Universidad de Valladolid
Facultad de Ciencias

Trabajo Fin de Grado

Grado en Matemáticas

Espacios de Hilbert con Núcleo Reproductor y Aplicaciones

Autor: Pablo Merino San José

Tutor: Javier Sanz Gil

Índice general

Introducción

En este trabajo, estudiamos la teoría y algunas aplicaciones de un tipo determinado de espacio de funciones con valores en $\mathbb{R}$ o $\mathbb{C}$: el *espacio de Hilbert con núcleo reproductor* o *RKHS*[1]. La caracterización de estos espacios viene dada por la continuidad de los funcionales evaluación en el RKHS. En otras palabras, denotando por H a uno de estos espacios, éstos se caraterizan por la continuidad de las aplicaciones lineales $\delta_x : H \to \mathbb{K}$ tales que $\delta_x(f) = f(x)$ para todo $f \in H$. Esta condición permite, a través del teorema de representación de Riesz, generalizar la evaluación de un elemento arbitrario $f \in H$ sobre un punto cualquiera de X como el producto escalar $\langle f, k_x \rangle_H$, con k_x un elemento de H que solo dependerá de x. Esta propiedad, llamada *propiedad reproductora*, genera un relación biunívoca entre los RKHSs y las funciones $k : X \times X \to \mathbb{K}$, dadas por $k_x(x') = k(x', x)$ para todo $x, x' \in X$. Aunque de simple concepción, en un RKHS nos encontramos con la extensión de propiedades, tales como acotación, medibilidad, continuidad o integrabilidad, satisfechas por todos los elementos del propio RKHS con tal de que el núcleo las satisfaga. La deducción de esto no es trivial, pero su estudio nos llevará a otro estadio relevante de estos espacios: su capacidad de representación en términos del núcleo. Esto simplifica aún más su análisis, y clarifica su futura aplicación sobre métodos de Aprendizaje Estadístico.

Los núcleos reproductores, tal y como los entendemos hoy, fueron estudiados por primera vez en la década de 1900 a 1910 por Zaremba, en su trabajo sobre problemas de valor frontera para funciones armónicas. Él fue el primero en relacionar un núcleo a una clase de funciones, y en formalizar su propiedad reproductora. No obstante, no fue hasta 1921 cuando se empezó a dar una teoría no tan particularizada de núcleos reproductores, mérito de Bergman, quien descubrió la propiedad reproductora de los núcleos construidos a partir de sistemas ortogonales de funciones armónicas y analíticas, en una o varias variables. Durante los 20 años siguientes, fueron apareciendo multitud de importantes resultados sobre el uso de estos núcleos hasta que, en 1943, Aronszajn aportó la primera teoría general de núcleos reproductores. Una de las claves de este trabajo fue la demostración de la correspondencia biunívoca entre núcleos reproductores y funciones semidefinidas positivas y simétricas. La teoría de estas últimas funciones fue desarrollada por Mercer en 1909, de lo que se siguieron muchas aplicaciones en la teoría de la transformada de Fourier (Bochner, 1932), y en la de grupos topológicos.

Sin embargo, fue en 1948 cuando el mismo Aronszajn publicó su importante trabajo (ver [6]), con el cual mostró por vez primera, y de una forma general, la relación entre núcleos y un determinado tipo de espacios de Hilbert. La última extensión de este formalismo data de 1962, cuando Schwartz introdujo la noción de subespacios de Hilbert de un espacio vectorial topológico, y probó la correspondencia entre dichos subespacios y los núcleos, generalizando aún más la teoría de Aronszajn. En los últimos 50 años, hemos visto un enorme desarrollo en el uso de núcleos reproductores, sobre todo en probabilidad y estadística matemática. El desarrollo más reciente responde a la aplicación de la teoría de RKHSs sobre métodos de Aprendizaje Estadístico, lo que ha suscitado una notable importancia de éstos en la teoría del *Machine Learning*.

En el planteamiento de este trabajo, se parte de las nociones básicas e indispensables de las funciones núcleo en general, particularizando sobre los núcleos a valores reales y su caracterización por simetría y por la condición de semidefinida positiva. A continuación, se sigue con el estudio de núcleos reproductores y RKHSs. Veremos que a todo RKHS le corresponde un único núcleo reproductor, y viceversa. Esto simplifica notablemente el análisis, centrando el foco en el estudio de estos núcleos. Una vez establecida la correspondencia biunívoca mencionada, estudiamos la acotación, continuidad, medibilidad e integrabilidad de los elementos de estos espacios, sin perder nunca de vista la propiedad reproductora formalizada por el núcleo. Para concluir el primer capítulo, mostramos algunos resultados notables de la representación general de funciones de los RKHSs, en el caso de núcleos a valores reales. Para la teoría fundamental de núcleos, de RKHSs, de la correspondencia entre estos

