K,T)$.*

Demostración. De la Observación 1.4.12 se ve que d_K es una métrica en T_0. Por la Proposición 1.3.6, $\mathcal{H}(K, T_0)$ es isomorfo a un cociente de $\mathcal{H}(K, T)$, que es separable por hipótesis. Luego $\mathcal{H}(K, T_0)$ es separable y por el Lema 1.4.13, (T_0, d_K) es separable.

Sea $f \in \mathcal{H}(K, T)$ tal que $f|_{S_0} = 0$. Como $f|_{T_0}$ es d_K-continua y $S_0 \subseteq T_0$ es d_K-denso, entonces $f|_{T_0} = 0$. Como T_0 es determinante para $\mathcal{H}(K, T)$, tenemos que $f = 0$. □

Dados $S \subseteq T$, $f : T \to \mathbb{R}$ y K un núcleo en T, notaremos

$$||f||_S := ||f|_S||_S = ||f|_S||_{\mathcal{H}(K,S)}, \tag{1.8}$$

donde la notación presupone que $f|_S \in \mathcal{H}(K, S)$.

Lema 1.4.17. *Sean T un conjunto, K un núcleo en T y $T_0 \subseteq T$ $K-$Hamel. Sea $S_0 = \{s_k\}_{k\in\mathbb{N}} \subseteq T_0$ un subconjunto d_K-denso y contable en T_0 y sea, para cada $n \in \mathbb{N}$, $S_n := \{s_k\}_{1\leq k\leq n}$. Entonces para cada $f \in \mathcal{H}(K, T)$ la sucesión $(||f||_{S_n})_{n\in\mathbb{N}}$ es creciente y*

$$\lim_{n\to\infty} ||f||_{S_n} = ||f||_T. \tag{1.9}$$

Recíprocamente, si $f : T \to \mathbb{R}$ es tal que

$$\lim_{n\to\infty} ||f||_{S_n} = \sup_{n\in\mathbb{N}} ||f||_{S_n} < +\infty, \tag{1.10}$$

entonces $f|_{S_0} \in \mathcal{H}(K, S_0)$ y existe una única $f_1 \in \mathcal{H}(K, T)$ que coincide con f en S_0. Además,

$$||f_1||_T = ||f||_{S_0} = \lim_{n\to\infty} ||f||_{S_n}. \tag{1.11}$$

Demostración. Si $f \in \mathcal{H}(K, T)$, entonces $||f||_{S_n} \leq ||f||_{S_{n+1}}$ por la Proposición 1.3.6. En vistas de la Proposición 1.4.16 sabemos que S_0 es un conjunto determinante para $\mathcal{H}(K, T)$, y luego, por la Proposición 1.3.10, vale que $||f||_T = ||f||_{S_0}$. Como todo subconjunto finito de S_0 está contenido en algún S_n, de la Proposición 1.3.2 se desprende que $||f||_{S_0} = \sup_{n\in\mathbb{N}} ||f||_{S_n} = \lim_{n\to\infty} ||f||_{S_n}$.

Si ahora $f : T \to \mathbb{R}$ verifica (1.10), nuevamente por la Proposición 1.3.2 vale que $f|_{S_0} \in \mathcal{H}(K, S_0)$. Como S_0 es determinante, la existencia y unicidad de f_1 y (1.11) se desprenden de la Proposición 1.3.10. □

Lema 1.4.18. *Sea T un conjunto, $S \subseteq T$, sean K, R dos núcleos en T tales que $K \leq R$ y llamemos $L : \mathcal{H}(R, T) \to \mathcal{H}(R, T)$ al operador de dominancia. Definamos*

$$F_1 = \mathcal{H}(R, T) \ominus \{f \in \mathcal{H}(R, T) : f|_S = 0\} \quad y$$

$$H_1 = \mathcal{H}(K, T) \ominus \{f \in \mathcal{H}(K, T) : f|_S = 0\}.$$

Entonces,

(1) $\mathcal{H}(K, S) \subseteq \mathcal{H}(R, S)$;

(2) $L(F_1) \subseteq H_1$.

Demostración.

(1) Si $\hat{f} \in \mathcal{H}(K,S)$, por la Proposición 1.3.6 existe $f \in \mathcal{H}(K,T)$ tal que $\hat{f} = f|_S$. Como $K \leq R$ tenemos que $f \in \mathcal{H}(R,T)$ y consecuentemente $\hat{f} = f|_S \in \mathcal{H}(R,S)$.

(2) Llamemos

$$F_0 = \{f \in \mathcal{H}(R,T) : f|_S = 0\}, \quad H_0 = \{f \in \mathcal{H}(K,T) : f|_S = 0\}.$$

Como $\mathrm{Im}(L) \subseteq \mathcal{H}(K,T)$, para ver que $L(F_1) \subseteq H_1$ basta ver que, dada $f \in F_1$, $Lf \perp H_0$. Sea entonces $g \in H_0$. En particular, $g \in F_0$ y luego, por la Observación 1.4.7,

$$\langle Lf, g\rangle_K = \langle f, g\rangle_R = 0,$$

lo cual termina la prueba. □

Sean T un conjunto, K, R dos núcleos en T tales que $K \leq R$ y $S \subseteq T$. Si denotamos K_S y R_S a las restricciones a S de K y R, entonces por la parte *(1)* del Lema 1.4.18 tenemos que $K_S \leq R_S$. El siguiente resultado da una caracterización del operador de dominancia de $K_S \leq R_S$ en términos del operador de dominancia de $K \leq R$. La caracterización es simple: salvo isometrías, el primero es una restricción y correstricción del último.

Proposición 1.4.19. *Sean T un conjunto, K, R dos núcleos en T con $K \leq R$ y $S \subseteq T$. Llamemos $L : \mathcal{H}(R,T) \to \mathcal{H}(R,T)$ y $L_S : \mathcal{H}(R,S) \to \mathcal{H}(R,S)$ a los respectivos operadores de dominancia, y recordemos de la Definición 1.4.5 que $\mathrm{Im}(L_S) \subseteq \mathcal{H}(K,S)$. Definamos F_1 y H_1 como en el Lema 1.4.18. Si $J : \mathbb{R}^T \to \mathbb{R}^S$ es el operador de restricción a S, entonces el diagrama*

$$\begin{tikzcd} F_1 \arrow[r, "L"] \arrow[d, "J"] & H_1 \arrow[d, "J"] \\ \mathcal{H}(R,S) \arrow[r, "L_S"] & \mathcal{H}(K,S) \end{tikzcd}$$

es conmutativo, y $\mathrm{Tr}(L_S) = \mathrm{Tr}(L|_{F_1})$. En particular, vale que si L es nuclear, entonces L_S también y $\mathrm{Tr}(L_S) \leq \mathrm{Tr}(L)$, con igualdad en el caso que S sea un conjunto determinante para $\mathcal{H}(R,T)$.

Demostración. Fijemos $t \in S$ y llamemos R^{F_1} al núcleo reproductor de F_1. Por la Observación 1.3.7, vale que $R_t^{F_1} = R_t$. Como las funciones JR_t y JK_t son los núcleos reproductores en t de $\mathcal{H}(R,S)$ y $\mathcal{H}(K,S)$ respectivamente (por la propia definición de estos espacios), tenemos por la Observación 1.4.6 que $L_S JR_t = JK_t$. Notemos $J^{-1} : \mathcal{H}(R,S) \to F_1$ a la inversa de la restricción de J a F_1 correstringida a $\mathcal{H}(R,S)$, que es un isomorfismo isométrico por la Proposición 1.3.6. Juntando todo, llegamos a que

$$L_S JR_t^{F_1} = L_S JR_t = JK_t = JLR_t = JLR_t^{F_1} = JLJ^{-1}JR_t^{F_1}.$$

Con esto, los operadores L_S y $JLJ^{-1} : \mathcal{H}(R,S) \to \mathcal{H}(K,S)$ coinciden en $[JR_t]_{t\in S}$, y luego son iguales por densidad. Es decir, $L_S = JLJ^{-1}$, o lo que es lo mismo, $L_S J = JL$ como queríamos ver.

Si $\{\hat{e}_i\}_{i\in I}$ es una base ortonormal de $\mathcal{H}(R,S)$, llamando $e_i = J^{-1}\hat{e}_i$ tenemos que $\{e_i\}_{i\in I}$ es una base ortonormal de F_1. Consecuentemente,

$$\begin{aligned}
\mathrm{Tr}(L_S) &= \sum_{i\in I} \langle L_S \hat{e}_i, \hat{e}_i\rangle_{\mathcal{H}(R,S)} = \sum_{i\in I} \langle JLe_i, \hat{e}_i\rangle_{\mathcal{H}(R,S)} \\
&= \sum_{i\in I} \langle J^{-1}JLe_i, J^{-1}\hat{e}_i\rangle_{F_1} = \sum_{i\in I} \langle Le_i, e_i\rangle_{\mathcal{H}(R,T)} \\
&= \mathrm{Tr}(L|_{F_1}).
\end{aligned}$$

Extendiendo $\{e_i\}_{i\in I}$ a una base ortonormal $\{e_i\}_{i\in I'}$ de $\mathcal{H}(R,T)$ con $I \subseteq I'$, y teniendo en cuenta que los operadores en cuestión son positivos, tenemos que

$$\mathrm{Tr}(L_S) = \mathrm{Tr}(L|_{F_1}) = \sum_{i\in I} \langle Le_i, e_i\rangle_{\mathcal{H}(R,T)} \leq \sum_{i\in I'} \langle Le_i, e_i\rangle_{\mathcal{H}(R,T)} = \mathrm{Tr}(L).$$

En el caso que S sea un conjunto determinante para $\mathcal{H}(R,T)$, vale que $F_1 = \mathcal{H}(R,T)$ y por lo tanto $\mathrm{Tr}(L_S) = \mathrm{Tr}(L)$. □

Proposición 1.4.20. *Sean T un conjunto y R un núcleo en T tal que $\mathcal{H}(R,T)$ es separable. Sea K otro núcleo en T tal que $K \ll R$ con operador de dominancia L. Sea $T_0 \subseteq T$ un subconjunto $R-$Hamel y sea $S_0 = \{s_k\}_{n\in\mathbb{N}} \subseteq T_0$ un subconjunto d_R-denso y contable. Denotemos K_n y R_n a las matrices obtenidas de restringir los núcleos K y R al conjunto $S_n = \{s_k\}_{1\leq k\leq n}$. Entonces*

$$Tr(L) = \lim_{n\to\infty} Tr(K_n R_n^{-1}). \tag{1.12}$$

Demostración. Como T_0 es un conjunto $R-$Hamel, $\{R_t\}_{t\in T_0}$ es linealmente independiente, y dado que $S_0 \subseteq T_0$, las matrices R_n son inversibles. Por la Proposición 1.4.16, S_0 es un conjunto determinante para $\mathcal{H}(R,T)$, y entonces, de la Proposición 1.4.19 se desprende que $\mathrm{Tr}(L) = \mathrm{Tr}(L_{S_0})$, donde L_{S_0} es el operador de dominancia de $\mathcal{H}(R,S_0)$ sobre $\mathcal{H}(K,S_0)$. Esto nos permite reducirnos al caso en que T es contable y $R-$Hamel. Luego, sin pérdida de generalidad, supongamos que $T = S_0 = \{s_k\}_{k\in\mathbb{N}}$.

Sea $\{e_i\}_{i\in\mathbb{N}}$ la base ortonormal de $\mathcal{H}(R,T)$ que se obtiene aplicando el algoritmo de Gram-Schmidt a $\{R_{s_k}\}_{k\in\mathbb{N}}$. Por definición de traza tenemos que $\mathrm{Tr}(L) = \sum_i \langle Le_i, e_i\rangle_{\mathcal{H}(R,T)}$. Para cada $n\in\mathbb{N}$ sea L_n el operador de dominancia de $\mathcal{H}(R,S_n)$ sobre $\mathcal{H}(K,S_n)$ y $J_n : \mathcal{H}(R,T) \to \mathcal{H}(R,S_n)$ la restricción a S_n. Notemos que si

$$F_0^n := \{f \in \mathcal{H}(R,T) : f|_{S_n} = 0\},$$

entonces $[R_{s_k}]_{1\leq k\leq n} = \mathcal{H}(R,T) \ominus F_0^n$, por lo que en vistas de la Proposición 1.3.6 vale que $J_n : [R_{s_k}]_{1\leq k\leq n} \to \mathcal{H}(R,S_n)$ es un isomorfismo isométrico. Por construcción de la base ortonormal $\{e_i\}_{i\in\mathbb{N}}$ vale que $[R_{s_k}]_{1\leq k\leq n} = [e_k]_{1\leq k\leq n}$ y, luego, $\{J_n e_i\}_{1\leq i\leq n}$ es una base ortonormal de $\mathcal{H}(R, S_n)$. Consecuentemente, por la Proposición 1.4.19, podemos decir que

$$\operatorname{Tr}(L_n) = \sum_{i=1}^{n} \langle L_n J_n e_i, J_n e_i\rangle_{\mathcal{H}(R,S_n)} = \sum_{i=1}^{n} \langle J_n L e_i, J_n e_i\rangle_{\mathcal{H}(R,S_n)} = \sum_{i=1}^{n} \langle L e_i, e_i\rangle_{\mathcal{H}(R,T)},$$

de donde obtenemos que $\operatorname{Tr}(L) = \operatorname{lím}_{n\to\infty} \operatorname{Tr}(L_n)$.

Por último, considerando L_n como una transformación lineal entre espacios de dimensión finita, de la Observación 1.4.6 obtenemos que vale la ecuación matricial $L_n R_n = K_n$. Por lo tanto $L_n = K_n R_n^{-1}$ y juntando esto con lo anterior, concluimos que $\operatorname{Tr}(L) = \operatorname{lím}_{n\to\infty} \operatorname{Tr}(L_n) = \operatorname{lím}_{n\to\infty} \operatorname{Tr}(K_n R_n^{-1})$, como queríamos probar. $\square$

Capítulo 2

Trayectorias Aleatorias en RKHSs

En este capítulo exploraremos las conexiones entre la teoría de probabilidades y los RKHSs. Concretamente, estudiaremos cómo asociarle RKHSs a procesos estocásticos de segundo orden, y qué propiedades del proceso se pueden deducir a partir de las propiedades del espacio asociado. En la Sección 2.1 damos la definición de proceso estocástico junto con algunas nociones básicas del área. La Sección 2.2 introduce nociones de probabilidad en el contexto de espacios de Hilbert, que serán ampliadas en la Sección 2.3 cuando estudiemos elementos aleatorios en espacios de Hilbert. Por último, en la Sección 2.4 combinaremos estas ideas con la teoría de RKHS para demostrar uno de los teoremas más interesantes del área: la ley cero-uno de Driscoll (o mejor dicho, una generalización de ella).

A lo largo de este capítulo, si $(\Omega, \mathcal{A}, \mathbb{P})$ es un espacio de probabilidad, se asume que $\mathbb{P}$ es completa. Es decir, si $A \in \mathcal{A}$ cumple que $\mathbb{P}(A) = 0$ y $B \subseteq A$, entonces $B \in \mathcal{A}$.

Si $(\Omega, \mathcal{A}, \mathbb{P})$ es un espacio de medida, $(\Omega', \mathcal{A}')$ es un espacio medible y $f : \Omega \to \Omega'$ es una función medible, denotamos $\mathbb{P}_f$ a la *medida pushforward* definida por

$$\mathbb{P}_f(A') := \mathbb{P}(f^{-1}(A')), \quad \forall A' \in \mathcal{A}'.$$

Si T es un espacio topológico, denotamos $\hat{B}(T)$ a la $\sigma-$álgebra de Borel de T. Es decir, a la $\sigma-$álgebra generada por los conjuntos abiertos de T.

Todos los espacios de Hilbert considerados en este capítulo son reales.

2.1. Procesos estocásticos

Definición 2.1.1. Sean Ω un espacio de probabilidad y T un conjunto. Un **proceso estocástico** (real) es una función $X : \Omega \times T \to \mathbb{R}$ tal que para cada $t \in T$ la función $X_t := X(\cdot, t)$ es una variable aleatoria. Equivalentemente, es una familia de variables aleatorias $\{X_t : \Omega \to \mathbb{R}\}_{t \in T}$. Llamamos **trayectorias** de X a las funciones $X(\omega, \cdot) : T \to \mathbb{R}$, y en ocasiones las denotaremos $X_\bullet(\omega)$.

En muchas aplicaciones, el conjunto T es interpretado como un dominio temporal. De esta forma, un proceso estocástico modela una familia de eventos aleatorios que van transcurriendo en el tiempo (como podría ser arrojar una moneda varias veces, o el ingreso de clientes a un local a lo largo del día). Cuando los eventos suceden a tiempo discreto, es usual tomar $T = \mathbb{N}$ y tenemos una sucesión de variables aleatorias. Si los eventos suceden a tiempo continuo, es usual tomar como T algún intervalo de la recta real. Sin embargo, nada impide considerar procesos estocásticos definidos en conjuntos más generales.

A continuación, damos algunas de las definiciones básicas asociadas a procesos estocásticos con las que trabajaremos durante el resto del capítulo.

Definición 2.1.2. Sea $X : \Omega \times T \to \mathbb{R}$ un proceso estocástico. Una **versión** de X es un proceso estocástico $\tilde{X} : \Omega \times T \to \mathbb{R}$ tal que para cada $t \in T$ se tiene que $X_t \overset{a.s.}{=} \tilde{X}_t$.

Definición 2.1.3. Sea $X : \Omega \times T \to \mathbb{R}$ un proceso estocástico. Si $p \geq 1$, decimos que X es un **proceso de orden** p si para todo $t \in T$ se tiene que $X_t \in L^p(\Omega)$. Para $p = 1$ y $p = 2$, decimos que X tiene **primer orden** y **segundo orden** respectivamente.

Si $X : \Omega \times T \to \mathbb{R}$ es un proceso de primer orden, está bien definida su **función de media** $m : T \to \mathbb{R}$ dada por

$$m(t) = \mathbb{E}[X_t].$$

Más aún, si X es además un proceso de segundo orden, está bien definida su **función de covarianza** $K : T \times T \to \mathbb{R}$ dada por

$$K(s,t) = Cov(X_s, X_t) = \mathbb{E}[(X_s - m(s))(X_t - m(t))] = \mathbb{E}[X_s X_t] - \mathbb{E}[X_s]\mathbb{E}[X_t].$$

Notemos que en este último caso, K es un núcleo en T. En efecto, es claro de la definición que K es simétrica. Además, para $s, t \in T$ vale que

$$K(s,t) = \mathbb{E}[(X_s - m(s))(X_t - m(t))] = \langle X_s - m(s), X_t - m(t) \rangle_{L^2(\Omega)}.$$

Por lo tanto, para $t_1, \ldots, t_n \in T$ y $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$,

$$\begin{aligned}
\sum_{i,j} \alpha_i \alpha_j K(t_i, t_j) &= \sum_{i,j} \alpha_i \alpha_j \langle X_{t_i} - m(t_i), X_{t_j} - m(t_j) \rangle_{L^2(\Omega)} \\
&= \left\langle \sum_i \alpha_i (X_{t_i} - m(t_i)), \sum_j \alpha_j (X_{t_j} - m(t_j)) \right\rangle_{L^2(\Omega)} \\
&= \left\| \sum_i \alpha_i (X_{t_i} - m(t_i)) \right\|_{L^2(\Omega)}^2 \geq 0.
\end{aligned}$$

Esta será la conexión entre procesos estocásticos y RKHSs que estudiaremos en la Sección 2.4. En adelante, todos los procesos estocásticos se asumirán de segundo orden, a menos que se indique explícitamente lo contrario.