[1]Siglas inglesas: Reproducing Kernel Hilbert Spaces

dos, y de las propiedades tales como continuidad o integrabilidad estudiadas a partir del núcleo, las referencias han sido, principalmente, [11], [4], [1], y [6]. En cuanto a la parte de representación a través del teorema de Mercer, se han tratado [3] y [4].
Este no sería un trabajo completo si no se expusieran ejemplos de algunos de los RKHSs hoy conocidos. En el capítulo 2, nos centramos en el estudio de algunos de estos, elegidos por su elegancia y su calado teórico. En concreto, hablaremos de los espacios de Paley-Wiener, de Bergman y de Hardy. Incluiremos en este capítulo, con énfasis en el espacio de Paley-Wiener, una breve incursión en la teoría del muestreo, en la que la estructura de RKHS tiene interesantes consecuencias. Las referencias de este capítulo han sido [13], [14], [10] y [12] para el estudio de los espacios de Hardy y Bergman, así como las mismas dos primeras referencias junto a [2] para los apartados de teoría del muestro y espacio de Paley-Wiener.
En el tercer y último capítulo, investigamos la aplicación de la teoría de RKHSs sobre el Aprendizaje Estadístico, cuya meta es la de encontrar una relación funcional, de cara a predicciones futuras, entre datos de entrada (*input*) y valores de salida o de respuesta (*output*). No entraremos en la complejidad de la teoría estadística, sino que nos centraremos en su utilidad desde el punto de vista del Análisis Matemático. La bibliografía de este capítulo incluye [8], [9], [7] y [5], los cuales cubren el espectro de todos los resultados de análisis funcional, topología y análisis real tratados. Los resultados de Teoría de la Probabilidad que hemos incluido en el capítulo siguen las líneas de [8] y [1].
Ejemplos reales en los que nos podemos encontrar esta situación podrían ser: los diagnósticos (*output*) a partir de determinadas medidas clínicas (*input*), o la evaluación en el futuro del precio de un vivienda (*output*) a partir de determinadas caracterísitcas actuales del mercado (*input*). En la teoría de Aprendizaje Estadístico, el *aprendizaje supervisado* consiste en la tarea de obtener una función que relacione datos input del conjunto de entrada, y datos output del conjunto de salida, a partir de pares conocidos del producto de estos dos conjuntos, infiriendo una relación funcional entre estos. En contraposición, y en esta misma teoría, nos encontramos con el *aprendizaje no supervisado*, cuyo objetivo es el de encontrar una relación funcional análoga, pero no a partir de muestras conocidas de datos de entrada y de salida, sino de conjuntos de datos que no se pueden etiquetar a priori. Esta última modalidad no es de interés en este trabajo.
Los métodos en que nos centraremos en este último capítulo están enmarcados en las llamadas *Máquinas de soporte vectorial* o *SVM* (*Support Vector Machines*). En particular, tocaremos principalmente procesos de clasificación, por no ser de nuestro interés la complejidad del propio método estadístico, sino su relación con los RKHSs. Esencialmente, un método de SVM lo caracterizamos a través de tres fases: enviamos los datos de entrada de X a un RKHS H, a través de una aplicación característica $\phi : X \to H$; establecemos un criterio de clasificación en H, por medio de un hiperplano definido con el producto interno de este espacio; y formalizamos, a través de propiedades del núcleo como su propiedad reproductora o su universalidad, criterios de separación que permitan clasificar. Veremos que esta construcción va de la mano con las llamadas *funciones pérdida* y *funciones riesgo*. Ellas son las que, sobre un espacio de medida en $X \times Y$ con Y el conjunto de valores de salida, darán la forma analítica a estos criterios de clasificación.
Es oportuno remarcar que, si bien trataremos algunos de los resultados más importantes que relacionan RKHSs y la teoría que hay detrás del Aprendizaje Estadístico, no entraremos en la aplicación o implementación del método, sino en la teoría funcional subyacente. Es decir, no se tratará de desarrollar el código de estos procesos, sino, por ejemplo, de demostrar la existencia y unicidad de soluciones de SVM a partir de la estructura reproductora de los RKHSs.
Nosotros nos pararemos aquí, una vez entendida la importancia de los RKHS como espacios de búsqueda de minimizadores de funciones riesgo, y después de haber visto, en los dos capítulos anteriores, la teoría fundamental de los RKHSs, y algunos ejemplos reconocidos.
Las asignaturas del Grado con que mejor entronca este trabajo serían, fundamentalmente, Introducción a los Espacios de Funciones, Variable Compleja, Ampliación de Análisis Matemático y Análisis Matemático. Esto se aprecia notablemente en los capítulos primero y segundo, en que se presenta la teoría, enmarcada en el análisis funcional clásico, de los RKHSs y las funciones núcleo, así como

ejemplos importantes que tratan análisis complejo en el caso de los espacios de Bergman y Hardy, y análisis real y funcional en el caso de Paley-Wiener. También en el primer capítulo, y ampliamente en el tercero, se hacen referencias, aunque sin demasiada complejidad, a resultados de teoría de la medida. En cuanto al tercer capítulo, además del análisis funcional recurrente a lo largo de todo el trabajo, es obligada la incursión en teoría de la probabilidad de cara a formalizar y presentar los modelos de SVM que nosotros estudiamos. Esto último se correspondería, principalmente, con la asignatura de Teoría de la Probabilidad y Estadística Matemática.
Al final de este trabajo hay dos apéndices con resultados auxiliares, uno referente al análisis funcional y otro al análisis real, que por su extensión no se han incluido en el cuerpo de nuestro desarrollo, pero que son utilizados. En el primero de ellos, se trata la compleción a un espacio de Hilbert a partir de un espacio con producto interno, continuidad y ortogonalidad en espacios de Hilbert, la compacidad de operadores, el teorema de representación de Fréchet-Riesz, algunas disquisiciones sobre el adjunto de un operador, y resultados relevantes y empleados sobre topologías débiles en un espacio de Banach. En cuanto al segundo apéndice, recogemos algunas nociones, imprescindibles en nuestro desarrollo, sobre teoría de la medida y la teoría de la transformada de Fourier.