En general, si m denota a la función de media de un proceso de segundo orden X, el proceso $Y_t := X_t - m(t)$ es un proceso de media 0 cuya función de covarianza coincide con la de X. Esto permitirá reducir muchos de los argumentos al caso en que la media del proceso es 0.

Por último, introducimos la clase central de procesos estocásticos que estudiaremos: los procesos gaussianos.

Definición 2.1.4. Un proceso estocástico $X : \Omega \times T \to \mathbb{R}$ se dice un **proceso gaussiano** si para cualesquiera $t_1, \dots, t_n \in T$ distintos y $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, se tiene que la variable aleatoria $\alpha_1 X_{t_1} + \cdots + \alpha_n X_{t_n}$ es gaussiana. Dicho de otro modo, si el vector aleatorio $(X_{t_1}, \dots, X_{t_n})$ es conjuntamente gaussiano para cualquier elección de puntos distintos $t_1, \dots, t_n \in T$.

2.2. Medidas de probabilidad en espacios de Hilbert

Comencemos recordando la siguiente noción en teoría de la medida. Sean X un conjunto y $\{\Omega_i\}_{i\in I}$ una familia de espacios medibles. Supongamos, además, que para cada $i \in I$ se tiene una función $f_i : X \to \Omega_i$. La **$\sigma-$álgebra inicial** en X respecto de la familia $\{f_i\}_{i\in I}$ es la $\sigma-$álgebra en X generada por los conjuntos de la forma $f_i^{-1}(A)$ con $i \in I$ y $A \subseteq \Omega_i$ medible. Equivalentemente, es la menor $\sigma-$álgebra en X que hace medibles a todas las funciones f_i.

Proposición 2.2.1. *Sean X un conjunto, $\mathcal{A}$ una $\sigma-$álgebra en X y $\{f_i : X \to \Omega_i\}_{i\in I}$ una familia de funciones, donde cada Ω_i es un espacio medible. Son equivalentes:*

(1) $\mathcal{A}$ es la $\sigma-$álgebra inicial respecto de $\{f_i\}_{i\in I}$;

(2) si Ω es un espacio medible, una función $f : \Omega \to X$ es medible si y sólo si para todo $i \in I$ la función $f_i \circ f : \Omega \to \Omega_i$ es medible.

Demostración.

(1) $\implies$ (2) Si $f : \Omega \to X$ es medible entonces $f_i \circ f$ es medible para todo $i \in I$ por ser composición de funciones medibles. Ahora supongamos que $f_i \circ f$ es medible para todo $i \in I$. Dados $i \in I$ y $B \subseteq \Omega_i$ medible, tenemos

$$f^{-1}(f_i^{-1}(B)) = (f_i \circ f)^{-1}(B),$$

que es medible pues $f_i \circ f$ lo es. Como $\Sigma := \{A \subseteq X : f^{-1}(A) \text{ es medible}\}$ es una $\sigma-$álgebra en X que contiene a los conjuntos generadores de $\mathcal{A}$, entonces $\mathcal{A} \subseteq \Sigma$ y por lo tanto f es medible.

(2) $\implies$ (1) Llamemos $\hat{\mathcal{A}}$ a la $\sigma-$álgebra inicial respecto de $\{f_i\}_{i\in I}$. Ver que $\mathcal{A} \subseteq \hat{\mathcal{A}}$ es equivalente a ver que la identidad $id : (X, \hat{\mathcal{A}}) \to (X, \mathcal{A})$ es medible. Por (2), esto a su vez equivale a que para todo $i \in I$ la función $f_i \circ id = f_i : (X, \hat{\mathcal{A}}) \to \Omega_i$ sea medible, lo cual vale por definición de $\hat{\mathcal{A}}$. Por otra parte, para ver que $\hat{\mathcal{A}} \subseteq \mathcal{A}$, basta ver que para cada $i \in I$ la

función $f_i : (X, \mathcal{A}) \to \Omega_i$ es medible. Ahora bien, como la identidad $id' : (X, \mathcal{A}) \to (X, \mathcal{A})$ es medible, por (2) vale que para todo $i \in I$ la función $f_i \circ id' = f_i : (X, \mathcal{A}) \to \Omega_i$ es medible, lo cual termina la prueba. □

La propiedad (y demostración) anterior es completamente análoga a la propiedad que cumple una topología inicial. De hecho, se puede obtener una de la otra simplemente intercambiando las nociones de medibilidad con las correspondientes nociones topológicas. Es decir, cambiando conjuntos medibles por conjuntos abiertos, $\sigma-$álgebras por topologías, etc.

Definición 2.2.2. Sea $\mathcal{H}$ un espacio de Hilbert. Llamamos $\sigma-$**álgebra cilíndrica** de $\mathcal{H}$ y denotamos $\hat{C}(\mathcal{H})$ a la $\sigma-$álgebra inicial respecto de la familia $\{\langle \cdot, h\rangle : \mathcal{H} \to \mathbb{R}\}_{h\in\mathcal{H}}$, donde en $\mathbb{R}$ se considera la $\sigma-$álgebra de Borel.

La $\sigma-$álgebra cilíndrica busca definir una noción de medibilidad compatible con la estructura Hilbert, pero que sea más flexible que, digamos, la $\sigma-$álgebra de Borel. Las ventajas (y desventajas) de este enfoque se verán en la sección siguiente. Sin embargo, para espacios de Hilbert separables, estas dos nociones coinciden, como veremos en el siguiente resultado.

Proposición 2.2.3. *Si $\mathcal{H}$ es un espacio de Hilbert separable, entonces $\hat{C}(\mathcal{H})$ coincide con la $\sigma-$álgebra de Borel $\hat{B}(\mathcal{H})$.*

Demostración. Como las funciones $\langle \cdot, h\rangle$ son continuas, en particular son $\hat{B}(\mathcal{H})-$medibles y por lo tanto $\hat{C}(\mathcal{H}) \subseteq \hat{B}(\mathcal{H})$. Veamos la otra inclusión.

Sea $\{e_n\}_{n\in\mathbb{N}}$ una base ortonormal de $\mathcal{H}$. Notemos que la función $||\cdot||^2 = \sum_{n\in\mathbb{N}} |\langle\cdot, e_n\rangle|^2$ es $\hat{C}(\mathcal{H})-$medible por ser límite puntual de una sucesión de funciones $\hat{C}(\mathcal{H})-$medibles. Por otra parte, si $x \in \mathcal{H}$, la traslación $\tau_x : (\mathcal{H}, \hat{C}(\mathcal{H})) \to (\mathcal{H}, \hat{C}(\mathcal{H}))$ definida por $\tau_x(h) = h - x$ es medible pues para $h \in \mathcal{H}$ se tiene que $\langle \tau_x, h\rangle = \langle \cdot, h\rangle - \langle x, h\rangle$ es medible. Consecuentemente, para $x \in \mathcal{H}$ y $\varepsilon > 0$ tenemos que

$$B_\varepsilon(x) = \{h \in \mathcal{H} : \ ||h - x||^2 < \varepsilon^2\} \in \hat{C}(\mathcal{H}),$$

ya que $||\cdot - x||^2$ es $\hat{C}(\mathcal{H})-$medible. Como $\mathcal{H}$ es separable, todo abierto de $\mathcal{H}$ es unión contable de bolas abiertas y luego todo abierto de $\mathcal{H}$ es $\hat{C}(\mathcal{H})-$medible, lo cual termina de probar que $\hat{B}(\mathcal{H}) \subseteq \hat{C}(\mathcal{H})$. □

Observación 2.2.4. Si $\mathcal{H}$ es un espacio de Hilbert y $F \subseteq \mathcal{H}$ es un subespacio cerrado, entonces

$$\hat{C}(F) = \{A \cap F : A \in \hat{C}(\mathcal{H})\}.$$

En efecto, llamemos $\Sigma := \{A \cap F : A \in \hat{C}(\mathcal{H})\}$. Como $\{A \in \hat{C}(\mathcal{H}) : A \cap F \in \hat{C}(F)\}$ es una $\sigma-$álgebra que contiene a los conjuntos generadores de $\hat{C}(\mathcal{H})$, tenemos que $\Sigma \subseteq \hat{C}(F)$. Por otra parte, Σ es una $\sigma-$álgebra que contiene a los conjuntos generadores de $\hat{C}(F)$, por lo que $\hat{C}(F) \subseteq \Sigma$.

Finalizamos esta sección con algunos resultados técnicos sobre regularidad de medidas de probabilidad en espacios métricos. Éstos resultados serán necesarios para atacar la ley cero-uno de Driscoll.

Definición 2.2.5. Sea T un espacio métrico y sea $\mathbb{P}$ una medida boreliana en T. Decimos que $\mathbb{P}$ es **regular** si para cada $A \in \hat{B}(T)$ y $\varepsilon > 0$ existen conjuntos $F \subseteq A \subseteq G$, F cerrado y G abierto tales que $\mathbb{P}(G \setminus F) < \varepsilon$. Si además $\mathbb{P}(T) < +\infty$, decimos que $\mathbb{P}$ es una **medida de Radon** si para cada $A \in \hat{B}(T)$ vale que $\mathbb{P}(A) = \sup\{\mathbb{P}(K) : K \subseteq A, K \text{ compacto}\}$.

Nuestro próximo objetivo es ver que toda medida de probabilidad boreliana en un espacio de Hilbert separable es de Radon (ver Proposición 2.2.8). Para esto, necesitamos los siguientes lemas técnicos.

Lema 2.2.6. *Sean T un espacio métrico y $\mathbb{P}$ una medida boreliana finita en T. Entonces $\mathbb{P}$ es regular.*

Demostración. Sea Σ la colección de subconjuntos de T que cumple la definición de regularidad. Es decir, un subconjunto $A \subseteq T$ es un elemento de Σ si y sólo si para todo $\varepsilon > 0$ existen conjuntos $F \subseteq A \subseteq G$, F cerrado y G abierto tales que $\mathbb{P}(G \setminus F) < \varepsilon$. Debemos probar que $\hat{B}(T) \subseteq \Sigma$. Para ello, veamos que Σ es una $\sigma-$álgebra que contiene a los cerrados de T. Comencemos probando que Σ es una $\sigma-$álgebra.

Tomando $G = F = \varnothing$ se ve que $\varnothing \in \Sigma$. Por otra parte, si $A \in \Sigma$ y $\varepsilon > 0$, tenemos que existen $F \subseteq A \subseteq G$, F cerrado y G abierto tales que $\mathbb{P}(G \setminus F) < \varepsilon$. Pero entonces

$$T \setminus G \subseteq T \setminus A \subseteq T \setminus F.$$

Además, $T \setminus G$ es cerrado, $T \setminus F$ es abierto y $\mathbb{P}((T \setminus F) \setminus (T \setminus G)) = \mathbb{P}((T \setminus F) \cap G) = \mathbb{P}(G \setminus F) < \varepsilon$, lo cual prueba que $T \setminus A \in \Sigma$. Por último, si $(A_k)_{k\in\mathbb{N}}$ es una sucesión de conjuntos de Σ, dado $\varepsilon > 0$ tenemos que para cada $k \in \mathbb{N}$ existen $F_k \subseteq A_k \subseteq G_k$, F_k cerrado y G_k abierto tales que $\mathbb{P}(G_k \setminus F_k) < \varepsilon/2^{k+1}$. Sean

$$G = \bigcup_{k\in\mathbb{N}} G_k \quad \text{y} \quad E = \bigcup_{k\in\mathbb{N}} F_k.$$

Luego $E \subseteq A \subseteq G$, G es abierto y

$$\begin{aligned} \mathbb{P}(G \setminus E) &= \mathbb{P}\Big[\Big(\bigcup_{k\in\mathbb{N}} G_k\Big) \setminus \Big(\bigcup_{k\in\mathbb{N}} F_k\Big)\Big] \leq \mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} G_k \setminus F_k\Big) \\ &\leq \sum_{k\in\mathbb{N}} \mathbb{P}(G_k \setminus F_k) < \sum_{k\in\mathbb{N}} \frac{\varepsilon}{2^{k+1}} = \frac{\varepsilon}{2}. \end{aligned}$$

Por la continuidad de la medida, y teniendo en cuenta que $\mathbb{P}$ es finita, existe $m \in \mathbb{N}$ tal que $\mathbb{P}\big(E \setminus \bigcup_{k=1}^m F_k\big) < \varepsilon/2$. Tomando $F = \bigcup_{k=1}^m F_k$, $F \subseteq E \subseteq A$, F es cerrado y

$$\mathbb{P}(G \setminus F) \leq \mathbb{P}(G \setminus E) + \mathbb{P}(E \setminus F) < \varepsilon,$$

lo cual prueba que $\bigcup_{k\in\mathbb{N}} A_k \in \Sigma$. Esto termina de probar que Σ es una $\sigma-$álgebra.

Ahora, si $K \subseteq T$ es cerrado y no vacío, definamos para $k \in \mathbb{N}$ el conjunto $G_k := \{t \in T : d(t, K) < 1/k\}$. Luego G_k es abierto y $K = \bigcap_{k\in\mathbb{N}} G_k$. Como los conjuntos G_k son decrecientes, por la continuidad de la medida existe $m \in \mathbb{N}$ tal que $\mathbb{P}(G_m \setminus K) < \varepsilon$. Tomando $G = G_m$ y $F = K$, esto prueba que $K \in \Sigma$. □

Lema 2.2.7. *Sean T un espacio métrico separable y $\mathbb{P}$ una medida de probabilidad boreliana en T. Entonces para cada $\varepsilon > 0$ existe $K \subseteq T$ cerrado y totalmente acotado tal que $\mathbb{P}(K) > 1 - \varepsilon$.*

Demostración. Sean $\varepsilon > 0$ y $A \subseteq T$ un subconjunto denso y contable. Para cada $n \in \mathbb{N}$ tenemos que $T = \bigcup_{a\in A} \overline{B}_{1/n}(a)$. Por lo tanto, para cada $n \in \mathbb{N}$ existe $J_n \subseteq A$ finito tal que $\mathbb{P}\big(\bigcup_{a\in J_n} \overline{B}_{1/n}(a)\big) > 1 - \varepsilon/2^n$. Más aún, podemos suponer que $J_n \subseteq J_{n+1}$ para cada $n \in \mathbb{N}$. Definamos

$$K = \bigcap_{n\in\mathbb{N}} \bigcup_{a\in J_n} \overline{B}_{1/n}(a).$$

Notemos que K es cerrado por ser intersección de conjuntos cerrados, y es totalmente acotado pues para cada $n \in \mathbb{N}$, K está contenido en una unión finita de bolas de radio $1/n$. Por último,

$$\mathbb{P}(T \setminus K) = \mathbb{P}\Big[\bigcup_{n\in\mathbb{N}} T \setminus \Big(\bigcup_{a\in J_n} \overline{B}_{1/n}(a)\Big)\Big] \leq \sum_{n\in\mathbb{N}} \mathbb{P}\Big[T \setminus \Big(\bigcup_{a\in J_n} \overline{B}_{1/n}(a)\Big)\Big] < \sum_{n\in\mathbb{N}} \frac{\varepsilon}{2^n} = \varepsilon,$$

de donde $\mathbb{P}(K) > 1 - \varepsilon$. □

Proposición 2.2.8. *Si T es un espacio métrico completo y separable, y $\mathbb{P}$ es una medida de probabilidad boreliana en T, entonces $\mathbb{P}$ es una medida de Radon.*

Demostración. En vistas del Lema 2.2.7 existe una sucesión creciente $(K_n)_{n\in\mathbb{N}}$ de conjuntos cerrados y totalmente acotados (y por lo tanto compactos al ser T completo) tales que $\mathbb{P}\big(\bigcup_{n\in\mathbb{N}} K_n\big) = 1$. Si $A \subseteq T$ es boreliano y $\varepsilon > 0$, por el Lema 2.2.6 existe $F \subseteq A$ cerrado tal que $\mathbb{P}(A \setminus F) < \varepsilon/2$. Definiendo $F_n = K_n \cap F$, vale que F_n es compacto y $\mathbb{P}\big(F \setminus \bigcup_{n\in\mathbb{N}} F_n\big) = 0$ por lo que existe $m \in \mathbb{N}$ tal que $\mathbb{P}(F \setminus F_m) < \varepsilon/2$ (recordemos que los conjuntos K_n son crecientes). Luego $\mathbb{P}(A \setminus F_m) < \varepsilon$ y entonces $\mathbb{P}(A) < \varepsilon + \mathbb{P}(F_m)$, lo cual termina la prueba. □

2.3. Elementos aleatorios

Definición 2.3.1. Sea $\mathcal{H}$ un espacio de Hilbert. Un **elemento aleatorio** en $\mathcal{H}$ es una función $\xi : \Omega \to \mathcal{H}$ que es $(\mathcal{A}, \hat{C}(\mathcal{H}))-$medible, donde $(\Omega, \mathcal{A}, \mathbb{P})$ es un espacio de probabilidad. Decimos que ξ es **separablemente valuado** si $\xi(\Omega)$ es un subconjunto separable de $\mathcal{H}$. Si $\eta : \Omega \to \mathcal{H}$ es otro elemento aleatorio, decimos que η es una **versión** de ξ si para cada $h \in \mathcal{H}$ vale que $\langle \xi, h\rangle \overset{a.s.}{=} \langle \eta, h\rangle$.

Trabajar con la $\sigma-$álgebra cilíndrica permite que la definición de elemento aleatorio abarque más funciones que si se trabajara, digamos, con la $\sigma-$álgebra de Borel, manteniendo a su vez una cierta compatibilidad con la estructura de espacio de Hilbert. Más aún, verificar que una función $\xi : \Omega \to \mathcal{H}$ es efectivamente un elemento aleatorio es sencillo: simplemente hay que ver que $\langle \xi, h \rangle$ es una variable aleatoria (real) para todo $h \in \mathcal{H}$. Sin embargo, una de las grandes desventajas que surge de trabajar con esta $\sigma-$álgebra es que, en general, la norma $|| \cdot || : \mathcal{H} \to \mathbb{R}$ no es medible.

En varias ocasiones a lo largo del trabajo, exploramos las ventajas que proporciona trabajar en espacios de Hilbert separables. Ejemplos de esto son las Proposiciones 1.4.16 y 2.2.3. El interés por los elementos aleatorios separablemente valuados viene de que, al tomar valores en un espacio de Hilbert separable, éstos heredan algunas de las bondades de estos espacios. El siguiente teorema es una instancia de lo antedicho.

Teorema 2.3.2. *Si ξ es un elemento aleatorio separablemente valuado en un espacio de Hilbert $\mathcal{H}$, entonces la medida pushforward asociada $\mathbb{P}_\xi$ es una medida de Radon.*

Demostración. Sea $\mathcal{H}_0$ el subespacio cerrado de $\mathcal{H}$ generado por $\mathrm{Im}(\xi)$. Como por hipótesis $\mathrm{Im}(\xi)$ es separable, $\mathcal{H}_0$ es un espacio de Hilbert separable. Si $A \subseteq \mathcal{H}$ es boreliano, entonces $A \cap \mathcal{H}_0 \in \hat{B}(\mathcal{H}_0) = \hat{C}(\mathcal{H}_0)$ por la Proposición 2.2.3. Luego, por la Observación 2.2.4, existe $A' \in \hat{C}(\mathcal{H})$ tal que $A' \cap \mathcal{H}_0 = A \cap \mathcal{H}_0$. Con esto,

$$\xi^{-1}(A) = \xi^{-1}(A \cap \mathcal{H}_0) = \xi^{-1}(A' \cap \mathcal{H}_0) = \xi^{-1}(A'),$$

que resulta medible. Esto prueba que ξ es medible Borel y, por lo tanto, $\mathbb{P}_\xi$ es una medida de probabilidad boreliana en $\mathcal{H}$.

Ahora bien, si μ denota a la restricción de $\mathbb{P}_\xi$ a $\hat{B}(\mathcal{H}_0)$, entonces μ es una medida de probabilidad boreliana en $\mathcal{H}_0$ y, por la Proposición 2.2.7, es de Radon. Como $\mathbb{P}_\xi(A) = \mathbb{P}_\xi(A \cap \mathcal{H}_0) = \mu(A \cap \mathcal{H}_0)$, deducimos que $\mathbb{P}_\xi$ es de Radon. $\square$

A simple vista, podría parecer que la hipótesis de ser separablemente valuado es demasiado restrictiva como para ser aplicable a elementos aleatorios con algún grado de generalidad. El siguiente teorema muestra que esto no es del todo así. Si bien no daremos su demostración, mencionamos que el resultado se puede obtener uniendo [15, Corollary 5] con [15, Theorem IV.2.7].

Teorema 2.3.3. *Sea $\mathcal{H}$ un espacio de Hilbert y sea $\xi : \Omega \to \mathcal{H}$ un elemento aleatorio. Entonces ξ admite una versión separablemente valuada.*

Definición 2.3.4. Sean $\mathcal{H}$ un espacio de Hilbert, $0 < p < \infty$. Una medida μ en $\hat{C}(\mathcal{H})$ se dice de **orden p débil** si para todo $h \in \mathcal{H}$ se tiene que $\langle \cdot, h \rangle \in L^p(\mathcal{H}, \mu)$. Una medida μ boreliana en $\mathcal{H}$ se dice de **orden p fuerte** si $|| \cdot ||_\mathcal{H} \in L^p(\mathcal{H}, \mu)$ (notar que se requiere que μ sea boreliana para que $|| \cdot ||_\mathcal{H}$ sea medible). Si ξ es un elemento aleatorio en $\mathcal{H}$, decimos que ξ es de orden p débil (fuerte) si la medida *pushforward* $\mathbb{P}_\xi$ lo es. Para los casos $p = 1$ y $p = 2$, diremos que la medida, o el elemento aleatorio, es de primer o segundo orden débil (fuerte) respectivamente.

Recordemos que si $\xi : \Omega \to \mathbb{R}^n$ es un vector aleatorio, en general su esperanza (de existir) se define como $\mathbb{E}\xi := (\mathbb{E}\xi_1, \dots, \mathbb{E}\xi_n)$. Similarmente, se define su matriz de covarianzas como $\Sigma = (Cov(\xi_i, \xi_j))_{i,j}$. Los siguientes resultados buscan extender estas definiciones al contexto de elementos aleatorios en espacios de Hilbert.

Proposición 2.3.5. *Sean $\mathcal{H}$ un espacio de Hilbert, $1 \leq p < \infty$ y μ una medida en $\hat{C}(\mathcal{H})$ de orden p débil. Entonces el operador $T : \mathcal{H} \to L^p(\mathcal{H}, \mu)$ dado por $Th = \langle \cdot, h \rangle$ es acotado.*

Demostración. Por definición de orden débil, T es un operador lineal bien definido. Para ver que es acotado, veamos que su gráfico es cerrado.

Supongamos que $(h_n)_{n\in\mathbb{N}}$ es una sucesión en $\mathcal{H}$ tal que $h_n \to h$ y $Th_n \xrightarrow{L^p} f$ para cierta $f \in L^p(\mathcal{H}, \mu)$. Luego, existe una subsucesión $(Th_{n_k})_{k\in\mathbb{N}}$ que converge $\mu-$a.e. a f. Pero por la convergencia de $(h_n)_{n\in\mathbb{N}}$ vale que $\langle \cdot, h_{n_k} \rangle \to \langle \cdot, h \rangle$ puntualmente, de donde concluimos que $Th = \langle \cdot, h \rangle = f$ $\mu-$a.e., lo cual muestra que el gráfico de T es cerrado. □

Corolario 2.3.6. *Sean $\mathcal{H}$ un espacio de Hilbert, $1 \leq p < \infty$ y μ una medida de probabilidad en $\hat{C}(\mathcal{H})$ de orden p débil. Entonces, el funcional $\varphi : \mathcal{H} \to \mathbb{R}$ definido por*

$$\varphi(h) = \int_{\mathcal{H}} \langle x, h \rangle d\mu(x)$$

es acotado.

Demostración. Si llamamos $I : L^p(\mathcal{H}, \mu) \to \mathbb{R}$ al operador de integración dado por $If = \int_{\mathcal{H}} f d\mu$ y T al operador de la Proposición 2.3.5, entonces $\varphi = IT$ es acotado pues I y T lo son. □

Definición 2.3.7. Sean $\mathcal{H}$ un espacio de Hilbert y ξ un elemento aleatorio en $\mathcal{H}$ de primer orden débil. En vistas del Corolario 2.3.6, se tiene que el funcional $\mathcal{H} \to \mathbb{R}$ dado por $h \mapsto \int_{\mathcal{H}} \langle x, h \rangle d\mathbb{P}_\xi(x)$ es acotado. Luego, por el teorema de representación de Riesz, es de la forma $\langle \cdot, h_\xi \rangle$ para algún $h_\xi \in \mathcal{H}$. Definimos la **esperanza** de ξ como h_ξ, y la denotamos $\mathbb{E}\xi$.

Observación 2.3.8. Notemos que la esperanza de un elemento aleatorio ξ queda determinada por la propiedad

$$\int_{\mathcal{H}} \langle x, h \rangle d\mathbb{P}_\xi(x) = \langle \mathbb{E}\xi, h \rangle, \quad \forall h \in \mathcal{H}.$$

Como la primera integral no es otra cosa que la esperanza de una variable aleatoria, eso se puede reescribir como

$$\mathbb{E}_{\mathbb{P}_\xi} \langle \cdot, h \rangle = \langle \mathbb{E}\xi, h \rangle, \quad \forall h \in \mathcal{H}. \tag{2.1}$$

Ahora bien, un resultado estándar sobre medidas *pushforward* [6, Theorem 1.6.9] nos da

$$\int_{\mathcal{H}} \langle x, h \rangle d\mathbb{P}_\xi(x) = \int_{\Omega} \langle \xi(\omega), h \rangle d\mathbb{P}(\omega).$$

Entonces, otra forma de escribir la propiedad que define a la esperanza de ξ es

$$\mathbb{E}\langle \xi, h\rangle = \langle \mathbb{E}\xi, h\rangle, \quad \forall h \in \mathcal{H}. \tag{2.2}$$

La diferencia entre las expresiones (2.1) y (2.2) es que $\mathbb{E}_{\mathbb{P}_\xi}\langle \cdot, h\rangle$ es la esperanza de una variable aleatoria definida en $\mathcal{H}$ respecto de la medida $\mathbb{P}_\xi$, mientras que $\mathbb{E}\langle \xi, h\rangle$ es la esperanza de una variable aleatoria definida en Ω respecto de la medida $\mathbb{P}$.

Antes de seguir con la definición de covarianza de un elemento aleatorio, debemos probar dos resultados generales sobre formas bilineales. Primero, recordemos que si X, Y son espacios normados, una forma bilineal $B : X \times Y \to \mathbb{K}$ se dice **acotada** si existe $C > 0$ tal que para cualesquiera $x \in X$, $y \in Y$, vale que

$$|B(x, y)| \leq C||x||_X||y||_Y.$$

Llamamos **norma** de B al valor

$$||B|| := \sup_{\substack{||x||_X \leq 1 \\ ||y||_Y \leq 1}} |B(x, y)|.$$

Proposición 2.3.9. *Sean X, Y espacios de Banach, y sea $B : X \times Y \to \mathbb{K}$ una forma bilineal. Entonces B es acotada si y sólo si $B(x, \cdot)$ y $B(\cdot, y)$ son acotados para cualesquiera $x \in X$, $y \in Y$.*

Demostración.

($\Longrightarrow$) Como B es bilineal, tenemos que para $x \in X$ la función $B(x, \cdot) : Y \to \mathbb{K}$ es lineal. Además, si $y \in Y$,

$$|B(x, \cdot)(y)| = |B(x, y)| \leq ||B||\ ||x||_X\ ||y||_Y,$$

de donde $B(x, \cdot)$ es un funcional acotado de Y con $||B(x, \cdot)||_{Y'} \leq ||B||\ ||x||_X$. Simétricamente se ve que $B(\cdot, y)$ es un funcional acotado de X para cualquier $y \in Y$.

($\Longleftarrow$) Por hipótesis, tenemos que el funcional $B(x, \cdot)$ es un elemento de Y' para cualquier $x \in X$. Sea entonces $\mathcal{F} = \{B(x, \cdot) : x \in \overline{B}_X\} \subseteq Y'$. Dados $x \in \overline{B}_X$ e $y \in Y$, tenemos que $|B(x, y)| \leq ||B(\cdot, y)||_{X'}$. Dicho de otra forma, para cada $y \in Y$ vale que $\sup_{x \in \overline{B}_X} |B(x, y)| \leq ||B(\cdot, y)||_{X'} < +\infty$. Por el principio de acotación uniforme, existe $C > 0$ tal que para todo $x \in \overline{B}_X$ vale que $||B(x, \cdot)||_{Y'} \leq C$. Consecuentemente, para $x \in X$ vale que $||B(x, \cdot)||_{Y'} \leq C||x||_X$. Juntando todo, llegamos a que

$$|B(x, y)| \leq ||B(x, \cdot)||_{Y'}||y||_Y \leq C||x||_X||y||_Y,$$

como queríamos probar. □

Proposición 2.3.10. *Sea $\mathcal{H}$ un espacio de Hilbert real y sea $B : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ una forma bilineal acotada. Entonces valen las siguientes afirmaciones.*

(1) Existe un único operador acotado $T : \mathcal{H} \to \mathcal{H}$ tal que $B(x, y) = \langle x, Ty \rangle$ para todos $x, y \in \mathcal{H}$.

(2) Si B es simétrica, el operador T es autoadjunto.

(2) Si B es semidefinida positiva, entonces $T \geq 0$.

Demostración.

(1) Por el teorema de representación de Riesz, para cada $y \in \mathcal{H}$ existe un único $Ty \in \mathcal{H}$ tal que $B(\cdot, y) = \langle \cdot, Ty \rangle$, pues B es acotada (notar que de la unicidad de Ty se desprende la unicidad de T). Por la bilinealidad de B, $T : \mathcal{H} \to \mathcal{H}$ resulta un operador lineal. Además, si $h \in \mathcal{H}$ es tal que $Th \neq 0$, tenemos que

$$||Th||^2 = \langle Th, Th \rangle = B(Th, h) = |B(Th, h)| \leq ||B|| \, ||Th|| \, ||h|| \implies ||Th|| \leq ||B|| \, ||h||.$$

Como la desigualdad también vale si $Th = 0$, esto prueba que T es acotado.

(2) Dados $x, y \in \mathcal{H}$,

$$\langle x, Ty \rangle = B(x, y) = B(y, x) = \langle y, Tx \rangle = \langle Tx, y \rangle.$$

(3) Si $x \in \mathcal{H}$,

$$\langle x, Tx \rangle = B(x, x) \geq 0,$$

lo cual termina la prueba. □

Si $\mathcal{H}$ es un espacio de Hilbert y μ es una medida de probabilidad de segundo orden débil en $\hat{C}(\mathcal{H})$, tenemos una forma bilineal bien definida dada por $(f, g) \mapsto Cov(\langle \cdot, f \rangle, \langle \cdot, g \rangle)$. Aplicaremos los dos resultados anteriores a esta forma bilineal, para finalmente llegar a la definición de covarianza en este contexto.

Proposición 2.3.11. *Sea $\mathcal{H}$ un espacio de Hilbert, y sea μ una medida de probabilidad definida en $\hat{C}(\mathcal{H})$ de segundo orden débil. Consideramos la forma bilineal*

$$r_\mu(f, g) = \int_{\mathcal{H}} \langle x, f \rangle \langle x, g \rangle d\mu(x) - \int_{\mathcal{H}} \langle x, f \rangle d\mu(x) \int_{\mathcal{H}} \langle x, g \rangle d\mu(x).$$

Entonces r_μ está bien definida, es acotada, simétrica y semidefinida positiva.

Demostración. Como μ es de segundo orden débil, tenemos que $\langle \cdot, f \rangle, \langle \cdot, g \rangle \in L^2(\mathcal{H}, \mu) \subseteq L^1(\mathcal{H}, \mu)$. De la desigualdad de Hölder se deduce que $\langle \cdot, f \rangle \langle \cdot, g \rangle \in L^1(\mathcal{H}, \mu)$ y, consecuentemente, todas las integrales involucradas en la definición de μ son finitas. Esto justifica la buena definición de μ. La simetría de μ se desprende de la definición, y la positividad de que $r_\mu(f, f) = \mathbb{V}\langle \cdot, f \rangle \geq 0$, donde $\mathbb{V}X$ denota a la varianza de una variable aleatoria X. Veamos la acotación.

En vistas de la Proposición 2.3.9 y de la simetría de r_μ, basta con probar que para cada $g \in \mathcal{H}$, el funcional $r_\mu(\cdot, g)$ es acotado. Para ello, llamemos $T : \mathcal{H} \to L^2(\mathcal{H}, \mu)$ al operador dado por $Th = \langle \cdot, h \rangle$, que es acotado por la Proposición 2.3.5, y $M_g : L^2(\mathcal{H}, \mu) \to L^1(\mathcal{H}, \mu)$ al operador dado por $M_g f = \langle \cdot, g \rangle f$, que es acotado por la desigualdad de Hölder. Si $I : L^1(\mathcal{H}, \mu) \to \mathbb{R}$ denota al operador de integración y $J : L^2(\mathcal{H}, \mu) \to L^1(\mathcal{H}, \mu)$ es la inclusión, tenemos entonces que

$$r_\mu(\cdot, g) = IM_g T - \int_{\mathcal{H}} \langle x, g \rangle d\mu(x) IJT,$$

por lo que $r_\mu(\cdot, g)$ es acotado. □

Definición 2.3.12. Sea $\mathcal{H}$ un espacio de Hilbert, y sea μ una medida de probabilidad definida en $\hat{C}(\mathcal{H})$ de segundo orden débil. Si r_μ denota a la forma bilineal de la Proposición 2.3.11, entonces por la Proposición 2.3.10 existe un único operador $\Theta \geq 0$ tal que $r_\mu(f, g) = \langle f, \Theta g \rangle, \forall f, g \in \mathcal{H}$. El operador Θ se denomina el **operador de covarianza** de la medida μ. Si ξ es un elemento aleatorio en $\mathcal{H}$ de segundo orden débil, el operador de covarianza de ξ se define como el operador de covarianza de $\mathbb{P}_\xi$. En este último caso, el operador Θ queda determinado por la propiedad

$$\langle f, \Theta g \rangle = \mathbb{E}\langle \xi, f \rangle \langle \xi, g \rangle - \mathbb{E}\langle \xi, f \rangle \mathbb{E}\langle \xi, g \rangle = Cov(\langle \xi, f \rangle, \langle \xi, g \rangle), \forall f, g \in \mathcal{H}. \tag{2.3}$$

El lector puede verificar que, si ξ es un elemento aleatorio en un espacio de Hilbert $\mathcal{H}$ y η es una versión de ξ, entonces $\mathbb{E}\xi = \mathbb{E}\eta$ y sus operadores de covarianza coinciden.

2.4. La ley cero-uno de Driscoll

En esta sección combinaremos finalmente las nociones probabilísticas de las secciones anteriores con los RKHSs para llegar a la generalización de la ley cero-uno de Driscoll, que esencialmente dice que la probabilidad de que un proceso gaussiano pertenezca a un RKHS dado es 0 ó 1. Concretamente, el objetivo de esta sección es demostrar el siguiente teorema.

Teorema 2.4.1 (Generalización del teorema de Driscoll)**.** *Sea $X : \Omega \times T \to \mathbb{R}$ un proceso gaussiano con función de media m y función de covarianza K, y sea R un núcleo en T tal que $K \leq R$ y $m \in \mathcal{H}(R)$.*

1. *Si $K \ll R$, entonces existe una versión $\tilde{X}$ de X tal que $\tilde{X}_\bullet \in \mathcal{H}(R)$ casi seguramente.*
2. *Si $K \not\ll R$, entonces $X_\bullet \in \mathcal{H}(R)$ con probabilidad 0.*

Para tal fin, debemos estudiar las relaciones entre procesos estocásticos y elementos aleatorios en RKHSs. Comencemos viendo cómo es posible definir uno a partir del otro.

Lema 2.4.2. *Sean T un conjunto, $\mathcal{H}(R)$ un RKHS en T y $X : \Omega \times T \to \mathbb{R}$ un proceso estocástico (no necesariamente de segundo orden) con casi todas sus trayectorias en $\mathcal{H}(R)$. Si $\Omega' = \{X_\bullet \in \mathcal{H}(R)\}$, entonces cualquier extensión de*

$$\xi(\omega) = X_\bullet(\omega), \quad \omega \in \Omega' \tag{2.4}$$

a Ω define un elemento aleatorio en $\mathcal{H}(R)$, y vale que

$$X_t \overset{a.s.}{=} \langle \xi, R_t \rangle, \quad t \in T. \tag{2.5}$$

Recíprocamente, si $\xi : \Omega \to \mathcal{H}(R)$ es un elemento aleatorio, entonces (2.5) define un proceso estocástico (no necesariamente de segundo orden) que verifica (2.4).

Demostración. Primero supongamos que tenemos un proceso estocástico X como en el enunciado y sea ξ como en (2.4). De la definición de núcleo reproductor se ve que vale (2.5). En particular, $\langle \xi, R_t \rangle$ es una variable aleatoria para todo $t \in T$ (recordemos que estamos asumiendo que $\mathbb{P}$ es completa), y entonces por linealidad vale que $\langle \xi, h \rangle$ es una variable aleatoria para todo $h \in \mathcal{W} = [R_t]_{t \in T}$. Por último, para $h \in \mathcal{H}(R)$, si $(h_n)_{n \in \mathbb{N}}$ es una sucesión en $\mathcal{W}$ tal que $h_n \to h$, entonces $\langle \xi, h_n \rangle \xrightarrow{a.s.} \langle \xi, h \rangle$ por lo que $\langle \xi, h \rangle$ resulta ser una variable aleatoria. En vistas de la Proposición 2.2.1, esto termina de probar que ξ es un elemento aleatorio en $\mathcal{H}(R)$.