Capítulo 1

Núcleos reproductores y RKHS

Empezaremos dando una serie de resultados indispensables en la comprensión y desarrollo de la teoría de los espacios de Hilbert con núcleo reproductor. En adelante, nos referiremos a este tipo de espacios como **RKHS**. Definiciones y resultados ya presentados en el Grado, o que podrían interrumpir nuestro desarrollo, han sido incluidos en dos apéndices, uno dedicado al análisis funcional, y otro al análisis real. Daremos esto por sabido (por ejemplo, la definición de producto interno de los espacios que consideremos, cuando lo haya, o qué es un espacio de Hilbert), simplemente recordando que $\mathbb{K}$ denota al cuerpo sobre el que trabajamos, cuando no queramos especificar si $\mathbb{K} = \mathbb{R}$ ó $\mathbb{K} = \mathbb{C}$. También, H denotará, salvo que se diga lo contrario, al $\mathbb{K}$ - espacio vectorial de Hilbert en cuestión. A lo largo de este trabajo, cuando se esté trabajando sobre el espacio $\mathbb{R}^n$ o $\mathbb{C}^n$ para algún $n \in \mathbb{N}$, y no se haga notar lo contrario, el producto interno $\langle x, y \rangle$ será el producto escalar usual en estos espacios para todo par de elementos x, y en dicho espacio. También denotaremos este producto, alternativamente, como $\langle x, y \rangle = x \cdot y = x_1\overline{y}_1 + \cdots + x_n\overline{y}_n$. Claramente, estas consideraciones se extienden a las normas inducidas por estos productos internos. Note el lector que, mientras no especifiquemos si trabajamos en valores complejos o en valores reales, usaremos la definición de producto interno con llegada en $\mathbb{C}$.

1.1. Núcleos. Primeras propiedades

Definimos, entrando ya en nuestra materia, lo que entenderemos por núcleo sobre un conjunto arbitrario y no vacío.

Definición 1.1.1. Sea X un conjunto no vacío. Se dice que una función $k : X \times X \to \mathbb{K}$ es un **núcleo** en X si existen un $\mathbb{K}$ - espacio de Hilbert H con producto interno $\langle \cdot, \cdot \rangle$, y una aplicación $\phi : X \to H$ tales que, para todos $x, x' \in X$, se tiene que

$$k(x, x') = \langle \phi(x'), \phi(x) \rangle. \tag{1.1.1}$$

Llamaremos **aplicación característica** a ϕ y **espacio característico** a H.

Así definido, no podemos decir nada de la unicidad de la aplicación caracterísitca ϕ ni del espacio característico H. Sea, por ejemplo, el conjunto $X := \mathbb{R}$ y $k(x, x') := xx'$, con x y x' en $\mathbb{R}$, y con el producto escalar usual en $\mathbb{R}$. Basta coger $\phi := \mathrm{id}_{\mathbb{R}} : X \to H$ con $X := \mathbb{R}$ y $H := \mathbb{R}$, de forma que

$$\langle id_{\mathbb{R}}(x'), id_{\mathbb{R}}(x) \rangle = xx' = k(x, x').$$

Sin embargo, también podemos escoger, manteniendo $X = \mathbb{R}$, la aplicación característica $\phi : X \to \mathbb{R}^2$ dada por $\phi(x) := (\frac{x}{\sqrt{2}}, \frac{x}{\sqrt{2}})$ para todo $x \in X$, de modo que, escogiendo ahora el producto escalar usual en $\mathbb{R}^2$, se tiene

$$\langle \phi(x'), \phi(x) \rangle = \frac{x'}{\sqrt{2}} \frac{x}{\sqrt{2}} + \frac{x'}{\sqrt{2}} \frac{x}{\sqrt{2}} = xx' = k(x, x').$$

De hecho, bastaría tomar un $n \in \mathbb{N}$ arbitrario, de nuevo $X = \mathbb{R}$, y $\phi : X \to \mathbb{R}^n$ dado por $\phi(x) = (\frac{x}{\sqrt{n}}, \ldots, \frac{x}{\sqrt{n}})$, de forma que, ahora con el producto escalar usual en $\mathbb{R}^n$,

$$\langle \phi(x'), \phi(x) \rangle = \frac{x'}{\sqrt{n}}\frac{x}{\sqrt{n}} + \cdots + \frac{x'}{\sqrt{n}}\frac{x}{\sqrt{n}} = \sum_{k=1}^{n} \frac{x'}{\sqrt{n}}\frac{x}{\sqrt{n}} = xx' = k(x, x').$$

Daremos ahora una serie de resultados sencillos que nos permiten dar ejemplos algo menos triviales de núcleo en un conjunto arbitrario y no vacío X.
Para el lema que sigue, aclaramos la notación que utilizaremos para los espacios $l^p(\mathbb{K})$, con $\mathbb{K}$ un cuerpo. Para $p \in [1, \infty)$, denotaremos $l^p = l^p(\mathbb{K})$ al espacio de sucesiones con términos en el cuerpo $\mathbb{K}$ tales que $x = (x_n)_{n=1}^{\infty} \in l^p$ si y sólo si $||x||_p = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} < \infty$. En el caso $p = 2$, dicha norma proviene de un producto interno dado como $\langle x, y \rangle = \sum_{n=1}^{\infty} x_n \overline{y}_n$, para $x = (x_n)_{n=1}^{\infty}$, $y = (y_n)_{n=1}^{\infty}$ en $l^2(\mathbb{K})$. En el caso $p = \infty$, $l^\infty(\mathbb{K})$ denota al conjunto de sucesiones acotadas con valores en $\mathbb{K}$, i.e., de sucesiones $x = (x_n)_{n=1}^{\infty}$ tales que $||x||_\infty = \sup\{|x_n| : n \in \mathbb{N}\} < \infty$. También, denotaremos a estos espacios l^p como $l^p(I)$, con I la familia de índices.