Ahora supongamos que ξ es un elemento aleatorio en $\mathcal{H}(R)$ y definamos X como en (2.5) (donde ahora la igualdad es en todo punto). Nuevamente, por la Proposición 2.2.1, cada X_t es una variable aleatoria y luego X es efectivamente un proceso estocástico. Por último, si $\omega \in \Omega$ y $t \in T$, entonces $\xi(\omega)(t) = \langle \xi(\omega), R_t \rangle = X_t(\omega) = X(\omega, t)$, de donde se ve que vale (2.4) con $\Omega' = \Omega$. □

El siguiente teorema muestra que, si tenemos un elemento aleatorio ξ en un RKHS $\mathcal{H}(R)$, entonces el proceso estocástico definido por ξ según el Lema 2.4.2 hereda algunas de las propiedades de ξ.

Teorema 2.4.3. *Sea $\mathcal{H}(R)$ un RKHS en un conjunto T, ξ un elemento aleatorio en $\mathcal{H}(R)$ y X el proceso estocástico que define según (2.5).*

(1) Si ξ es de primer orden débil, entonces X es de primer orden, y si llamamos m a su función de media, vale que

$$m(t) = \mathbb{E}[X_t] = \langle \mathbb{E}\xi, R_t \rangle = (\mathbb{E}\xi)(t) \quad \forall t \in T.$$

En particular, m coincide con $\mathbb{E}\xi$, y por lo tanto es un elemento de $\mathcal{H}(R)$.

(2) Si ξ es de segundo orden débil, entonces X es de segundo orden. En este caso, si Θ es el operador de covarianza de ξ y K es la función de covarianza de X, vale que $K(s,t) = \langle \Theta R_s, R_t \rangle$. En particular, $K \leq R$ y Θ es el operador de dominancia.

Demostración.

(1) Como ξ es de primer orden débil vale que $\mathbb{E}|X_t| = \mathbb{E}|\langle \xi, R_t \rangle| < +\infty$, lo cual prueba que X es de primer orden. Además, para $t \in T$, por la Observación 2.3.8 tenemos que $m(t) = \mathbb{E}[X_t] = \mathbb{E}\langle \xi, R_t \rangle = \langle \mathbb{E}\xi, R_t \rangle = (\mathbb{E}\xi)(t)$ de donde se desprende el resto del enunciado.

(2) Como ξ es de segundo orden débil vale que $\mathbb{E}|X_t|^2 = \mathbb{E}|\langle \xi, R_t\rangle|^2 < +\infty$, lo cual prueba que X es de segundo orden. Por otra parte, si $s, t \in T$ entonces $K(s,t) = Cov(X_s, X_t) = Cov(\langle \xi, R_s\rangle, \langle \xi, R_t\rangle) = \langle \Theta R_s, R_t\rangle$, donde en la última igualdad usamos (2.3). El resto del enunciado es consecuencia inmediata de la Proposición 1.4.9. $\square$

Ninguna de las recíprocas del Teorema 2.4.3 es cierta. En [11, Section 2] se exhiben contraejemplos.

Recordemos que, a lo largo de este capítulo, todos los procesos estocásticos se asumen de segundo orden.

Dado un proceso estocástico $X : \Omega \times T \to \mathbb{R}$, existen dos espacios de Hilbert naturalmente asociados a él. Uno de ellos es $\mathcal{H}(K)$, donde K es la función de covarianza de X. El otro es el subespacio cerrado de $L^2(\Omega)$ generado por $\{X_t\}_{t\in T}$. El siguiente resultado establece que, si el proceso tiene media 0, éstos son esencialmente el mismo espacio.

Proposición 2.4.4 (isometría de Loève)**.** *Sea $\{X_t\}_{t\in T}$ un proceso estocástico con media 0 y función de covarianza K, definido en un espacio de probabilidad Ω. Llamemos $\mathcal{H}$ al subespacio cerrado de $L^2(\Omega)$ generado por $\{X_t\}_{t\in T}$. Entonces se tiene un isomorfismo isométrico $J : \mathcal{H}(K) \to \mathcal{H}$ que cumple*

$$JK_t = X_t. \tag{2.6}$$

Demostración. Definimos $J : [K_t]_{t\in T} \to [X_t]_{t\in T}$ como $JK_t = X_t$ y extendido por linealidad. Notemos que, como

$$\begin{aligned}\left\|\sum_i \alpha_i K_{t_i}\right\|_K^2 &= \sum_{i,j} \alpha_i\alpha_j K(t_j, t_i) = \sum_{i,j} \alpha_i\alpha_j Cov(X_{t_j}, X_{t_i}) = \sum_{i,j} \alpha_i\alpha_j \mathbb{E}[X_{t_j} X_{t_i}] \\ &= \left\|\sum_i \alpha_i X_{t_i}\right\|_{L^2(\Omega)}^2,\end{aligned}$$

J está bien definido y es de hecho un isomorfismo isométrico. Por lo tanto, se extiende a un isomorfismo isométrico $J : \mathcal{H}(K) \to \mathcal{H}$. $\square$

Antes de seguir con el próximo resultado, debemos hacer una serie de aclaraciones.

Definición 2.4.5. Si $\mathcal{H}$ un espacio de Hilbert, una medida de probabilidad $\mathbb{P}$ en $\hat{C}(\mathcal{H})$ se dice **gaussiana** si para cada $h \in \mathcal{H}$ la medida *pushforward* $\mathbb{P}_{\langle \cdot, h\rangle}$ es una medida gaussiana en $\mathbb{R}$. Por otra parte, un elemento aleatorio $\xi : \Omega \to \mathcal{H}$ se dice **gaussiano** si para cada $h \in \mathcal{H}$ la variable aleatoria $\langle \xi, h\rangle$ es gaussiana.

De las definiciones se desprende que si ξ es un elemento aleatorio gaussiano en $\mathcal{H}$, entonces $\mathbb{P}_\xi$ es una medida de probabilidad gaussiana en $\mathcal{H}$.

Un resultado técnico pero importante del área es el *teorema de Mourier-Prokhorov*. Su enunciado preciso requeriría desarrollar la teoría de funcionales característicos de medidas de probabilidad en espacios de Banach, lo cual está por fuera del alcance de este trabajo. Sin embargo, de este teorema se desprende el siguiente hecho: si $\mathcal{H}$ es un espacio de Hilbert y $\mathbb{P}$ es una medida de probabilidad gaussiana y de Radon en la $\sigma-$álgebra boreliana $\hat{B}(\mathcal{H})$, entonces el operador de covarianza asociado es nuclear. Para un tratamiento completo sobre el tema, se puede consultar [15] (el teorema de Mourier-Prokhorov está enunciado como Theorem IV.2.4). La segunda afirmación del siguiente teorema se basa en este resultado.

Teorema 2.4.6. *Sea $X = \{X_t\}_{t\in T}$ un proceso gaussiano con función de media m y función de covarianza K. Sea $\mathcal{H}(R)$ un RKHS en T tal que $m \in \mathcal{H}(R)$. Si las trayectorias de X pertenecen a $\mathcal{H}(R)$ casi seguramente, entonces el elemento aleatorio que define X es gaussiano, y vale que $R \gg K$.*

Demostración. Cambiando X por $X - m$, podemos suponer que $m = 0$. Sea ξ el elemento aleatorio definido por X. Queremos ver que $\langle \xi, f\rangle$ es una variable aleatoria gaussiana para toda $f \in \mathcal{H}(R)$.

Notemos que si $f = \sum_i \alpha_i R_{t_i}$, entonces

$$\langle \xi, f\rangle = \sum_i \alpha_i \langle \xi, R_{t_i}\rangle = \sum_i \alpha_i X_{t_i}$$

es gaussiana pues el proceso X es gaussiano. Para $f \in \mathcal{H}(R)$, tomemos $(f_n)_{n\in\mathbb{N}} \subseteq [R_t]_{t\in T}$ tal que $f_n \xrightarrow{\|\cdot\|} f$. Luego, $\langle \xi, f_n\rangle \xrightarrow{a.s.} \langle \xi, f\rangle$ y entonces $\langle \xi, f\rangle$ es gaussiana por ser límite casi seguro de variables gaussianas.

Por los Teoremas 2.3.2 y 2.3.3, podemos reducirnos al caso en que la medida $\mathbb{P}_\xi$ es (gaussiana y) de Radon. Si Θ denota al operador de covarianza de ξ, por un lado Θ es nuclear por el teorema de Mourier-Prokhorov. Por el otro, por la parte (2) del Teorema 2.4.3, vale que $R \geq K$ y que Θ es el operador de dominancia de R sobre K, lo cual prueba que $R \gg K$. □

Para atacar la segunda mitad de la ley cero-uno de Driscoll, hacen falta dos resultados más. El primero es el siguiente teorema.

Teorema 2.4.7 (Kallianpur)**.** *Sean T un espacio métrico completo y separable, $H \subseteq \mathbb{R}^T$ un subespacio real y consideremos en H la $\sigma-$álgebra inicial respecto de las evaluaciones $\{ev_t : H \to \mathbb{R}\}_{t\in T}$. Sea $\mathbb{P}$ una medida de probabilidad en H que hace de $\{ev_t\}_{t\in T}$ un proceso gaussiano de media 0. Llamemos $K(s,t) = \mathbb{E}[ev_t ev_s]$ a la función de covarianza del proceso $\{ev_t\}_{t\in T}$ y supongamos que*

(i) K *es una función continua en* $T \times T$;

(ii) $\mathcal{H}(K) \subseteq H$.

Si M es un $\mathbb{Q}-$subespacio medible de H, entonces $\mathbb{P}(M) = 0$ o 1.

El Teorema 2.4.7 se encuentra en [10, Theorem 1], y es una de las primeras “leyes cero-uno” que fueron demostradas para procesos gaussianos.

El segundo resultado necesario es el siguiente lema técnico sobre dominancia de núcleos.

Lema 2.4.8. *Sea T un conjunto, y sean K, R dos núcleos definidos en T tales que $K \leq R$. Para cada $S \subseteq T$, llamemos K_S y R_S a las restricciones de K y R a S. Si $K_S \ll R_S$ para todo $S \subseteq T$ contable, entonces $K \ll R$.*

Demostración. Supongamos que $K \not\ll R$ y sea $\{e_i\}_{i\in I}$ una base ortonormal de $\mathcal{H}(R,T)$. Si $L : \mathcal{H}(R,T) \to \mathcal{H}(R,T)$ denota al operador de dominancia de R sobre K, tenemos que

$$\sum_{i\in I} \langle Le_i, e_i\rangle_R = +\infty,$$

ya que L no es nuclear. Luego, para cada $n \in \mathbb{N}$, existe $I_n \subseteq I$ finito tal que $\sum_{i\in I_n} \langle Le_i, e_i\rangle_R \geq n$. Definiendo $J = \bigcup_{n\in\mathbb{N}} I_n$, vale que J es contable y, como la serie es de términos positivos, resulta

$$\sum_{i\in J} \langle Le_i, e_i\rangle_R = +\infty.$$

Para cada $i \in J$, sea $(f_k^i)_{k\in\mathbb{N}}$ una sucesión en $[R_t]_{t\in T}$ tal que $f_k^i \xrightarrow{k\to\infty} e_i$. Cada f_k^i es de la forma

$$f_k^i = \sum_{t\in T_k^i} \alpha_{k,t}^i R_t,$$

con $\alpha_{k,t}^i \in \mathbb{R}$ y $T_k^i \subseteq T$ finito. Definamos

$$S = \bigcup_{i\in J} \bigcup_{k\in\mathbb{N}} T_k^i.$$

El conjunto S es contable por ser una unión contable de conjuntos contables. La prueba termina viendo que $K_S \not\ll R_S$.

Llamemos $F := \overline{[R_t]}_{t\in S}$. Primero notemos que, por construcción, vale que $\{e_i\}_{i\in J} \subseteq F$. Al ser $\{e_i\}_{i\in J}$ un conjunto ortonormal en F, se extiende a una base ortonormal $\{e_i\}_{i\in J'}$ de F, con $J \subseteq J'$. Por otra parte, como una función $f \in \mathcal{H}(R,T)$ se anula en S si y sólo si $\langle f, R_t\rangle_R = 0$ para todo $t \in S$, concluimos que $F = \mathcal{H}(R,T) \ominus \{f \in \mathcal{H}(R,T) : f|_S = 0\}$. Por el Lema 1.4.18, sabemos que $K_S \leq R_S$, y si L_S es el operador de dominancia de R_S sobre K_S, por la Proposición 1.4.19 tenemos que

$$\mathrm{Tr}(L_S) = \mathrm{Tr}(L|_F) = \sum_{i\in J'} \langle Le_i, e_i\rangle_R \geq \sum_{i\in J} \langle Le_i, e_i\rangle_R = +\infty,$$

lo cual prueba que $K_S \not\ll R_S$. □

El siguiente resultado es el punto 2 del Teorema 2.4.1.

Teorema 2.4.9. *Sea $X = \{X_t\}_{t\in T}$ un proceso gaussiano definido en un espacio de probabilidad Ω con función de media m y función de covarianza K. Sea R un núcleo en T tal que $K \leq R$ y $m \in \mathcal{H}(R)$. Si $K \not\ll R$, entonces $\mathbb{P}(X_\bullet \in \mathcal{H}(R)) = 0$.*

Demostración. Cambiando X por el proceso $X - m$, podemos suponer que $m = 0$. Como $K \not\ll R$, por el Lema 2.4.8, existe $S \subseteq T$ contable tal que $K_S \not\ll R_S$, donde K_S y R_S son las restricciones de los operadores K y R a S. Llamemos Y a la restricción del proceso X a S. En vistas de la Proposición 1.3.6, vale que

$$\{X_\bullet \in \mathcal{H}(R,T)\} \subseteq \{Y_\bullet \in \mathcal{H}(R,S)\},$$

y luego basta probar que $\mathbb{P}(Y_\bullet \in \mathcal{H}(R,S)) = 0$. Esto nos permite reducirnos al caso en que T es contable. Más aún, si T fuera finito, entonces los espacios $\mathcal{H}(R,T)$ y $\mathcal{H}(K,T)$ tendrían dimensión finita, por lo que necesariamente debería suceder que $K \ll R$, contrario a nuestra hipótesis. Luego, nos podemos reducir al caso en que T es numerable, digamos $T = \{t_n\}_{n\in\mathbb{N}}$.

Buscamos aplicar el Teorema 2.4.7. Tomemos $H = \mathbb{R}^T$ con la $\sigma-$álgebra inicial respecto de las evaluaciones $\{ev_t\}_{t\in T}$, y denotemos $X_\bullet : \Omega \to \mathbb{R}^T$ a la función dada por $\omega \mapsto X_\bullet(\omega) = X(\omega, \cdot)$. Como para cada $t \in T$ vale que $ev_t \circ X_\bullet = X_t$ es una función medible, la función $X_\bullet$ es medible. Podemos considerar entonces en H la medida *pushforward* $\mathbb{P}_{X_\bullet}$. Si $t_1, \ldots, t_n \in T$, $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ y $B \subseteq \mathbb{R}$ es boreliano, entonces

$$\begin{aligned} \mathbb{P}_{X_\bullet}(\alpha_1 ev_{t_1} + \cdots + \alpha_n ev_{t_n} \in B) &= \mathbb{P}((\alpha_1 ev_{t_1} + \cdots + \alpha_n ev_{t_n}) \circ X_\bullet \in B) \\ &= \mathbb{P}(\alpha_1 X_{t_1} + \cdots + \alpha_n X_{t_n} \in B). \end{aligned}$$

Como X es un proceso gaussiano, deducimos que $\mathbb{P}_{X_\bullet}$ hace de $\{ev_t\}_{t\in T}$ un proceso gaussiano. Además, utilizando nuevamente [6, Theorem 1.6.9], tenemos que, para $t \in T$,

$$\int_H ev_t d\mathbb{P}_{X_\bullet} = \int_\Omega ev_t \circ X_\bullet d\mathbb{P} = \int_\Omega X_t d\mathbb{P} = \mathbb{E}[X_t] = 0,$$

de donde el proceso $\{ev_t\}_{t\in T}$ tiene media 0.

Equipando a T con la métrica discreta, T resulta un espacio métrico completo y separable (por ser numerable). La función de covarianza del proceso $\{ev_t\}_{t\in T}$ es continua pues el espacio $T \times T$ es discreto, y como $H = \mathbb{R}^T$, es inmediato que $\mathcal{H}(K) \subseteq H$. Por último, veamos que $M = \mathcal{H}(R) \subseteq \mathbb{R}^T$ es medible.

Por la Proposición 1.3.2, una función $f \in \mathbb{R}^T$ es un elemento de $\mathcal{H}(R,T)$ si y sólo si verifica la condición de Fortet (1.3). Por un argumento de densidad, el supremo de (1.3) se puede tomar sobre elecciones $\alpha_i \in \mathbb{Q}$, de donde tenemos que

$$\mathcal{H}(R) = \bigcup_{k\in\mathbb{N}} \bigcap_{n\in\mathbb{N}} \bigcap_{\alpha_i\in\mathbb{Q}} \left\{ f \in \mathbb{R}^T : \frac{|\sum_{i=1}^n \alpha_i f(t_i)|^2}{\sum_{i,j=1}^n \alpha_i\alpha_j R(t_i,t_j)} \le k \right\}. \tag{2.7}$$

Para $k, n \in \mathbb{N}$ fijos, y una elección posible de coeficientes $\alpha_i \in \mathbb{Q}$, el conjunto interior de (2.7) es medible. Como todas las uniones e intersecciones involucradas son numerables, concluimos que $\mathcal{H}(R)$ es medible.

Por el Teorema 2.4.7, tenemos que $\mathbb{P}(X_\bullet \in \mathcal{H}(R)) = \mathbb{P}_{X_\bullet}(\mathcal{H}(R)) = 0$ ó 1. Si esta probabilidad fuera 1, por el Teorema 2.4.6 valdría que $K \ll R$, contradiciendo nuestra hipótesis. Luego, $\mathbb{P}(X_\bullet \in \mathcal{H}(R)) = 0$ como queríamos probar. $\square$

La prueba anterior corrige una omisión cometida en [11, Theorem 7.3]. En ese trabajo, los autores asumen implícitamente que, si K, R son dos núcleos en un conjunto T tales que $K \leq R$, $K \not\ll R$, $\mathcal{H}(R,T)$ es de dimensión infinita y $S \subseteq T$ es un subconjunto contable tal que $\mathcal{H}(R,S)$ sigue siendo de dimensión infinita, entonces $K|_{S\times S} \not\ll R_{S\times S}$. Más aún, esa misma omisión se comete en [8, Proposition 4.5.1] a la hora de demostrar el mismo teorema. El siguiente ejemplo muestra que esto no es cierto, lo cual justifica la necesidad del Lema 2.4.8. No conocemos ninguna otra fuente que reconozca este error.

Ejemplo 2.4.10. Sea $T = \mathbb{N}$ y consideremos en T los núcleos $R(m,n) = \delta_{m,n}$ y

$$K(m,n) = \begin{cases} 2^{-n}\delta_{m,n} & \text{si } n \text{ es par} \\ \delta_{m,n} & \text{si } n \text{ es impar} \end{cases}.$$

Por el comentario hecho al final del Ejemplo 1.4.8, vale que $K \leq R$ pero $K \not\ll R$. Tomando $S = 2\mathbb{N}$, de ese mismo comentario se deduce que $K|_{S\times S} \ll R|_{S\times S}$. Es claro que los espacios $\mathcal{H}(R,T)$ y $\mathcal{H}(R,S)$ son ambos de dimensión infinita.