Lema 1.1.2. *Sea X un conjunto no vacío, y sean $f_n : X \to \mathbb{K}$, $n \in \mathbb{N}$ funciones tales que $(f_n(x)) \in l^2(\mathbb{K})$ para todo $x \in X$. Entonces,*

$$k(x, x') := \sum_{n=1}^{\infty} f_n(x)\overline{f_n(x')}, \tag{1.1.2}$$

para todo $x, x' \in X$, define un núcleo en X.

Demostración. Usamos la desigualdad de Hölder en el espacio de sucesiones l^2 como sigue

$$\sum_{n=1}^{\infty} |f_n(x) f_n(x')| \leq ||f_n(x)||_{l^2} ||f_n(x')||_{l^2}. \tag{1.1.3}$$

Entonces, como la sucesión $(f_n(x))$ está en l^2 para todo $x \in X$, la serie (1.1.2) converge absolutamente para todo $x, x' \in X$, lo que prueba la buena definición de la suma enunciada. Así pues, escribiendo $H := l^2$ y definiendo $\psi : X \to H$ con $\psi(x) := (\overline{f_n(x)})$ para todo $x \in X$, es inmediato que (1.1.2) nos da un núcleo en X: $k(x, x') = \langle \psi(x'), \psi(x) \rangle_{l^2}$. □

Seguimos con los resultado básicos para la construcción de núcleos.

Lema 1.1.3 (Restricción de núcleos)**.** *Sea X un conjunto no vacío, y sea k un núcleo en X. Sea ahora $\tilde{X}$ otro conjunto no vacío. Consideramos una aplicación $A : \tilde{X} \to X$. Entonces, la función $\tilde{k}$ definida por $\tilde{k}(x, x') := k(A(x), A(x'))$, $x, x' \in X$, es un núcleo en $\tilde{X}$. En particular, si $\tilde{X} \subset X$, entonces la restricción $k_{|\tilde{X}\times\tilde{X}}$ es un núcleo.*

Demostración. Denotamos por H y $\phi : X \to H$ a los espacio y aplicación característicos para el núcleo k. Consideramos la composición $\tilde{\phi} = \phi \circ A : \tilde{X} \to H$. Es inmediato ver que, tal y como se nos da la función $\tilde{k}$, se tiene que

$$\tilde{k}(x, x') = k(A(x'), A(x)) = \langle \phi \circ A(x'), \phi \circ A(x) \rangle_H = \langle \tilde{\phi}(x'), \tilde{\phi}(x) \rangle_H, \tag{1.1.4}$$

donde, en la última igualdad, hemos usado la definición de núcleo k en X. Así pues, manteniendo el espacio característico y considerando $\tilde{\phi}$ como la aplicación característica, tenemos que $\tilde{k}$ cumple la definición de núcleo con H y $\tilde{\phi}$.
En el caso de $\tilde{X} \subset X$, tomando la inclusión canónica $A = i : \tilde{X} \hookrightarrow X$, es inmediato que la restricción de k a $\tilde{X} \times \tilde{X}$ es un núcleo en $\tilde{X}$. □

El lema 1.1.3 nos permite afirmar que, dado $k : \mathbb{C}^d \times \mathbb{C}^d \to \mathbb{C}$, entonces su restricción a las d-uplas reales $k_{|\mathbb{R}^d \times \mathbb{R}^d}$ es también un núcleo en *sentido complejo*. De hecho, como probaremos ahora, si $k(x, x') \in \mathbb{R}$ para todo $x, x' \in X$, también será un núcleo en X en el *sentido real*.

Lema 1.1.4. *Sea $k : X \times X \to \mathbb{C}$ un núcleo, H uno de sus $\mathbb{C}$-espacios de Hilbert característicos, y $\phi : X \to H$ una de sus aplicaciones características. Supongamos que $k(x, x') \in \mathbb{R}$ para todo $x, x' \in X$. Entonces, $H_0 := H$ junto al producto interno que sigue*

$$\langle w, w' \rangle_{H_0} := \mathrm{Re}\langle w, w' \rangle_H, \qquad w, w' \in H_0, \tag{1.1.5}$$

es un $\mathbb{R}$-espacio de Hilbert, y $\phi : X \to H_0$ es una aplicación característica del núcleo k con espacio característico H_0.

Demostración. No es difícil probar que $\langle \cdot, \cdot \rangle_{H_0}$ es un producto escalar real, heredado del hecho de que $\langle \cdot, \cdot \rangle_H$ es un producto escalar con valores complejos. Además, se tiene que

$$k(x, x') = \langle \phi(x'), \phi(x) \rangle_H = \mathrm{Re}\langle \phi(x'), \phi(x) \rangle_H + i\mathrm{Im}\langle \phi(x'), \phi(x) \rangle_H.$$

Como, por hipótesis, $k(x, x') \in \mathbb{R}$ en todos los puntos de $X \times X$, entonces $\mathrm{Im}\langle \phi(x'), \phi(x) \rangle_H = \mathrm{Im}k(x, x') = 0$. Entonces

$$k(x, x') = \mathrm{Re}\langle \phi(x'), \phi(x) \rangle_H = \langle \phi(x'), \phi(x) \rangle_{H_0},$$

donde la última igualdad se deduce de la definición de $\langle \cdot, \cdot \rangle_{H_0}$. Así, una vez probemos que $(H_0, \langle \cdot, \cdot \rangle_{H_0})$ es espacio de Hilbert, ya podremos concluir que k es un núcleo con espacio y aplicación característicos H_0 y ϕ.