El siguiente teorema es debido a M. Lukić y J. Beder [11, Theorem 5.1]. Lo sobresaliente del resultado es que logran probar la otra mitad de la ley cero-uno para procesos estocásticos generales. Los enfoques previos a su trabajo lograban únicamente lidiar con procesos gaussianos definidos sobre espacios métricos separables y con función de covarianza continua.

Teorema 2.4.11. *Sea $X = \{X_t\}_{t\in T}$ un proceso estocástico con función de covarianza K y sea R un núcleo en T tal que $R \gg K$ con operador de dominancia L. Supongamos además que la función de media m de X pertenece a $\mathcal{H}(R)$. Entonces existe una versión Y de X cuyas trayectorias pertenecen a $\mathcal{H}(R)$ casi seguramente.*

Demostración. Cambiando X por $X - m$ podemos suponer que $m = 0$. Más aún, por el Lema 1.4.10 existe un núcleo R_1 en T tal que $K \ll R_1 \leq R$ y $\mathcal{H}(R_1)$ es separable. Esto nos permite suponer sin pérdida de generalidad que $\mathcal{H}(R)$ es separable. Denotemos $(\Omega, \mathcal{A}, \mathbb{P})$ al espacio de probabilidad sobre el que está definido X.

Fijemos $T_0 \subseteq T$ un subconjunto $R-$Hamel y $S_0 = \{s_k\}_{k\in\mathbb{N}} \subseteq T_0$ un subconjunto d_R-denso y contable, cuya existencia está garantizada por la Proposición 1.4.16. Veamos que las trayectorias del proceso $\{X_s\}_{s\in S_0}$ pertenecen a $\mathcal{H}(R, S_0)$ con probabilidad 1.

Para cada $n \in \mathbb{N}$ sea $S_n = \{s_k\}_{1\leq k\leq n}$ y denotemos a las restricciones de K y R a S_n como K_n y R_n respectivamente. En vistas de la Proposición 1.3.5, vale que

$$||X||^2_{S_n} = \sum_{i,j=1}^{n} X_{s_i} X_{s_j} R_n^{-1}(s_i, s_j). \tag{2.8}$$

Definamos $Z_n := ||X||^2_{S_n}$, que es una variable aleatoria por (2.8). La sucesión de variables aleatorias $(Z_n)_{n\in\mathbb{N}}$ es creciente por la Proposición 1.3.6, por lo que está bien definida la variable aleatoria

$$Z := \lim_{n\to\infty} Z_n,$$

donde eventualmente Z podría valer $+\infty$. Sin embargo, por el teorema de convergencia monótona y la Proposición 1.4.20 vale que

$$\begin{aligned}\mathbb{E}[Z] &= \lim_{n\to\infty} \mathbb{E}[Z_n] = \lim_{n\to\infty} \sum_{i,j=1}^{n} \mathbb{E}[X_{s_i} X_{s_j}] R_n^{-1}(s_i, s_j) \\ &= \lim_{n\to\infty} \sum_{i,j=1}^{n} K_n(s_i, s_j) R_n^{-1}(s_i, s_j) = \lim_{n\to\infty} \operatorname{Tr}(K_n R_n^{-1}) \\ &= \operatorname{Tr}(L) < +\infty.\end{aligned}$$

Por lo tanto, $\mathbb{P}(Z < +\infty) = 1$. Si $\Omega' = \{Z < +\infty\}$, para cada $\omega \in \Omega'$ vale que la trayectoria $X_\bullet(\omega)$ del proceso $\{X_s\}_{s\in S_0}$ pertenece a $\mathcal{H}(R, S_0)$ por el Lema 1.4.17, lo cual termina de probar que las trayectorias del proceso $\{X_s\}_{s\in S_0}$ pertenecen a $\mathcal{H}(R, S_0)$ con probabilidad 1.

Como S_0 es un conjunto determinante para $\mathcal{H}(R, T)$ por la Proposición 1.4.16, de la Proposición 1.3.10 se tiene que cada trayectoria $X_\bullet(\omega)$ de $\{X_s\}_{s\in S_0}$ con $\omega \in \Omega'$ se extiende de forma única a una función $f(\omega, \cdot) \in \mathcal{H}(R)$. Definamos

$$\xi(\omega) = \begin{cases} f(\omega, \cdot) & \text{si } \omega \in \Omega' \\ 0 & \text{si no} \end{cases}.$$

Afirmamos que ξ es un elemento aleatorio en $\mathcal{H}(R)$ de segundo orden fuerte y esperanza cero.

Para $s \in S_0$ y $\omega \in \Omega'$ tenemos que $\langle \xi(\omega), R_s \rangle = f(\omega, s) = X_s(\omega)$. Es decir, $\langle \xi, R_s \rangle = X_s$ con probabilidad 1, por lo que $\langle \xi, R_s \rangle$ es una variable aleatoria (recordemos que estamos suponiendo que la medida $\mathbb{P}$ es completa). Luego, por la parte (4) de la Proposición 1.3.10, podemos decir que ξ es medible respecto de $\hat{C}(\mathcal{H}(R))$. Como $\mathcal{H}(R)$ es separable, esto lo mismo que decir que ξ es medible Borel por la Proposición 2.2.3 (esto es necesario para poder decir que ξ es de segundo orden fuerte). Consecuentemente, $||\xi||$ es medible. Además, $||\xi||_{S_n}^2 \overset{a.s.}{=} ||X||_{S_n}^2 = Z_n$. Por el Lema 1.4.17 y el teorema de convergencia monótona llegamos a que

$$\mathbb{E}||\xi||^2 = \mathbb{E}\left[\lim_{n\to\infty} ||\xi||_{S_n}^2\right] = \lim_{n\to\infty} \mathbb{E}||\xi||_{S_n}^2 = \lim_{n\to\infty} \mathbb{E}[Z_n] = \operatorname{Tr}(L) < +\infty.$$

Es decir, ξ es de segundo orden fuerte. En particular, es de primer orden débil y entonces está bien definida $\mathbb{E}\xi \in \mathcal{H}(R)$. Para $s \in S_0$, recordemos que $\langle \xi, R_s \rangle \overset{a.s.}{=} X_s$ y luego $\langle \mathbb{E}\xi, R_s \rangle = \mathbb{E}\langle \xi, R_s \rangle = \mathbb{E}[X_s] = 0$. Esto termina de probar que ξ es un elemento aleatorio en $\mathcal{H}(R)$ de segundo orden fuerte y esperanza cero.

Sea $Y = \{Y_t\}_{t\in T}$ el proceso definido por ξ vía el Lema 2.4.2. Por el Teorema 2.4.3 tenemos que Y es un proceso estocástico de segundo orden, función de media 0 y con trayectorias en $\mathcal{H}(R)$. La prueba termina viendo que para cada $t \in T$ se tiene que $\mathbb{P}(X_t = Y_t) = 1$.

Comencemos probando que $||X_t - Y_t||_{L^2}^2 = 0$ para todo $t \in T_0$. Notemos que si $t \in S_0$, esto vale pues $Y_t = \langle \xi, R_t \rangle \stackrel{a.s.}{=} X_t$. Si $t \in T_0$ es arbitrario, tenemos que para todo $s \in S_0$,

$$||X_t - Y_t||_{L^2} \leq ||X_t - X_s||_{L^2} + \underbrace{||X_s - Y_s||_{L^2}}_{=0} + ||Y_s - Y_t||_{L^2} = ||X_t - X_s||_{L^2} + ||Y_s - Y_t||_{L^2}.$$

Ahora bien, teniendo en cuenta la isometría de Loève (Proposición 2.4.4) y la Observación 1.4.6, tenemos que

$$\begin{aligned} ||X_t - X_s||^2_{L^2} &= ||K_t - K_s||^2_K = ||LR_t - LR_s||^2_K = \langle L(R_t - R_s), L(R_t - R_s)\rangle_K \\ &= \langle R_t - R_s, L(R_t - R_s)\rangle_R \leq ||L||\; ||R_t - R_s||^2_R \\ &= ||L|| d^2_R(t, s). \end{aligned}$$

Por otra parte,

$$||Y_s - Y_t||^2_{L^2} = \mathbb{E}(Y_s - Y_t)^2 = \mathbb{E}\langle \xi, R_s - R_t\rangle^2_R \leq \mathbb{E}||\xi||^2_R ||R_s - R_t||^2_R = \mathrm{Tr}(L) d^2_R(s, t).$$

Juntando ambas cotas, llegamos a que si $t \in T_0$ y $s \in S_0$,

$$||X_t - Y_t||_{L^2} \leq \left(\sqrt{||L||} + \sqrt{\mathrm{Tr}(L)}\right) d_R(s, t).$$

Por la d_R-densidad de S_0 en T_0, el miembro derecho de la desigualdad anterior es arbitrariamente pequeño, por lo que $||X_t - Y_t||_{L^2} = 0$ como queríamos ver.

Por último, si $t \in T$ es arbitrario, se tiene que $R_t = \sum_k \alpha_k R_{t_k}$ para finitos $t_k \in T_0$ y $\alpha_k \in \mathbb{R}$. Aplicando el operador de dominancia, tenemos que $K_t = \sum_k \alpha_k K_{t_k}$ y por la isometría de Loève resulta $X_t \stackrel{a.s.}{=} \sum_k \alpha_k X_{t_k}$. Por otra parte,

$$Y_t = \langle \xi, R_t \rangle_R = \sum_k \alpha_k \langle \xi, R_{t_k}\rangle_R = \sum_k \alpha_k Y_{t_k}$$

y como $X_{t_k} \stackrel{a.s.}{=} Y_{t_k}$ para cada k, concluimos que $X_t \stackrel{a.s.}{=} Y_t$. □

Con la demostración de los Teoremas 2.4.9 y 2.4.11 finaliza la demostración del Teorema 2.4.1, y con ella el capítulo presente.

Capítulo 3

Redes Neuronales y RKBSs

En este capítulo estudiamos problemas de minimización asociados al aprendizaje automático desde el punto de vista de la teoría de núcleos reproductores. Para ello, la noción de RKHS resulta insuficiente y por lo tanto, la ampliamos con el propósito de poder estudiar espacios más generales, como los que estarán asociados a las redes neuronales. Esta generalización nos llevará naturalmente al concepto de espacio de Banach con núcleo reproductor.

En la Sección 3.1 introducimos la noción de espacio de Banach con núcleo reproductor junto con una caracterización alternativa que busca emular la caracterización de RKHSs en términos de núcleos. La Sección 3.2 está dedicada a definir lo que es una red neuronal de una capa oculta y, a partir de esa definición, motivar lo que en este trabajo se llamarán redes neuronales de ancho infinito. En la Sección 3.3 establecemos la conexión entre estos objetos de estudio y, por último, en la Sección 3.4 demostramos los resultados principales del capítulo, generalizando los problemas de optimización típicos del aprendizaje profundo al contexto de redes neuronales de ancho infinito, y probando que, de hecho, éstos tienen solución de ancho finito. Lo sorprendente de este último resultado es que muestra que, aún dentro del espacio de las redes neuronales de ancho infinito (que es un espacio mucho más grande que el de las redes neuronales clásicas), las redes neuronales clásicas siguen brindando soluciones óptimas a los problemas de *machine learning*. Las ideas principales de este capítulo son tomadas de [2, Section 3].

3.1. Espacios de Banach con Núcleo Reproductor

Definición 3.1.1. Sea T un conjunto. Diremos que un subespacio vectorial $\mathcal{B} \subseteq \mathbb{K}^T$ con una norma $||\cdot||$ es un **espacio de Banach con núcleo reproductor** (abreviado **RKBS** por sus siglas en inglés) en T si cumple las siguientes propiedades:

(i) $(\mathcal{B}, ||\cdot||)$ es un espacio de Banach;

(ii) para cada $t \in T$, el funcional de evaluación $E_t : \mathcal{B} \to \mathbb{K}$ dado por $E_t f = f(t)$ resulta acotado.

La definición anterior simplemente toma la definición de RKHS y cambia "Hilbert" por "Banach" en todos los lugares necesarios.

Como $\mathcal{B}$ no tiene por qué ser un espacio de Hilbert, ya no contamos con una caracterización en términos de un núcleo en T, como sucedía en el caso de los RKHSs. Para conseguir una caracterización alternativa, debemos observar lo siguiente: si $\mathcal{H}$ es un RKHS en T con núcleo K, tenemos una función $\phi : T \to \mathcal{H}$ dada por $\phi(t) = K_t$. Vía ϕ, tenemos una forma de que T actúe sobre los elementos de $\mathcal{H}$ por $\langle h, \phi(t)\rangle$. Es decir, podemos pensar que, en realidad, ϕ le asigna a cada $t \in T$ un elemento de $\mathcal{H}'$, el espacio dual de $\mathcal{H}$. En esto vamos a basar la caracterización alternativa de los RKBSs.

Proposición 3.1.2. *Un subespacio lineal $\mathcal{B} \subseteq \mathbb{K}^T$ con una norma $||\cdot||_{\mathcal{B}}$ es un RKBS en T si y sólo si existen un espacio de Banach $\mathcal{F}$ y una función $\phi : T \to \mathcal{F}'$ tales que*

(i) $\mathcal{B} = \{f_\mu : \mu \in \mathcal{F}\}$ *donde* $f_\mu(t) = \phi(t)(\mu)$;

(ii) $||f||_{\mathcal{B}} = \text{ínf}\{||\mu||_{\mathcal{F}} : \mu \in \mathcal{F}, f = f_\mu\} \ \ \forall f \in \mathcal{B}$.

La función $\phi : T \to \mathcal{F}'$ se denomina **función característica**.

Demostración. ($\Longrightarrow$) Tomemos $\mathcal{F} = \mathcal{B}$ y $\phi : T \to \mathcal{B}'$ como $\phi(t) = E_t$. Si $\mu \in \mathcal{F} = \mathcal{B}$, entonces $f_\mu(t) = \phi(t)(\mu) = E_t(\mu) = \mu(t)$. De esto se deducen fácilmente las condiciones (i) y (ii).

($\Longleftarrow$) Fijemos $t \in T$. Dada $f \in \mathcal{B}$, sea $\mu \in \mathcal{F}$ tal que $f = f_\mu$. Entonces

$$|E_t f| = |f(t)| = |f_\mu(t)| = |\phi(t)(\mu)| \leq ||\phi(t)||_{\mathcal{F}'}||\mu||_{\mathcal{F}}.$$

Tomando ínfimo sobre las $\mu \in \mathcal{F}$ tales que $f = f_\mu$ y usando (ii), llegamos a que

$$|E_t f| \leq ||\phi(t)||_{\mathcal{F}'}||f||_{\mathcal{B}},$$

probando así que E_t es acotada con $||E_t|| \leq ||\phi(t)||_{\mathcal{F}'}$.

Para ver que $(\mathcal{B}, ||\cdot||_{\mathcal{B}})$ es un espacio de Banach, definamos $L : \mathcal{F} \to \mathcal{B}$ como $L\mu = f_\mu$. Claramente, L es un operador lineal y de la condición (ii) se deduce que es acotado con $||L|| \leq 1$. Más aún, L es sobreyectivo por la condición (i) y, consecuentemente, induce un isomorfismo $\bar{L} : \mathcal{F}/\ker(L) \to \mathcal{B}$. Pero de la condición (ii) y la definición de la norma cociente obtenemos que $\bar{L}$ es una isometría. Como $\mathcal{F}/\ker(L)$ es un espacio de Banach, concluimos que $\mathcal{B}$ también. $\square$

Para motivar el estudio de cierta clase de RKBSs definidos de forma integral, debemos introducir primero la noción de red neuronal.

3.2. Redes neuronales

Esta sección busca cubrir los conceptos básicos de redes neuronales estrictamente necesarios para el resto del trabajo. Una exposición más completa del tema desde el punto de vista matemático se puede encontrar en [4].

Definición 3.2.1. Una **red neuronal** (con una capa oculta) es una función $f : \mathbb{R}^n \to \mathbb{R}$ de la forma

$$f(x) = \sum_{k=1}^{N} \alpha_k \sigma(w_k \cdot x - b_k), \tag{3.1}$$

con $N \in \mathbb{N},\ \alpha_k, b_k \in \mathbb{R},\ w_k \in \mathbb{R}^n$ y donde $\sigma : \mathbb{R} \to \mathbb{R}$ es una función medible. Decimos que N es el **ancho** de la red y que σ es la **función de activación**.

Algunas de las funciones de activación más utilizadas en la práctica son la *Rectified Linear Unit* (ReLU) y la función sigmoide σ, definidas como

$$\mathrm{ReLU}(x) = \max\{x, 0\} \quad \text{y} \quad \sigma(x) = \frac{1}{1 + e^{-x}}$$

respectivamente.

Con suficientes hipótesis sobre la función de activación, se puede probar que las redes neuronales con una capa oculta pueden aproximar a cualquier función razonable en el sentido adecuado. Es decir, que pueden aproximar uniformemente funciones continuas, o que pueden aproximar funciones integrables en $|| \cdot ||_{L^1}$, etc. (ver [4, Chapter 9]). En general, el costo de aproximar tales funciones se paga con redes de ancho muy grande. Esto nos lleva a preguntarnos lo siguiente: ¿y si pudieran tener ancho infinito?

Para f como en (3.1), notemos $\rho(x, \theta_k) = \sigma(w_k \cdot x - b_k)$, donde $\theta_k = (w_k, b_k) \in \mathbb{R}^{n+1}$. Esto nos permite escribir

$$f(x) = \sum_{k=1}^{N} \alpha_k \rho(x, \theta_k) = \int_{\mathbb{R}^{n+1}} \rho(x, \theta) d\mu(\theta)$$

donde $\mu = \sum_{k=1}^{N} \alpha_k \delta_{\theta_k}$, con δ_{θ_k} la medida de Dirac concentrada en θ_k. De esta forma, podemos pensar que en lugar de pesar a los términos $\rho(x, \theta_k)$ con finitos coeficientes escalares, los pesamos con una medida Borel. Si nos permitimos pensar al ancho de la red como el cardinal del soporte de la medida μ, este enfoque es lo que nos permite precisar qué es una red "de ancho infinito".

3.3. Una familia de RKBSs de redes neuronales

Para esta discusión, fijemos Θ un espacio topológico localmente compacto, Hausdorff y N_2 (es decir, que admite una base contable de abiertos). Este espacio se interpreta como el espacio donde viven los parámetros que determinan a una red neuronal y a fines prácticos, se puede pensar que es algún subconjunto adecuado de $\mathbb{R}^n$. Denotamos $\mathcal{M}(\Theta)$ al espacio de las medidas signadas de Radon definidas sobre la σ-álgebra de Borel de Θ, dotado de la norma de la **variación total**

$$||\mu||_{TV} = |\mu|(\Theta).$$

Por el teorema de representación de Riesz-Markov, sabemos que $\mathcal{M}(\Theta)$ es isométricamente isomorfo al espacio dual de $C_0(\Theta)$, el espacio de funciones continuas $f : \Theta \to \mathbb{R}$ que se anulan en infinito.

Sea T un conjunto y supongamos que tenemos funciones $\rho : T \times \Theta \to \mathbb{R}$ y $\beta : \Theta \to \mathbb{R}$ con β medible tales que

(i) para cada $t \in T$,

$$D_t := \sup_{\theta \in \Theta} |\rho(t, \theta)\beta(\theta)| < +\infty; \tag{3.2}$$

(ii) para cada $t \in T$, la función $\rho(t, \cdot) : \Theta \to \mathbb{R}$ es medible.

Esto nos permite definir un RKBS de redes neuronales de ancho infinito. Concretamente, definimos la función característica $\phi : T \to \mathcal{M}(\Theta)'$ por

$$\phi(t)(\mu) = \int_\Theta \rho(t, \theta)\beta(\theta)d\mu(\theta),$$

cuya buena definición está dada por las hipótesis (i) y (ii). Por la Proposición 3.1.2, ϕ define un RKBS $\mathcal{B}$ en T dado por

$$\mathcal{B} = \{f_\mu : \mu \in \mathcal{M}(\Theta)\};$$

$$f_\mu(t) = \int_\Theta \rho(t, \theta)\beta(\theta)d\mu(\theta);$$

$$||f||_{\mathcal{B}} = \text{ínf}\{||\mu||_{TV} : f = f_\mu\}.$$

Comentario. La condición (3.2) es necesaria y suficiente para garantizar la integrabilidad de $\rho(t, \cdot)\beta$ para todo $t \in T$ respecto de cualquier medida finita boreliana definida en Θ. La inclusión de la función β es para poder garantizar el cumplimiento de esta condición sin asumir hipótesis sobre la función ρ. En su lugar, las condiciones de decaimiento necesarias se le imponen a β. Como ρ representa lo que esencialmente es la función de activación de la red neuronal, para que el enfoque pueda ser aplicado a las clásicas elecciones de funciones de activación (como por ejemplo la ReLU), es importante tener flexibilidad sobre las hipótesis impuestas sobre ella.

3.4. Problemas de minimización

Los problemas de aproximación por redes neuronales se suelen plantear de la siguiente forma: existe una función $f : \mathbb{R}^n \to \mathbb{R}$ de la cual solamente se conocen valores en finitos puntos $\{x_i\}_{i=1}^N$ (dicho de otra forma, se conocen finitas *muestras* de la función f), y a partir de esa información se busca aproximar a f por una red neuronal. Para ello, primero se prefija la arquitectura de la red, que para este trabajo es simplemente el ancho de la red y la función de activación (las ventajas y desventajas de cada arquitectura son algo propio de la disciplina y en este trabajo no se ahondará en ello). Llamemos $M \in \mathbb{N}$ al ancho y σ a la función de activación. Para cada posible elección de parámetros $\theta = (\alpha, w, b)$, se define la función

$$g_\theta(x) = \sum_{k=1}^{M} \alpha_k \sigma(w_k \cdot x - b_k).$$

Para hallar una red neuronal que, dados los datos sobre f, la aproxime lo mejor posible, un primer acercamiento al problema podría ser buscar θ que minimice el error cuadrático medio

$$C(\theta) = \frac{1}{N} \sum_{k=1}^{N} (g_\theta(x_i) - f(x_i))^2.$$

Sin embargo, se ha visto que este enfoque en general produce redes que sobreajustan los datos. Es decir, produce redes que en los puntos x_i aproximan bien a la función f, pero en los demás puntos no. Usualmente, este comportamiento se ve reflejado en que los parámetros θ son grandes y, por lo tanto, resultan en una función excesivamente complicada. Para contrarrestar esto, una de las técnicas estándar en el área es agregar a la función C un término de la forma $\lambda ||\theta||^2$, con $\lambda > 0$ un parámetro que debe ser prefijado, y $|| \cdot ||$ alguna norma en $\mathbb{R}^d$ (en la literatura esto se conoce como un *término regularizante*). Es decir, se buscan parámetros θ que minimicen la función

$$C(\theta) = \frac{1}{N} \sum_{k=1}^{N} (g_\theta(x_i) - f(x_i))^2 + \lambda ||\theta||^2.$$

Los siguientes resultados buscan probar resultados de existencia para los problemas de minimización análogos que resultan de estudiar redes neuronales de ancho infinito. Manteniendo la notación de la sección anterior, el objetivo principal de esta sección es probar el siguiente resultado, mencionado al principio del capítulo presente.

Teorema 3.4.1. *Supongamos que para cada $t \in T$ se tiene que $\rho(t, \cdot)\beta \in C_0(\Theta)$. Si $L(\cdot, y)$ es coerciva para cada $y \in \mathbb{R}$, entonces existen minimizantes de*

$$\inf_{f \in \mathcal{B}} \frac{1}{N} \sum_{i=1}^{N} L(f(t_i), y_i) + ||f||_{\mathcal{B}} \tag{3.3}$$

de la forma

$$f^*(t) = \sum_{k=0}^{K} \alpha_k \rho(t, \theta_k)$$

con $K \leq N$ *y*

$$||f^*||_{\mathcal{B}} = \sum_{k=0}^{K} |\alpha_k \beta(\theta_k)^{-1}|.$$

El teorema anterior dice, esencialmente, que bajo ciertas hipótesis sobre el decaimiento de la función β, el problema (3.3) admite minimizantes, y de hecho admite minimizantes de ancho finito. Más aún, el ancho de los minimizantes está controlado por la cantidad de puntos de muestreo. Lo sorprendente del resultado es que muestra que, aún dentro del espacio de las redes neuronales de ancho infinito (que a priori es mucho más extenso que el espacio de redes neuronales clásicas), las redes neuronales clásicas brindan soluciones óptimas.

Lo primero que haremos en pos de demostrar el Teorema 3.4.1 será ver que estos problemas de minimización son equivalentes a problemas de minimización sobre el espacio de las medidas. La ventaja de este enfoque es que en el espacio de las medidas contamos con la topología débil*, respecto de la cual la bola cerrada es compacta. Mantenemos la notación de la sección anterior a lo largo de esta sección.

Proposición 3.4.2. *Sean* $\rho : T \times \Theta \to \mathbb{R}$*,* $\beta : \Theta \to \mathbb{R}$ *y* $\mathcal{B}$ *como antes. Si* $L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ *es cualquier función y* $(t_i, y_i)_{i=1}^N \subseteq T \times \mathbb{R}$ *son* N *pares de puntos, entonces*

$$\inf_{f \in \mathcal{B}} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f(t_i)) + ||f||_{\mathcal{B}} \right) = \inf_{\mu \in \mathcal{M}(\Theta)} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + ||\mu||_{TV} \right).$$

Más aún, si $\mu^* \in \mathcal{M}(\Theta)$ *es un minimizante de*

$$\inf_{\mu \in \mathcal{M}(\Theta)} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + ||\mu||_{TV} \right), \tag{3.4}$$

entonces $f^* = f_{\mu^*}$ *es un minimizante de*

$$\inf_{f \in \mathcal{B}} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f(t_i)) + ||f||_{\mathcal{B}} \right) \tag{3.5}$$

con $||f^*||_{\mathcal{B}} = ||\mu^*||_{TV}$.

Demostración. Por definición de $\mathcal{B}$ tenemos que

$$\begin{aligned}
\inf_{f \in \mathcal{B}} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f(t_i)) + ||f||_{\mathcal{B}} \right) &= \inf_{\mu \in \mathcal{M}(\Theta)} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + ||f_\mu||_{\mathcal{B}} \right) \\
&= \inf_{\mu \in \mathcal{M}(\Theta)} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + \inf_{f_\nu = f_\mu} ||\nu||_{TV} \right) \\
&= \inf_{\substack{\mu, \nu \in \mathcal{M}(\Theta) \\ f_\mu = f_\nu}} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + ||\nu||_{TV} \right) \\
&= \inf_{\nu \in \mathcal{M}(\Theta)} \left(\frac{1}{N} \sum_{i=1}^{N} L(y_i, f_\nu(t_i)) + ||\nu||_{TV} \right).
\end{aligned}$$

Ahora supongamos que μ^* es un minimizante de (3.4). Luego, para toda $\nu \in \mathcal{M}(\Theta)$ vale que

$$\frac{1}{N}\sum_{i=1}^{N} L(y_i, f^*(t_i)) + ||\mu^*||_{TV} \leq \frac{1}{N}\sum_{i=1}^{N} L(y_i, f_\nu(t_i)) + ||\nu||_{TV}.$$

Fijemos $\mu \in \mathcal{M}(\Theta)$. Tomando ínfimo sobre todas las $\nu \in \mathcal{M}(\Theta)$ tales que $f_\mu = f_\nu$ obtenemos que

$$\frac{1}{N}\sum_{i=1}^{N} L(y_i, f^*(t_i)) + ||\mu^*||_{TV} \leq \frac{1}{N}\sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + ||f_\mu||_{\mathcal{B}}.$$

Sustituyendo en la igualdad anterior $\mu = \mu^*$ llegamos a que $||\mu^*||_{TV} \leq ||f^*||_{\mathcal{B}}$ y como por definición $||f^*||_{\mathcal{B}} \leq ||\mu^*||_{TV}$, concluimos que $||f^*||_{\mathcal{B}} = ||\mu^*||_{TV}$. Consecuentemente,

$$\frac{1}{N}\sum_{i=1}^{N} L(y_i, f^*(t_i)) + ||f^*||_{\mathcal{B}} \leq \frac{1}{N}\sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + ||f_\mu||_{\mathcal{B}}$$

para toda $\mu \in \mathcal{M}(\Theta)$. □

El siguiente corolario es un análogo de un resultado clásico del contexto hilbertiano, comúnmente llamado *Representer Theorem*, muy utilizado en ciertos problemas de regresión [13, Theorem 8.7].

Corolario 3.4.3. *En las condiciones de antes, sea*

$$\mathcal{V} = \{\mu \in \mathcal{M}(\Theta) : \ f_\mu(t_i) = 0 \ \ \textit{para} \ \ 1 \leq i \leq N\} = \{\rho(t_i, \cdot)\beta(\cdot)\}^{\perp}_{1\leq i\leq N},$$

donde el anulador $^\perp$ *se toma respecto del emparejamiento* ${}_{\mathcal{M}(\Theta)'}\langle \cdot, \cdot \rangle_{\mathcal{M}(\Theta)}$. *Entonces existe un subespacio* $\mathcal{W} \subseteq \mathcal{M}(\Theta)$ *con* $\dim \mathcal{W} \leq N$ *tal que* $\mathcal{M}(\Theta) = \mathcal{V} \oplus \mathcal{W}$ *y*

$$\inf_{\mu \in \mathcal{M}(\Theta)} \left(\frac{1}{N}\sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + ||\mu||_{TV} \right) = \inf_{\nu \in \mathcal{W}} \left(\frac{1}{N}\sum_{i=1}^{N} L(y_i, f_\nu(t_i)) + \inf_{\tau \in \mathcal{V}} ||\nu + \tau||_{TV} \right).$$

Demostración. Por definición de la norma en $\mathcal{B}$, sabemos que la asignación $\mu \mapsto f_\mu$ es continua. A su vez, por ser $\mathcal{B}$ un RKBS en T, tenemos que la asignación $f_\mu \mapsto f_\mu(t)$ es continua para todo $t \in T$. Por lo tanto, para cada $t \in T$ la asignación $\mu \mapsto f_\mu(t)$ es un funcional lineal de $\mathcal{M}(\Theta)$. En consecuencia, $\mathcal{V}$ es un subespacio cerrado de $\mathcal{M}(\Theta)$ de codimensión a lo sumo N, por ser intersección de N hiperplanos cerrados. Como los subespacios de codimensión finita siempre son complementados, existe $\mathcal{W} \subseteq \mathcal{M}(\Theta)$ subespacio de dimensión a lo sumo N tal que $\mathcal{M}(\Theta) = \mathcal{V} \oplus \mathcal{W}$.

Llamemos $F : \mathcal{M}(\Theta) \to \mathbb{R}$ a la función

$$F(\mu) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f_\mu(t_i)) + ||\mu||_{TV}.$$

Luego, si $\mu = \nu + \tau$ con $\nu \in \mathcal{W}$ y $\tau \in \mathcal{V}$, entonces

$$F(\mu) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f_\nu(t_i)) + ||\nu + \tau||_{TV},$$

de donde se desprende el resto del enunciado. $\square$

A continuación, probaremos algunos resultados técnicos sobre minimización en espacios localmente convexos, que serán necesarios para la demostración del Teorema 3.4.1. Las versiones generales de estos resultados pueden hallarse en [3]. Nosotros daremos versiones simplificadas, pero que son suficientes para nuestros propósitos.

Primero, recordemos lo siguiente: si X es un espacio topológico, una función $f : X \to \mathbb{R}$ se dice **semicontinua inferiormente** si cumple alguna de las siguientes condiciones equivalentes:

(i) para cada $a \in \mathbb{R}$, el conjunto $f^{-1}((-\infty, a])$ es cerrado;

(ii) para toda red $(x_\lambda)_{\lambda\in\Lambda}$ en X que converge a algún $x \in X$, vale que

$$\liminf f(x_\lambda) := \sup_{\lambda\in\Lambda} \inf_{\alpha\geq\lambda} f(x_\alpha) \geq f(x).$$

Definición 3.4.4. Si X es un espacio localmente convexo, decimos que una función $f : X \to \mathbb{R}$ es **coerciva** si para todo $a \in \mathbb{R}$ el conjunto de subnivel

$$S^-(f, a) := \{x \in X : f(x) \leq a\}$$

es compacto. Notar que si f es coerciva, en particular todos los conjuntos de subnivel son cerrados y, luego, f es semicontinua inferiormente.

Proposición 3.4.5. *Sean X un espacio localmente convexo y $f, g : X \to \mathbb{R}$ funciones.*

(1) Si f es coerciva, entonces f es acotada inferiormente.

(2) Si f es coerciva y g es semicontinua inferiormente con $\inf g > -\infty$ *, entonces $f+g$ es coerciva. En particular, suma de funciones coercivas es coerciva.*

Demostración. (1) La familia $\{S^-(f,a)\}_{a\in\mathbb{R}}$ es una familia no vacía de subconjuntos compactos de $\mathcal{H}$ cuya intersección es vacía. Por compacidad, deben existir finitos miembros de la familia cuya intersección es vacía. Pero como la familia es totalmente ordenada respecto de la inclusión, esto quiere decir que existe un elemento de la familia que es vacío. Es decir, $S^-(f,a) = \{h \in \mathcal{H} : F(h) \leq a\} = \varnothing$ para cierto $a \in \mathbb{R}$, de donde se deduce que $f > a$.

(2) Como f y g son semicontinuas inferiormente, $f + g$ también lo es. Además, tenemos que $S^-(f+g, a) \subseteq S^-(f, a - \inf g)$. Como $S^-(f, a - \inf g)$ es compacto por ser f coerciva, y $S^-(f + g, a)$ es cerrado por ser $f + g$ semicontinua inferiormente, esto prueba que $S^-(f + g, a)$ es compacto. $\square$

Antes de probar el siguiente lema, recordemos la definición de punto extremal.

Definición 3.4.6. Sea X un espacio vectorial, $\varnothing \neq A \subseteq X$. Decimos que un punto $x \in A$ es un **punto extremal** de A si vale la siguiente implicación para cualesquiera $y, z \in A$ y $0 < t < 1$:

$$ty + (1-t)z = x \implies x = y = z.$$

Denotamos por $\mathrm{Ext}(A)$ al conjunto de puntos extremales de A.

Lema 3.4.7. *Sean X un espacio localmente convexo y $K \subseteq X$ un conjunto convexo. Sea Y un espacio vectorial topológico real y $L : X \to Y$ una aplicación lineal.*

(i) Si L es continua y K es compacto, entonces $Ext(L(K)) \subseteq L(Ext(K))$.

(ii) Si L es inyectiva entonces $Ext(L(K)) = L(Ext(K))$.

Demostración. (i) Sea $k \in K$ tal que Lk es un punto extremal de $L(K)$. Queremos ver que existe $\bar{k} \in \mathrm{Ext}(K)$ tal que $Lk = L\bar{k}$. Como L es continua, el conjunto $k + \ker L$ es cerrado. Al ser K compacto, se tiene que el conjunto $(k + \ker L) \cap K$ es compacto y no vacío. Por el teorema de Krein-Milman (Teorema A.3.3), admite algún punto extremal $\bar{k}$. Como $Lk = L\bar{k}$, sólo resta ver que $\bar{k}$ es un punto extremal de K. Supongamos entonces que $\bar{k} = tk_1 + (1-t)k_2$ con $0 < t < 1$, $k_1, k_2 \in K$. Luego,

$$Lk = L\bar{k} = tLk_1 + (1-t)Lk_2,$$

y por extremalidad de Lk tenemos que $Lk_1 = Lk_2 = Lk$. Por lo tanto, k_1 y k_2 pertenecen a $(k + \ker L) \cap K$, y de la extremalidad de $\bar{k}$ en este conjunto concluimos que $k_1 = k_2 = \bar{k}$.

(ii) Sea $k \in K$ tal que Lk es un punto extremal de $L(K)$. Veamos que k es extremal en K. Supongamos que $k = tk_1 + (1-t)k_2$ con $0 < t < 1$, $k_1, k_2 \in K$. Luego,

$$Lk = tLk_1 + (1-t)Lk_2,$$

y por extremalidad de Lk tenemos que $Lk_1 = Lk_2 = Lk$. Como L es inyectiva, concluimos que $k = k_1 = k_2$.

Ahora sea $k \in \mathrm{Ext}(K)$. Veamos que $Lk \in \mathrm{Ext}(L(K))$. Si $Lk = tLk_1 + (1-t)Lk_2$ con $0 < t < 1$, $k_1, k_2 \in K$, entonces $Lk = L(tk_1 + (1-t)k_2)$ y por inyectividad de L tenemos que $k = tk_1 + (1-t)k_2$. Como k es extremal en K, concluimos que $k = k_1 = k_2$ y en consecuencia $Lk = Lk_1 = Lk_2$. □

La siguiente proposición es la mayor parte de la demostración del Teorema 3.4.1. Una vez probada, lo único que resta hacer es ver que el problema (3.3) es una instancia de esta proposición.

Proposición 3.4.8. *Sea X un espacio de Banach, y consideremos en su espacio dual X' la topología débil*. Supongamos que $F : \mathbb{R}^N \to \mathbb{R}$ es una función coerciva, y que $A : X' \to \mathbb{R}^N$ es un operador lineal y continuo. Entonces existen minimizantes de*

$$\operatorname*{ínf}_{x' \in X'} F(Ax') + ||x'||_{X'} \tag{3.6}$$

de la forma $x^* = \sum_{i=0}^{p} \gamma_i x'_i$ con $p \leq N$, $\gamma_i > 0$, $\sum_i \gamma_i = ||x^*||_{X'}$ y $x'_i \in Ext(\overline{B}_{X'})$.

Demostración. Llamemos $J(x') = F(Ax') + ||x'||_{X'}$. Como F es coerciva, por la parte (1) de la Proposición 3.4.5 vale que F es acotada inferiormente. Ahora bien, como F es semicontinua inferiormente (por ser coerciva) y A es continuo, entonces $F \circ A$ es semicontinua inferiormente. Por otra parte, $|| \cdot ||_{X'}$ es coerciva respecto de la topología débil* de X' por el teorema de Banach-Alaoglu. Por la parte (2) de la Proposición 3.4.5, J es coerciva. En particular, es acotada inferiormente.