Para probar la completitud, tomemos $(x_n) \subset H_0$ una sucesión de Cauchy en $(H_0, \langle \cdot, \cdot \rangle_{H_0} := \mathrm{Re}(\langle \cdot, \cdot \rangle_H))$. Entonces

$$\langle x_n - x_m, x_n - x_m \rangle_{H_0}^{1/2} = ||x_n - x_m||_{H_0} \to 0, \;\; n, m \to \infty. \tag{1.1.6}$$

Para ver que esto implica que (x_n) es también de Cauchy en $(H, \langle \cdot, \cdot \rangle_H)$, usamos el hecho de que las normas inducidas por productos internos, como normas que son, son reales, luego las normas en H_0 y H coinciden. Recordemos que $H_0 = H$, con lo que los elementos x_n también están en H.

$$\begin{aligned} &||x_n - x_m||_H^2 = \langle x_n - x_m, x_n - x_m \rangle_H = \mathrm{Re}\left(\langle x_n - x_m, x_n - x_m \rangle_H\right) \\ &= \langle x_n - x_m, x_n - x_m \rangle_{H_0} = ||x_n - x_m||_{H_0}^2 \to 0, \;\; n, m \to 0. \end{aligned}$$

Entonces, (x_n) es también de Cauchy en $(H, \langle \cdot, \cdot \rangle_H)$. Como sabemos que este último espacio es completo, entonces existe un $x \in H = H_0$ tal que $x_n \to x$ en norma de H para $n \to \infty$, es decir, tal que $||x_n - x||_H \to 0$ para $n \to \infty$. Igual que antes, se sigue que

$$||x_n - x||_{H_0} = ||x_n - x||_H \to 0, \;\; n \to \infty. \tag{1.1.7}$$

Esto prueba, ahora sí, que $(H_0, \langle \cdot, \cdot \rangle_{H_0})$ es un espacio de Hilbert. □

Ahora vamos con algunos resultados algebraicos elementales, imprescindibles para la construcción de núcleos.

Lema 1.1.5 (Suma de núcleos y producto por un escalar)**.** *Sea X un conjunto no vacío, $\alpha \geq 0$, y k_0 y k_1 dos núcleos en X. Entonces, αk_0 y $k_0 + k_1$ son núcleos en X.*

Demostración. En primer lugar, dado $\alpha \geq 0$, probemos que αk_0 es núcleo. Para ello, denotamos por H_0 y por $\phi_0 : X \to H_0$ a sus espacio y aplicación característicos, y denotamos por $\langle \cdot, \cdot \rangle$ al producto interno definido sobre el espacio de Hilbert H_0. Así, para todo $x, x' \in X$, tenemos que

$$k_0(x, x') = \langle \phi_0(x'), \phi_0(x) \rangle. \tag{1.1.8}$$

Definimos una nueva aplicación característica como $\phi_0' := \sqrt{\alpha}\phi_0 : X \to H_0$ (notemos que $\alpha \geq 0$, luego la raíz cuadrada positiva está bien definida), con el mismo espacio característico H_0. De este modo, para todo $x, x' \in X$, es claro que

$$\alpha k_0(x, x') = \alpha\langle\phi_0(x'), \phi_0(x)\rangle = \langle\sqrt{\alpha}\phi_0(x'), \sqrt{\alpha}\phi_0(x)\rangle, \tag{1.1.9}$$

donde se ha usado que $\sqrt{\alpha} \in \mathbb{R}$, de tal forma que coincide con su conjugado.
De forma análoga, probamos que $k_0 + k_1$ es un núcleo. Además de los elementos anteriores, ahora denotamos por H_1 y por $\phi_1 : X \to H_1$ a los espacio y aplicación característicios del núcleo k_1. Así pues, basta tomar una aplicación característica suma $\phi_0 + \phi_1 : X \to H_0 \times H_1$ (donde $H_0 \times H_1$ denota al espacio de Hilbert suma directa de estos dos espacios de Hilbert H_0 y H_1). De esta forma, para todo $x, x' \in X$, tenemos que

$$\begin{aligned}(k_0 + k_1)(x, x') = k_0(x, x') + k_1(x, x') &= \langle\phi_0(x'), \phi_0(x)\rangle_{H_0} + \langle\phi_1(x'), \phi_1(x)\rangle_{H_1} \\ &= \langle\phi_0(x') + \phi_1(x'), \phi_0(x) + \phi_1(x)\rangle_{H_0\times H_1},\end{aligned}$$

donde, en la última igualdad, hemos usado la definición del producto interno sobre un espacio de Hilbert que es suma directa finita de espacios de Hilbert. □

El último lema demuestra que los núcleos en X forman un cono convexo[1]. Hay que remarcar que en el lema 1.1.5 no se dice nada de la diferencia de núcleos. De hecho, tomemos k_0 y k_1 dos núcleos en X con valores reales, con espacio característico H, tales que existe $x \in X$ con $k_0(x, x) - k_1(x, x) < 0$. Así pues, $k_0 - k_1$ no es un núcleo, pues si existiera una aplicación característica para la resta, $\phi : X \to H$, tendríamos, usando la positividad del producto escalar en H,

$$0 \leq \langle\phi(x), \phi(x)\rangle = k_0(x, x) - k_1(x, x) < 0,$$

lo cual es absurdo. Observemos que bastaría tomar, por ejemplo, $k_0(x, x') = xx'$ y $k_1(x, x') = (\sqrt{2}x)(\sqrt{2}x) = 2xx'$, con $X = H = \mathbb{R}$, de modo que, con $x = 1$, tendríamos $k_0(1, 1) - k_1(1, 1) < 0$, y por tanto $k_0 - k_1$ no es núcleo.