Con todo esto, tenemos que $\{S^-(J, a)\}_{a > \text{ínf } J}$ es una familia de subconjuntos compactos de X' con la propiedad de intersección finita. Luego $\bigcap_{a > \text{ínf } J} S^-(J, a) \neq \varnothing$ y cualquier elemento de la intersección es un minimizante de (3.6). Tomemos z' algún tal minimizante. Como en el caso que $z' = 0$ no hay nada que hacer, supongamos que $z' \neq 0$.

Al ser $Az'/||z'||_{X'}$ un elemento de $A(\overline{B}_{X'})$, que es compacto y convexo, por el teorema de Minkowski (Teorema A.4.4) y el teorema de Carathéodory (Teorema A.2.2) tenemos que podemos escribir

$$\frac{1}{||z'||_{X'}} Az' = \sum_{i=0}^{p} t_i w_i$$

con $p \leq N$, $\sum_i t_i = 1$, $t_i > 0$ y $w_i \in \text{Ext}(A(\overline{B}_{X'}))$. Por la parte (i) del Lema 3.4.7, cada w_i es de la forma Ax'_i con $x'_i \in \text{Ext}(\overline{B}_{X'})$. Llamemos $\gamma_i = ||z'|| t_i$ y $x^* = \sum_{i=0}^{p} \gamma_i x'_i$. Afirmamos que x^* es un minimizante de (3.6).

Primero, notemos que

$$Ax^* = A\left(\sum_{i=0}^{p} \gamma_i x'_i\right) = \sum_{i=0}^{p} \gamma_i Ax'_i = ||z'||_{X'} \sum_{i=0}^{p} t_i w_i = Az'.$$

Por otra parte,

$$||x^*||_{X'} \leq \sum_{i=0}^{p} \gamma_i \underbrace{||x'_i||_{X'}}_{\leq 1} \leq \sum_{i=0}^{p} \gamma_i = ||z'||_{X'}.$$

Juntando ambas cosas, podemos decir que

$$J(x^*) = F(\underbrace{Ax^*}_{=Az'}) + \underbrace{||x^*||_{X'}}_{\leq ||z'||_{X'}} \leq F(Az') + ||z'||_{X'} = J(z')$$

y como z' era un minimizante de J, esto prueba que x^* es un minimizante de (3.6). Además, como $F(Ax^*) + ||x^*||_{X'} = J(x^*) = J(z') = F(Az') + ||z'||_{X'}$ y $F(Ax^*) = F(Az')$, entonces $||x^*||_{X'} = ||z'||_{X'}$ y consecuentemente $\sum_i \gamma_i = ||z||_{X'} = ||x^*||_{X'}$, lo cual prueba que x^* tiene la forma del enunciado. $\square$

Demostración del Teorema 3.4.1. Por la Proposición 3.4.2, basta considerar el problema (3.3) sobe el espacio de medidas $\mathcal{M}(\Theta)$. Aplicamos la Proposición 3.4.8. Consideramos $X = C_0(\Theta)$, e identificamos $X' = \mathcal{M}(\Theta)$ vía el teorema de representación de Riesz-Markov. Sean $\mathcal{H} = \mathbb{R}^N$, $A : X \to \mathbb{R}^N$ dado por $(A\mu)_i = f_\mu(t_i)$ y $F : \mathcal{H} \to \mathbb{R}$ definida por $F(w) = \frac{1}{N}\sum_{i=1}^N L(w_i, y_i)$.

Notemos que F es coerciva por ser suma de funciones coercivas (parte (2) de la Proposición 3.4.5). Por otra parte, para cada $1 \leq i \leq N$ tenemos que la aplicación $\mu \mapsto f_\mu(t_i)$ es débil* continua, ya que

$$f_\mu(t_i) = \int_\Theta \rho(t_i, \theta)\beta(\theta)d\mu(\theta) = {}_{\mathcal{M}(\Theta)}\langle \mu, \rho(t_i, \cdot)\beta\rangle_{C_0(\Theta)},$$

lo cual prueba que el operador A es continuo.

Por la Proposición 3.4.8, existe $\mu^* \in \mathcal{M}(\Theta)$ minimizante de (3.3) de la forma

$$\mu^* = \sum_{k=0}^K \gamma_k \nu_k$$

con $K \leq N$, $\gamma_k > 0$, $\sum_k \gamma_k = ||\mu^*||_{TV}$ y $\nu_k \in \operatorname{Ext}(\overline{B}_{\mathcal{M}(\Theta)})$. Pero por el Teorema A.1.3 tenemos que $\nu_k = \pm\delta_{\theta_k}$ para algún $\theta_k \in \Theta$.

Ahora bien, por la Proposición 3.4.2 sabemos que $f^* = f_{\mu^*}$ es un minimizante de (3.3) con $||f^*||_{\mathcal{B}} = ||\mu^*||_{TV}$. En consecuencia, tenemos un minimizante

$$f^* = \sum_{k=0}^K \pm\gamma_k\beta(\theta_k)\rho(\cdot, \theta_k)$$

que es de la forma del enunciado. Por último, llamando $\alpha_k = \pm\gamma_k\beta(\theta_k)$ tenemos que

$$||f^*||_{\mathcal{B}} = ||\mu^*||_{TV} = \sum_{k=1}^K \gamma_k = \sum_{k=1}^K |\alpha_k\beta(\theta_k)^{-1}|,$$

lo cual termina la prueba. $\square$

Apéndice A

Puntos extremales y convexidad

En este apéndice damos algunos resultados clásicos sobre puntos extremales que fueron utilizados en el Capítulo 3, particularmente en la Sección 3.4. Incluimos sus demostraciones para comodidad del lector.

A.1. Puntos extremales en espacios de Banach

Comencemos recordando las distintas nociones de extremalidad.

Definición A.1.1. Sea X un espacio vectorial, y sean $\varnothing \neq A \subseteq B \subseteq X$. Decimos que A es **extremal** en B si vale la siguiente implicación para cualesquiera $b, b' \in B$ y $0 < t < 1$:

$$tb + (1-t)b' \in A \implies b, b' \in A.$$

En este caso, notamos A ext B. Decimos que $b \in B$ es un **punto extremal** de B si vale que $\{b\}$ ext B (observemos que esto es consistente con la Definición 3.4.6). Al igual que en el Capítulo 3, denotamos por $\text{Ext}(B)$ al conjunto de puntos extremales de B.

Para este trabajo es de particular interés entender los puntos extremales de la bola cerrada de un espacio de Banach. Un primer resultado en esa dirección es el siguiente.

Lema A.1.2. *Sea $X \neq 0$ un espacio de Banach. Entonces $Ext(\overline{B}_X) = Ext(S_X)$.*

Demostración.

($\subseteq$) Sea $x \in \text{Ext}(\overline{B}_X)$. Para ver que x es un punto extremal de S_X, basta ver que $||x|| = 1$. Luego, veamos que si $||x|| < 1$, entonces x no es extremal en $\overline{B}_X$.

Si $x = 0$, para $v \in X$ con $||v|| = 1$ se llega a que $x = \frac{1}{2}v + \frac{1}{2}(-v)$ de donde x no es extremal en $\overline{B}_X$. Si $x \neq 0$, tomando $t = ||x||$ vale que $x = (1-t)0 + t\frac{x}{||x||}$ por lo que x tampoco es extremal en $\overline{B}_X$.

($\supseteq$) Supongamos que $x \in \text{Ext}(S_X)$ y sean $y, z \in \overline{B}_X$, $0 < t < 1$ tales que $x = ty + (1-t)z$. Luego

$$1 = ||x|| \leq t \underbrace{||y||}_{\leq 1} + (1-t) \underbrace{||z||}_{\leq 1} \leq 1,$$

de donde deducimos que $||y|| = ||z|| = 1$ y de la extremalidad de x en S_X se obtiene que $y = z = x$. □

En general, no es fácil caracterizar los puntos extremales de un conjunto. Sin embargo, hallar esas caracterizaciones es tremendamente útil en muchas aplicaciones. Una de ellas (que es lo que hicimos en este trabajo) es atacar problemas de optimización convexa. En lo que queda de esta sección, buscamos dar una caracterización de los puntos extremales de la bola cerrada del espacio de las medidas. Para ello, primero debemos recordar una noción general de teoría de la medida.

Si $(E, \mathcal{A})$ es un espacio medible y μ es una medida signada en E, una **descomposición de Hahn** para μ es un par (P, N) tal que:

(i) $P, N \in \mathcal{A}$;

(ii) $P \cup N = E$ y $P \cap N = \varnothing$;

(iii) para todo $A \subseteq P$ medible vale que $\mu(A) \geq 0$;

(iv) para todo $B \subseteq N$ medible vale que $\mu(B) \leq 0$.

Un resultado estándar en teoría de la medida dice que toda medida signada admite una descomposición de Hahn [16, Theorem 10.36]. Además vale, aunque no es necesario para este trabajo, que son únicas salvo conjuntos de medida 0. Es decir, si $(P, N), (P', N')$ son dos descomposiciones de Hahn de una medida signada μ, entonces $|\mu|(P \triangle P') = |\mu|(N \triangle N') = 0$, donde $\triangle$ denota a la diferencia simétrica de conjuntos. Es claro de las definiciones que $\mu = \mu(\cdot \cap P) + \mu(\cdot \cap N)$, y también vale que $|\mu| = \mu(\cdot \cap P) - \mu(\cdot \cap N)$.

A continuación, damos una caracterización de los puntos extremales de la bola cerrada del espacio de las medidas, que fue utilizada en la demostración del Teorema 3.4.1.

Teorema A.1.3. *Sea $\Theta \neq \varnothing$ un espacio topológico localmente compacto y Hausdorff con base contable de abiertos. Entonces Ext$(\overline{B}_{\mathcal{M}(\Theta)}) = \{\pm\delta_\theta : \theta \in \Theta\}$.*

Demostración. Denotemos $||\cdot|| = ||\cdot||_{\mathcal{M}(\Theta)}$.

($\supseteq$) Tomemos $\theta \in \Theta$ y $a = \pm 1$. Para ver que $a\delta_\theta$ es un punto extremal de $\overline{B}_{\mathcal{M}(\Theta)}$, por el Lema A.1.2 basta ver que es extremal en la esfera $S_{\mathcal{M}(\Theta)}$.

Supongamos que $\mu, \nu \in S_{\mathcal{M}(\Theta)}$ y $t \in (0,1)$ son tales que $t\mu + (1-t)\nu = a\delta_\theta$. Veamos que $|\mu| = |\nu| = \delta_\theta$.

Observemos que $(t|\mu| + (1-t)|\nu|)(\Theta) = 1$. Si $E \subseteq \Theta$ es un conjunto boreliano y $\theta \in E$ entonces

$$1 = \delta_\theta(E) \leq (t|\mu| + (1-t)|\nu|)(E) \leq (t|\mu| + (1-t)|\nu|)(\Theta) = 1,$$

de donde $(t|\mu| + (1-t)|\nu|)(E) = t|\mu|(E) + (1-t)|\nu|(E) = 1$. Como $|\mu|(E) \leq |\mu|(\Theta) = 1$ y lo mismo para $|\nu|$, necesariamente $|\mu|(E) = |\nu|(E) = 1$.

Por otra parte, si $\theta \notin E$, entonces $\theta \in \Theta \setminus E$. Por lo recién argumentado, $|\mu|(\Theta \setminus E) = 1$ y consecuentemente $|\mu|(E) = |\mu|(\Theta) - |\mu|(\Theta \setminus E) = 1 - 1 = 0$. Por el mismo motivo, $|\nu|(E) = 0$. Esto termina de probar que $|\mu| = |\nu| = \delta_\theta$.

Veamos ahora que $\mu = \pm\delta_\theta$. Sea (P, N) una descomposición de Hahn para μ. Luego para cada conjunto boreliano $E \subseteq \Theta$ tenemos que

$$\mu(E) = \mu(E \cap P) + \mu(E \cap N).$$

Si $\theta \in P$, entonces $\mu(E \cap N) = 0$ para cualquier boreliano E y consecuentemente

$$\mu(E) = \mu(E \cap P) = \mu(E \cap P) - \mu(E \cap N) = |\mu|(E) = \delta_\theta(E).$$

Es decir, $\mu = \delta_\theta$. Similarmente, si $\theta \in N$ obtenemos que $\mu = -\delta_\theta$. Del mismo modo, $\nu = \pm\delta_\theta$.

Finalmente, veamos que $\mu = \nu = a\delta_\theta$. Cambiando μ y ν por $-\mu$ y $-\nu$ si fuera necesario, podemos suponer que $a = 1$. Supongamos que no es el caso y sin pérdida de generalidad, supongamos que $\mu = -\delta_\theta$. Entonces

$$1 = \delta_\theta(\Theta) = t\mu(\Theta) + (1-t)\nu(\Theta) = -t \pm (1-t) = \begin{cases} -1 & \text{si } \nu = -\delta_\theta \\ 1 - 2t & \text{si } \nu = \delta_\theta \end{cases},$$

que en ambos casos es absurdo. Esto termina de probar que $\mu = \nu = a\delta_\theta$ y, en consecuencia, $a\delta_\theta$ es un punto extremal de $\overline{B}_{\mathcal{M}(\Theta)}$.

($\subseteq$) Sea $\mu \in \mathrm{Ext}(\overline{B}_{\mathcal{M}(\Theta)})$ y supongamos que $\mu \neq \pm\delta_\theta$ para todo $\theta \in \Theta$. Por el Lema A.1.2 tenemos que $||\mu|| = 1$. Notemos que si $E \subseteq \Theta$ es un boreliano tal que $0 < |\mu|(E) < 1$, podemos escribir

$$\mu = |\mu|(E)\frac{\mu(E \cap \cdot)}{|\mu(E)|} + (1 - |\mu|(E))\frac{\mu(\cdot \setminus E)}{1 - |\mu|(E)}.$$

Luego para llegar a un absurdo, comenzamos buscando un conjunto boreliano $E \subseteq \Theta$ tal que $|\mu|(E) \in (0, 1)$.

Supongamos que no existe tal conjunto. Es decir, que para todo boreliano E vale que $|\mu|(E) \in \{0, 1\}$. Afirmamos que $\mathrm{sop}(|\mu|) := \{x \in \Theta \mid \forall\, U \ni x \text{ abierto} : |\mu|(U) > 0\} \neq \varnothing$. En efecto, si $\mathrm{sop}(|\mu|) = \varnothing$, entonces para cada $x \in \Theta$ existe un entorno abierto U_x de x tal que $|\mu|(U_x) = 0$. Como $\{U_x\}_{x\in\Theta}$ es un cubrimiento de Θ por abiertos y Θ admite una base contable de abiertos, existe un subcubrimiento $\{U_x\}_{x\in J}$ contable. Pero entonces

$$1 = |\mu|(\Theta) = |\mu|\left(\bigcup_{x\in J} U_x\right) \leq \sum_{x\in J} \underbrace{|\mu|(U_x)}_{=0} = 0$$

lo cual es un absurdo. Luego $\mathrm{sop}(|\mu|) \neq \varnothing$. Tomemos $\theta \in \mathrm{sop}(|\mu|)$. Si U es un entorno abierto de θ, por definición de $\mathrm{sop}(|\mu|)$ vale que $|\mu|(U) > 0$ y en consecuencia $|\mu|(U) = 1$ ya que estamos suponiendo que $|\mu|$ sólo toma los valores 0 y 1. Por regularidad exterior, tenemos que

$$|\mu|(\{\theta\}) = \inf_{\substack{\theta \in U \\ U \text{ abierto}}} |\mu|(U) = 1.$$

Con esto, si E es un boreliano tal que $\theta \in E$, entonces

$$1 = |\mu|(\{\theta\}) \leq |\mu|(E) \leq 1.$$

Es decir, $|\mu|(E) = 1$. Por otra parte, si $\theta \notin E$, entonces $\theta \in \Theta \setminus E$. Por lo recién dicho, esto implica que $|\mu|(\Theta \setminus E) = 1$ y en consecuencia $|\mu|(E) = 0$. Esto prueba que $|\mu| = \delta_\theta$ y, argumentando como antes, podemos concluir que $\mu = \pm\delta_\theta$, contradiciendo nuestra suposición inicial. El absurdo provino de suponer que $|\mu|$ sólo asumía los valores 0 y 1. Luego, debe existir un boreliano E tal que $|\mu|(E) \in (0, 1)$.

Resta ver que, cambiando adecuadamente E, vale que $\mu \neq (|\mu|(E))^{-1}\mu(E \cap \cdot)$. Sea (P, N) una descomposición de Hahn para μ. Luego,

$$0 < |\mu|(E) = \mu(E \cap P) - \mu(E \cap N).$$

Esto implica que, o bien $\mu(E \cap P) > 0$, o bien $-\mu(E \cap N) > 0$. Como ambos casos son análogos, probemos solamente el caso en que $\mu(E \cap P) > 0$. Al ser

$$0 < \mu(E \cap P) \leq |\mu|(E) < 1,$$

cambiando E por $E \cap P$ podemos suponer que $E \subseteq P$ y en consecuencia $\mu(E) = |\mu|(E)$. Pero entonces tenemos por un lado que $\mu(E) = |\mu|(E) < 1$, y por otro lado

$$\frac{\mu(E \cap E)}{|\mu|(E)} = \frac{\mu(E)}{\mu(E)} = 1,$$

terminando así de probar que $\mu \neq (|\mu|(E))^{-1}\mu(E \cap \cdot)$. $\square$

A.2. El teorema de Carathéodory

Definición A.2.1. Sea X un espacio vectorial, $E \subseteq X$ un subconjunto cualquiera. Definimos la **cápsula convexa** de E (denotada $\mathrm{co}(E)$) como

$$\mathrm{co}(E) = \bigcap_{\substack{C \supseteq E \\ C \text{ convexo}}} C.$$

Es decir, $\mathrm{co}(E)$ es el menor convexo de X que contiene a E. Equivalentemente, $\mathrm{co}(E)$ es el conjunto de todas las posibles combinaciones convexas de elementos de E.

Si X es además un espacio vectorial topológico, definimos la **cápsula convexa cerrada** de E como $\overline{\mathrm{co}(E)}$, y la denotamos $\overline{\mathrm{co}}(E)$. Equivalentemente, $\overline{\mathrm{co}}(E)$ es el menor convexo cerrado de X que contiene a E.

Denotamos

$$\Delta_n := \left\{ t \in \mathbb{R}^{n+1} : t_i \geq 0 \, \forall i, \, \sum_i t_i = 1 \right\} = \operatorname{co}(\{e_0, \dots, e_n\}),$$

donde e_j denota al j−ésimo vector canónico de $\mathbb{R}^{n+1}$. Luego, decir que x es combinación convexa de $x_0, \dots, x_n$ es equivalente a decir que existe $t \in \Delta_n$ tal que $x = \sum_i t_i x_i$.

Teorema A.2.2 (Carathéodory)**.** *Sea $E \subseteq \mathbb{R}^n$ y sea $x \in co(E)$. Entonces existen $s \in \Delta_n$ y $x_0, \dots, x_n \in E$ tales que $x = \sum_i s_i x_i$.*

Demostración. Como $x \in \operatorname{co}(E)$, existen $x_0, \dots, x_k \in E$ y $t \in \Delta_k$ tales que $x = \sum_{i=0}^k t_i x_i$. Probamos el resultado por inducción en k. Si $k \leq n$ no hay nada que hacer. Supongamos ahora que $k > n$. Si algún $t_i = 0$ entonces el resultado vale por hipótesis inductiva. Supongamos entonces además que $t_i > 0$ para todo $0 \leq i \leq k$.