Lema 1.1.6. *Sea k_0 un núcleo en X_0 y k_1 un núcleo en X_1. Entonces, el producto $k_0 \cdot k_1$ es un núcleo en $X_0 \times X_1$. En particular, para $X_0 = X_1$, se tiene que $k(x, x') := k_0(x, x')k_1(x, x')$, $x, x' \in X$, es un núcleo en X.*

Demostración. Sea H_i un espacio característico y ϕ_i una aplicación característica del núcleo k_i, $i = 1, 2$. Utilizamos la noción de espacio con producto interno convencial en el producto tensorial de espacios de Hilbert $H_1 \otimes H_2$, y denotamos a su compleción a espacio de Hilbert como $H_1 \hat{\otimes} H_2$. Con ello, y por definición de los núcleos k_1 y k_2, se sigue que

$$\begin{aligned}k_1(x_1, x_1') \cdot k_2(x_2, x_2') &= \langle\phi_1(x_1'), \phi_1(x_1)\rangle_{H_1} \cdot \langle\phi_1(x_2'), \phi_1(x_2)\rangle_{H_2} \\ &= \langle\phi_1(x_1') \otimes \phi_2(x_2'), \phi_1(x_1) \otimes \phi_2(x_2)\rangle_{H_1\hat{\otimes}H_2}, \quad x_1, x_1' \in X_1, \quad x_2, x_2' \in X_2,\end{aligned}$$

lo que demuestra que la aplicación $\phi_1 \otimes \phi_2 : X_1 \times X_2 \to H_1 \hat{\otimes} H_2$ es característica del producto $k_1 \cdot k_2$, así como que $k_1 \cdot k_2$ es núcleo en $X_1 \times X_2$. i.e., es un núcleo que va de $(X_1 \times X_2) \times (X_1 \times X_2)$ al correspondiente cuerpo.
Probamos la última afirmación. Supongamos que $X_1 = X_2 = X$, y denotamos por $k = k_1 \cdot k_2$ al núcleo producto. Entonces, siendo rigurosos, el núcleo anterior quedaría como $k : (X \times X) \times (X \times X) \to \mathbb{K}$. Consideramos la restricción de este núcleo k al conjunto formado por los $((x_1, x_1'), (x_2, x_2')) \in (X \times X) \times (X \times X)$ tales que $(x_1, x_1') = (x_2, x_2')$. Entonces, por el último aserto del lema 1.1.3, se tiene que dicha restricción $k(x, x') = k_1(x, x') \cdot k_2(x, x') : X \times X \to \mathbb{K}$ es núcleo en X. □

[1] Definimos un cono convexo C sobre un cuerpo $\mathbb{K}$ como un conjunto con dos operaciones bien definidas, una entre elementos de C que denotaremos por $+$, y otra de elementos de $\mathbb{R}^+$ con elementos de C (en este orden), a la que denotaremos por $\cdot$. Esto es, tenemos que $x + y \in C$ y $\alpha \cdot x \in C$ para todos $x, y \in C$, y para todo $\alpha \geq 0$.

Ya tenemos algunas claves para construir núcleos no triviales. Por ejemplo, tomemos $X := \mathbb{R}$. Sabemos que $k(x, x') = xx'$ es un núcleo en X, con aplicación característica $\phi : X \to \mathbb{R}$ con $\phi = id_{\mathbb{R}}$. Pues bien, dado $n \in \mathbb{N}$, por el lema 1.1.6 aplicado $n-1$ veces, $k_n(x) = (xx')^n$ es un núcleo en X. Con esto y el lema 1.1.5, tenemos que el polinomio $p : X \to \mathbb{R}$, dado por $p(t) = a_m t^m + ... + a_1 t + a_0$, con coeficientes a_i no negativos, nos da el núcleo $k(x, x') = p(xx')$. Es inmediato comprobar que la aplicación $\phi : X \to H$ dada por $\phi(x) = (\sqrt{a_m}x^m, \ldots, \sqrt{a_1}x, \sqrt{a_0})$, con $H := \mathbb{R}^{m+1}$, es la aplicación característica. De ello se deduce que un núcleo polinomial, con polinomio original p de grado m, nos da un coste operativo[2] determinado por dicha m. Para generalizar esto, damos el siguiente resultado para una clase determinada de núcleos polinomiales.