Como $k > n$, el mapa lineal $\mathbb{R}^{k+1} \to \mathbb{R}^{n+1}$ dado por

$$a \mapsto \left(\sum_i a_i x_i, \sum_i a_i \right)$$

tiene núcleo no trivial. Entonces existe $a \in \mathbb{R}^{k+1} \setminus \{0\}$ tal que $\sum_i a_i x_i = 0$ y $\sum_i a_i = 0$. Llamemos

$$\mu = \max_{0 \leq i \leq k} \frac{|a_i|}{t_i},$$

que es positivo pues $a \neq 0$. Si $0 \leq j \leq k$ es donde se realiza el máximo, podemos suponer que $a_j > 0$, eventualmente cambiando a por $-a$. Sea

$$\lambda = \frac{1}{\mu} = \frac{t_j}{a_j}.$$

Entonces $\lambda a_j = t_j$ y $|\lambda a_i| \leq t_i$ para cada $0 \leq i \leq k$.

Tomamos $t_i' = t_i - \lambda a_i$. Notemos que $t_i' = t_i - \lambda a_i \geq t_i - |\lambda a_i| \geq 0$ y además

$$\sum_i t_i' = \underbrace{\sum_i t_i}_{=1} - \lambda \underbrace{\sum_i a_i}_{=0} = 1$$

por lo que $t' \in \Delta_k$. Más aún, tenemos que

$$x = \sum_i t_i x_i = \sum_i (t_i' + \lambda a_j) x_i = \sum_i t_i' x_i - \lambda \underbrace{\sum_i a_i x_i}_{=0} = \sum_i t_i' x_i.$$

Pero por construcción vale que $t_j' = t_j - \lambda a_j = 0$ por lo que la última escritura de x como combinación convexa de los x_i tiene un sumando menos, y luego el resultado se obtiene por hipótesis inductiva. □

Corolario A.2.3. *Si $K \subseteq \mathbb{R}^n$ es compacto, entonces $co(K)$ es compacto. En particular, $\overline{co}(K) = co(K)$.*

Demostración. La función continua $K^{n+1} \times \Delta_n \to \operatorname{co}(K)$ dada por $(x, t) \mapsto \sum_i t_i x_i$ es sobreyectiva por el Teorema A.2.2. Como $K^{n+1} \times \Delta_n$ es compacto, se sigue el resultado. □

A.3. El teorema de Krein-Milman

En esta sección demostramos uno de los resultados centrales en el estudio de puntos extremales: el teorema de Krein-Milman.

Lema A.3.1. *Sean X un espacio vectorial, $A \subseteq B \subseteq C \subseteq X$. Si A ext B y B ext C, entonces A ext C.*

Demostración. Sean $x, y \in C$ y $0 < t < 1$ tales que $tx + (1-t)y \in A$. Como $A \subseteq B$, vale que $tx + (1-t)y \in B$ y, por la extremalidad de B en C, tenemos que $x, y \in B$. Pero entonces $x, y \in B$ son tales que $tx + (1-t)y \in A$, de donde $x, y \in A$ por la extremalidad de A en B. □

Recordemos una de las versiones geométricas del teorema de Hahn-Banach.

Teorema A.3.2. *Sea X un espacio localmente convexo, y sean $A, B \subseteq X$ convexos disjuntos, con A abierto. Entonces existen $\varphi \in X'$ y $\beta \in \mathbb{R}$ tales que para todos $a \in A$ y $b \in B$ vale que*

$$Re\ \varphi(a) < \beta \leq Re\ \varphi(b).$$

Teorema A.3.3 (Krein-Milman)**.** *Sean X un espacio localmente convexo, $\varnothing \neq K \subseteq X$ compacto. Entonces valen las siguientes afirmaciones:*

(1) $Ext(K) \neq \varnothing$;

(2) $K \subseteq \overline{co}(Ext(K))$;

(3) *si K es convexo, entonces* $K = \overline{co}(Ext(K))$.

Demostración.

(1) Sea $\Omega := \{S \subseteq K : S \text{ extremal y cerrado}, S \neq \varnothing\}$ ordenado por inclusión. $\Omega \neq \varnothing$ pues $K \in \Omega$. Si $\mathcal{C} \subseteq \Omega$ es una cadena no vacía, afirmamos que $S = \bigcap_{A \in \mathcal{C}} A \in \Omega$. En efecto, S es cerrado por ser intersección de conjuntos cerrados. Por otra parte, como $\mathcal{C}$ es una cadena de conjuntos no vacíos, tiene la propiedad de intersección finita. Al ser K compacto y $\mathcal{C}$ una familia de cerrados de K, tenemos que $S \neq \varnothing$. Finalmente, S es extremal en K pues cada $A \in \mathcal{C}$ lo es. Esto termina de probar que $S \in \Omega$, y consecuentemente es una cota inferior de $\mathcal{C}$.

Por el lema de Zorn, Ω tiene algún elemento minimal S. Afirmamos que $\#S = 1$. Sabemos que $\#S \geq 1$ pues $S \neq \varnothing$. Supongamos que $\#S \geq 2$ y sean $x, y \in S$ distintos. Luego, existe $\varphi \in X'$ tal que Re $\varphi(y) <$ Re $\varphi(x)$. Como Re φ es continua y S es compacto, vale que Re φ alcanza máximo m en S. Llamemos $S_\varphi = \{s \in S : \text{Re } \varphi(s) = m\}$. El conjunto S_φ es cerrado pues S lo es y Re φ es continua, y es no vacío pues ya dijimos que Re φ alcanza máximo en S. Veamos que es extremal en K. Como S es extremal en K, por el

Lema A.3.1, basta ver que S_φ es extremal en S. Supongamos entonces que $x', y' \in S$ y $0 < t < 1$ son tales que $tx' + (1-t)y' \in S_\varphi$. Luego

$$m = \text{Re } \varphi(tx' + (1-t)y') = t \underbrace{\text{Re } \varphi(x')}_{\leq m} + (1-t) \underbrace{\text{Re } \varphi(y')}_{\leq m} \leq m$$

de donde Re $\varphi(x') =$ Re $\varphi(y') = m$ como queríamos ver. Por último, notemos que $y \notin S_\varphi$ pues Re $\varphi(y) <$ Re $\varphi(x) \leq m$. Pero entonces $S_\varphi \in \Omega$ cumple que $S_\varphi \subsetneq S$, lo cual contradice la minimalidad de S. El absurdo provino de suponer que $\#S \geq 2$, de donde $\#S = 1$. Esto termina de probar que $\text{Ext}(K) \neq \varnothing$.

(2) Sea $C = \overline{\text{co}}(\text{Ext}(K))$ y supongamos que $K \nsubseteq C$. Sea $x_0 \in K \setminus C$ y V entorno abierto convexo de 0 tal que $x_0 \notin C + V =: A$ (basta tomar V un entorno abierto, convexo y simétrico del 0 tal que $x_0 + V \subseteq X \setminus C$). Como A es un abierto convexo, por el Teorema A.3.2 existen $\varphi \in X'$ y $\beta \in \mathbb{R}$ tales que para todo $x \in A$ (y en particular, para todo $x \in \text{Ext}(K)$) vale que Re $\varphi(x) < \beta \leq$ Re $\varphi(x_0)$. Si llamamos γ al máximo de Re φ en K y $K_\varphi = \{x \in K : \text{Re } \varphi(x) = \gamma\}$, razonando igual que antes, tenemos que K_φ es compacto, no vacío y extremal en K. Por (1), existe $x_1 \in \text{Ext}(K_\varphi)$ y como K_φ ext K, entonces nuevamente por el Lema A.3.1, $x_1 \in \text{Ext}(K)$. Con esto,

$$\gamma = \text{Re } \varphi(x_1) < \beta \leq \text{Re } \varphi(x_0) \leq \gamma,$$

lo cual es absurdo. El absurdo provino de suponer que $K \nsubseteq C$, de donde $K \subseteq C$.

(3) $K \subseteq \overline{\text{co}}(\text{Ext}(K)) \subseteq \overline{\text{co}}(K) = K$, donde la última igualdad vale pues K es convexo y cerrado. □

Mostremos una aplicación clásica de este resultado.

Corolario A.3.4 (c_0 no es un dual)**.** *No existe ningún espacio de Banach X tal que X' sea isométricamente isomorfo a c_0.*

Demostración. Supongamos que sí. Luego, c_0 con la topología débil* es un espacio localmente convexo, y por el teorema de Banach-Alaoglu vale que $\overline{B}_{c_0}$ es compacta con esta topología. Por el teorema de Krein-Milman, $\overline{B}_{c_0}$ debe tener puntos extremales. Veamos que no es el caso.

Si $x \in \overline{B}_{c_0}$ es extremal, entonces necesariamente $||x||_{c_0} = 1$. Sea $N \in \mathbb{N}$ tal que para todo $n \geq N$ vale que $|x_n| < 1/2$. Definimos las sucesiones

$$y_n = \begin{cases} x_n + \frac{1}{4} & \text{si } n = N \\ x_n & \text{si } n \neq N \end{cases} \quad , \quad z_n = \begin{cases} x_n - \frac{1}{4} & \text{si } n = N \\ x_n & \text{si } n \neq N \end{cases} \quad .$$

Vale que $y, z \in \overline{B}_{c_0}$ y $x = \frac{y+z}{2}$, contradiciendo la extremalidad de x. □

Con un argumento similar se puede probar que $L^1[0,1]$ no es un dual.

A.4. El teorema de Minkowski

Por último, probamos el teorema de Minkowski que dice, esencialmente, que en espacios de dimensión finita, el teorema de Krein-Milman vale sin necesidad de tomar clausura.

Teorema A.4.1 (teorema del interior relativo)**.** *Sea $K \subseteq \mathbb{R}^n$ un convexo no vacío y sea $L \subseteq \mathbb{R}^n$ la variedad afín generada por K. Entonces el interior de K relativo a L es no vacío.*

Demostración. Como $K \neq \varnothing$, trasladando podemos suponer que $0 \in K$. Más aún, cambiando $\mathbb{R}^n$ por L podemos suponer que la variedad afín generada por K (que coincide con el subespacio generado por K pues $0 \in K$) es $\mathbb{R}^n$. En estas condiciones, debe existir una base $\{v_1, \ldots, v_n\} \subseteq K$ de $\mathbb{R}^n$. Llamemos $\{v_1', \ldots, v_n'\}$ a la correspondiente base dual.

Si
$$U = \left\{ x \in \mathbb{R}^n \;\middle|\; \forall i : v_i'(x) > 0, \ \sum_i v_i'(x) < 1 \right\}, \tag{A.1}$$
entonces vale que U es abierto y no vacío pues $\sum_i \frac{1}{2n} v_i \in U$. Además, si $x \in U$, entonces
$$x = \Big(1 - \sum_i v_i'(x)\Big) 0 + \sum_i v_i'(x) v_i \in \operatorname{co}(0, v_1, \ldots, v_n),$$
de donde $U \subseteq \operatorname{co}(0, v_1, \ldots, v_n) \subseteq K$. Esto termina de probar el interior de K es no vacío. □

Proposición A.4.2. *Si $K \subseteq \mathbb{R}^n$ es convexo y genera afínmente a $\mathbb{R}^n$, entonces $\overline{K} = \overline{K^\circ}$.*

Demostración. Basta probar que $\overline{K} \subseteq \overline{K^\circ}$. Sean $x \in \overline{K}$ y $\varepsilon > 0$. El conjunto $K_1 = K \cap B_\varepsilon(x)$ es convexo y no vacío. Trasladando, podemos suponer que $0 \in K_1$ (y en particular $0 \in K$). Sean $v_1, \ldots v_n \in K$ linealmente independientes, que existen pues $0 \in K$ y K genera afínmente a $\mathbb{R}^n$. Para cada $k > n$, sea
$$x_k := \sum_{i=1}^n \frac{1}{k} v_i.$$

Definiendo U como en (A.1), es claro que $x_k \in U \subseteq K^\circ$ y $x_k \xrightarrow{k \to \infty} 0$. Al ser $B_\varepsilon(x)$ un entorno abierto de 0 (pues $0 \in K_1$), para algún $k_0 > n$ se tiene que $x_{k_0} \in B_\varepsilon(x)$ y, como $\varepsilon > 0$ era arbitrario, esto prueba que $x \in \overline{K^\circ}$. □

Lema A.4.3. *Sean X un espacio vectorial real y $C \subseteq X$ convexo y no vacío. Supongamos que $\varphi : X \to \mathbb{R}$ es un funcional no nulo tal que φ alcanza un máximo α en C, y sea $H = \varphi^{-1}(\alpha)$. Entonces $Ext(C \cap H) = Ext(C) \cap H$.*

Demostración. La inclusión "$(\supseteq)$" es inmediata de la definición de punto extremal. Probemos la otra inclusión.

Sea $x \in \mathrm{Ext}(C \cap H)$ y supongamos que $y, z \in C$, $0 < t < 1$ son tales que $x = ty + (1-t)z$. Luego

$$\alpha = \varphi(x) = t\underbrace{\varphi(y)}_{\leq\alpha} + (1-t)\underbrace{\varphi(z)}_{\leq\alpha} \leq \alpha$$

de donde $\varphi(y) = \varphi(z) = \alpha$. Es decir, $y, z \in H$ y como x es extremal en $C \cap H$, concluimos que $y = z = x$. $\square$

Teorema A.4.4 (Minkowski)**.** *Sea $K \subseteq \mathbb{R}^n$ compacto, convexo y no vacío. Entonces $K = co(Ext(K))$.*

Demostración. Sea L la variedad afín generada por K. Probamos el resultado por inducción en $m = \dim(L)$. Para $m = 0$ no hay nada que hacer.

Supongamos que $m \geq 1$. Al igual que antes, podemos suponer que $m = n$ y que $0 \in K$. Como K es convexo, sólo debemos ver que $K \subseteq \mathrm{co}(\mathrm{Ext}(K))$. Tomemos $x \in K$ y consideremos los siguientes casos.

(1) $x \in \partial K = K \setminus K^\circ$. Como K es convexo, tenemos que K° es convexo y es no vacío por el Teorema A.4.1. Luego, por el Teorema A.3.2 existe $\varphi : \mathbb{R}^n \to \mathbb{R}$ lineal tal que para todo $y \in K^\circ$ vale que $\varphi(y) < \varphi(x)$. En vistas de la Proposición A.4.2, esto implica que para todo $y \in K$ se tiene que $\varphi(y) \leq \varphi(x)$. Dicho de otro modo, $\varphi(x)$ es el máximo de φ en K. Sea $H = x + \ker\varphi$ y $K_1 = K \cap H$. Luego K_1 es compacto, convexo, no vacío y genera una variedad lineal de dimensión a lo sumo $m-1$. Como $x \in K_1$, por hipótesis inductiva vale que $x \in \mathrm{co}(\mathrm{Ext}(K_1))$. Pero por el Lema A.4.3 tenemos que $\mathrm{Ext}(K_1) = \mathrm{Ext}(K) \cap H \subseteq \mathrm{Ext}(K)$, y entonces $x \in \mathrm{co}(\mathrm{Ext}(K))$.

(2) $x \in K^\circ$. Sea $A := \{t \in \mathbb{R}_{\neq 0} : x + te_1 \in K\}$. $A \neq \varnothing$ pues $x \in K^\circ$, y es acotado pues K es compacto. Sean $a := \sup A > 0$ y $b := \operatorname{ínf} A < 0$. Llamemos $x_1 = x + ae_1$ y $x_2 = x + be_1$. Como K es cerrado, $x_1, x_2 \in K$ y, por las definiciones de a, b, vale que $x_1, x_2 \notin K^\circ$. Es decir, $x_1, x_2 \in \partial K$. Observemos que

$$x = \frac{-b}{a-b}x_1 + \frac{a}{a-b}x_2,$$

y como $a > 0$, $b < 0$, esto significa que x es combinación convexa de x_1 y x_2. Pero por (1), sabemos que x_1 y x_2 son a su vez combinaciones convexas de $\mathrm{Ext}(K)$. Por lo tanto, $x \in \mathrm{co}(\mathrm{Ext}(K))$ como queríamos probar. $\square$

Bibliografía

[1] N. Aronszajn. Theory of reproducing kernels. *Trans. Am. Math. Soc.*, 68:337–404, 1950.

[2] Francesca Bartolucci, Ernesto De Vito, Lorenzo Rosasco, and Stefano Vigogna. Understanding neural networks with reproducing kernel Banach spaces. *Appl. Comput. Harmon. Anal.*, 62:194–236, 2023.

[3] Kristian Bredies and Marcello Carioni. Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calc. Var. Partial Differ. Equ.*, 59(1):26, 2020. Id/No 14.

[4] Ovidiu Calin. *Deep learning architectures. A mathematical approach.* Springer Ser. Data Sci. Cham: Springer, 2020.

[5] Michael F. Driscoll. The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 26:309–316, 1973.

[6] Rick Durrett. *Probability. Theory and examples*, volume 49 of *Camb. Ser. Stat. Probab. Math.* Cambridge: Cambridge University Press, 5th edition edition, 2019.

[7] R. Fortet. Espaces à noyau reproduisant et lois de probabilités des fonctions aléatoires. *Ann. Inst. Henri Poincaré, Nouv. Sér., Sect. B*, 9:41–58, 1973.

[8] Antonio F. Gualtierotti. *Detection of random signals in dependent Gaussian noise.* Cham: Springer, 2015.

[9] Christopher Heil. What is ... a frame? *Notices Am. Math. Soc.*, 60(6):748–750, 2013.

[10] Gopinath Kallianpur. Zero-one laws for Gaussian processes. *Trans. Am. Math. Soc.*, 149:199–211, 1970.

[11] Milan N. Lukic and Jay H. Beder. Stochastic processes with sample paths in reproducing kernel Hilbert spaces. *Trans. Am. Math. Soc.*, 353(10):3945–3969, 2001.

[12] Reinhold Meise and Dietmar Vogt. *Introduction to functional analysis. Transl. from the German by M. S. Ramanujan*, volume 2 of *Oxf. Grad. Texts Math.* Oxford: Clarendon Press, 1997.

[13] Vern I. Paulsen and Mrinal Raghupathi. *An introduction to the theory of reproducing kernel Hilbert spaces*, volume 152 of *Camb. Stud. Adv. Math.* Cambridge: Cambridge University Press, 2016.

[14] Ingo Steinwart. Reproducing kernel Hilbert spaces cannot contain all continuous functions on a compact metric space. *Arch. Math.*, 122(5):553–557, 2024.

[15] N. N. Vakhaniya, V. I. Tarieladze, and S. A. Chobanyan. *Probability distributions on Banach spaces. Transl. from the Russian by Wojbor A. Woyczynski*, volume 14 of *Math. Appl., Sov. Ser.* Dordrecht etc.: D. Reidel Publishing Company, 1987.

[16] Richard L. Wheeden and Antoni Zygmund. *Measure and integral. An introduction to real analysis*. Boca Raton, FL: CRC Press, 2nd ed. edition, 2015.