Lema 1.1.7. *Sea $m \geq 0$ y $d \geq 1$ enteros, y $c \geq 0$ real. Entonces, $k(z, z') := (\langle z, z' \rangle + c)^m$, con $z, z' \in \mathbb{C}^d$, es un núcleo en $X := \mathbb{C}^d$. Además, su restricción a $\mathbb{R}^d$ nos da un núcleo real, esto es, con llegada en $\mathbb{K} = \mathbb{R}$.*

Demostración. Observamos que $\tilde{k} : X \times X \to \mathbb{C}$, con $\tilde{k}(x, x') = \langle x', x \rangle$ y $X := \mathbb{C}^d$, es un núcleo. Esto es trivial, sin más que tomar como espacio característico a $\mathbb{C}^d$ y como aplicación característica a $\phi = id_{\mathbb{C}^d}$. Así, usando el binomio de Newton obtenemos

$$k(z, z') = \sum_{k=0}^{m} \binom{m}{k} \langle z, z' \rangle^k c^{m-k} = \sum_{k=0}^{m} \binom{m}{k} \tilde{k}(z, z')^k c^{m-k}, \tag{1.1.10}$$

lo cual es una suma finita de núcleos en X multiplicados por escalares, ya que $\tilde{k}$ es núcleo en X. Esto implica, a través de los lemas 1.1.5 y 1.1.6, que k es un núcleo en X.
El hecho de que la restricción a $\mathbb{R}^d$ nos da un núcleo con valores reales es consecuencia directa de 1.1.4. □

En el caso particular de $m = 1$ y $c = 0$, obtenemos los llamados **núcleos lineales**, que no son más que el núcleo $\tilde{k}$ usado en la última demostración. Vamos más allá. Construimos núcleos usando series de Taylor.

Lema 1.1.8. *Sean $\mathring{B}_{\mathbb{C}}$ y $\mathring{B}_{\mathbb{C}^d}$ las bolas unidad abiertas de $\mathbb{C}$ y $\mathbb{C}^d$, respectivamente. Sea $r \in (0, \infty]$ y $f : r\mathring{B}_{\mathbb{C}} \to \mathbb{C}$ una función holomorfa con serie de Taylor*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \qquad z \in r\mathring{B}_{\mathbb{C}^d}. \tag{1.1.11}$$

Si $a_n \geq 0$ para todo $n \geq 0$, entonces

$$k(z, z') := f(\langle z, z' \rangle_{\mathbb{C}^d}) = \sum_{n=0}^{\infty} a_n \langle z, z' \rangle_{\mathbb{C}^d}^n, \qquad z, z' \in \sqrt{r}\mathring{B}_{\mathbb{C}^d},$$

*es un núcleo en $\sqrt{r}\mathring{B}_{\mathbb{C}^d}$. Además, su restricción a $X := \{x \in \mathbb{R}^d : ||x||_2 < \sqrt{r}\}$ es un núcleo con valores en $\mathbb{R}$. Denotaremos a este tipo de núcleo como de **tipo Taylor**.*

Demostración. Sean $z, z' \in \sqrt{r}\mathring{B}_{\mathbb{C}^d}$. Dado que $|\langle z, z' \rangle| \leq ||z||_2 ||z'||_2 < r$, entonces la serie de Taylor de f evaluada en $\langle z, z' \rangle$ está bien definida, dado el radio de la bola en que se ha supuesto bien definida la serie de Taylor de f. Recordemos la fórmula multinomial generalizada para cualquier $n \in \mathbb{N}$ y $z_1, \ldots, z_n \in \mathbb{C}$:

$$(z_1 + \cdots + z_d)^n = \sum_{\substack{j_1, \ldots, j_d \geq 0 \\ j_1 + \cdots + j_d = n}} n! \prod_{i=1}^{d} \frac{z_i^{j_i}}{j_i!}. \tag{1.1.12}$$

[2]Hablar de coste operativo en términos de núcleos polinomiales es hablar del número de operaciones necesarias para su evaluación.

Volviendo a nuestra demostración, sea z_i la componente i-ésima de $z \in \mathbb{C}^d$. Por (1.1.11), como esta serie es absolutamente convergente para $z \in r\mathring{B}_{\mathbb{C}^d}$, si cogemos los z, z' anteriores y usamos (1.1.12):

$$k(z,z') = \sum_{n=0}^{\infty} a_n (\sum_{j=1}^{d} z_j \overline{z}'_j)^n = \sum_{n=0}^{\infty} a_n \sum_{\substack{j_1,\dots,j_d \geq 0 \\ j_1+\dots+j_d=n}} c_{j_1,\dots,j_d} \prod_{i=1}^{d} (z_i \overline{z}'_i)^{j_i}$$
$$= \sum_{j_1,\dots,j_d \geq 0} a_{j_1+\dots+j_d} c_{j_1,\dots,j_d} \prod_{i=1}^{d} (\overline{z}'_i)^{j_i} \prod_{i=1}^{d} (z_i)^{j_i},$$

donde $c_{j_1,\dots,j_d} := \frac{n!}{\prod_{i=1}^{d} j_i!}$. Con los cálculos realizados, es inmediato que, si definimos[3] $\phi : X \to l^2(\mathbb{N}_0^d)$ de la forma

$$\phi(z) := (\sqrt{a_{j_1+\dots+j_d} c_{j_1,\dots,j_d}} \prod_{i=1}^{d} \overline{z}_i^{j_i})_{j_1,\dots,j_d}, \qquad z \in \sqrt{r}\mathring{B}_{\mathbb{C}^d},$$

entonces $k(z,z') = \langle \phi(z'), \phi(z) \rangle_{l^2(\mathbb{N}_0^d)}$ para cualesquiera $z, z' \in \sqrt{r}\mathring{B}_{\mathbb{C}^d}$. Notemos que, en esta construcción, se ha escogido la aplicación ϕ como característica, y el espacio de Hilbert $l^2(\mathbb{N}_0^d)$ como característico.
En cuanto a la restricción, basta usar el lema 1.1.3 aplicado al subconjunto $X := \{x \in \mathbb{R}^d : ||x||_2 < \sqrt{r}\}$ del enunciado, y tener en cuenta el resultado 1.1.4 de paso de núcleo con valores complejos a núcleo con valores reales, para concluir que k, restringido a $X \times X$, es un núcleo con valores reales. □

Un primer ejemplo inmediato de núcleo no trivial con un núcleo de tipo Taylor sería la función $k(x,x') := \exp(\langle x, x' \rangle)$, dados $x, x' \in \mathbb{K}^d$, con $\mathbb{K} = \mathbb{R}$ ó $\mathbb{K} = \mathbb{C}$ los cuerpos habituales, y $\langle \cdot, \cdot \rangle$ el producto escalar usual en el espacio $\mathbb{K}^d$. Tomamos $f(z) = \exp(z)$, la función exponencial compleja, que es holomorfa en todo $\mathbb{C}$, luego, en particular, es analítica en $z_0 = 0$. Por tanto, tomando un $r > 0$ arbitrario, podemos expandir $f(z)$, para todo $z \in \mathbb{C}^d$ (tomando $\mathbb{K} = \mathbb{C}$ en el caso general descrito), con el bien conocido desarrollo de Taylor de la función exponencial, dado por

$$f(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Luego, en paralelo con la demostración del lema 1.1.8, podemos afirmar que la función

$$k(z,z') = \sum_{n=0}^{\infty} \frac{\langle z, z' \rangle^n}{n!}, \qquad z, z' \in \sqrt{r}\mathring{B}_{\mathbb{C}^d}, \tag{1.1.13}$$

es un núcleo en $X = \mathbb{C}^d$, con aplicación característica

$$\phi(z) = (\sqrt{\frac{n!}{(j_1 + \dots + j_d)! \prod_{i=1}^{d} j_i!}} \prod_{i=1}^{d} \overline{z}_i^{j_i})_{j_1,\dots,j_d \geq 0}, \quad z \in \sqrt{r}\mathring{B}_{\mathbb{C}^d}. \tag{1.1.14}$$

Con lo que sabemos hasta ahora, estamos en condiciones de aportar los siguientes resultados, que nos darán unas nociones introductorias de algunos de los núcleos más extendidos.

Proposición 1.1.9. *Sean $d \in \mathbb{N}$, $\gamma > 0$, $z = (z_1, \dots, z_d) \in \mathbb{C}^d$, y $z' = (z'_1, \dots, z'_d) \in \mathbb{C}^d$. La función*

$$k_{\gamma,\mathbb{C}^d}(z,z') := \exp\left(-\frac{\sum_{j=1}^{d} (z_j - \overline{z}'_j)^2}{\gamma^2}\right), \qquad z, z' \in \mathbb{C}^d, \tag{1.1.15}$$

[3] En adelante, denotaremos $\mathbb{N}_0 = \{0, 1, 2, \dots\} = \mathbb{N} \cup \{0\}$, y $\mathbb{N} = \{1, 2, \dots\}$.

es un núcleo en $X = \mathbb{C}^d$, y su restricción a $\mathbb{R}^d \times \mathbb{R}^d$ es un núcleo con valores reales, llamado ***núcleo de función base radial*** *(o* ***núcleo RBF gaussiano****, donde RBF son siglas inglesas: Radial Basis Function), con anchura $\gamma > 0$. Además, el núcleo RBF puede calcularse como*

$$k_{\gamma,\mathbb{R}^d}(x, x') = k_\gamma(x, x') = \exp\left(-\frac{||x - x'||_2^2}{\gamma^2}\right), \qquad x, x' \in \mathbb{R}^d \tag{1.1.16}$$

Demostración. Fijamos $z, z' \in \mathbb{C}^d$. Si desarrollamos el cuadrado del exponente, es directo que

$$k_{\gamma,\mathbb{C}^d}(z, z') = \frac{\exp\left(2\gamma^{-2}\sum_{j=1}^d z_j\overline{z}_j\right)}{\exp\left(\gamma^{-2}\sum_{j=1}^d \overline{z}_j^2\right) \cdot \exp\left(\gamma^{-2}\sum_{j=1}^d z_j^2\right)}. \tag{1.1.17}$$

Con esta factorización, solo resta aplicar los lemas 1.1.6 y el ejemplo antes dado, a raíz de los núcleos de tipo Taylor, del núcleo $k(z, z') := \exp(\langle z, z'\rangle)$, para concluir con $k_{\gamma,\mathbb{R}^d}$ es núcleo. Además, el último aserto sobre su restricción a $\mathbb{R}^d \times \mathbb{R}^d$ es consecuencia inmediata del lema 1.1.3. □

El siguiente resultado concierne al llamado *núcleo binomial*, cuya condición de núcleo se sigue directamente de las propiedades fundamentales.

Proposición 1.1.10. *Sea $X := \mathring{B}_{\mathbb{R}^d}$ la bola unidad abierta en $\mathbb{R}^d$ bajo la norma usual $||\cdot||_2$ en $\mathbb{R}^d$, y sea $\alpha > 0$. Entonces, la función $k(x, x') := (1 - \langle x, x'\rangle)^{-\alpha}$ es un núcleo en $\mathbb{R}^d$ llamado* ***núcleo binomial****.*

Demostración. Consideremos el desarrollo general de un binomio de exponente $-\alpha$, con el α del enunciado:

$$(1 - z)^{-\alpha} = \sum_{n=0}^{\infty} \binom{-\alpha}{n}(-1)^n z^n, \tag{1.1.18}$$

lo cual es cierto para todo z contenido en la bola $\mathring{B}_{\mathbb{R}^d}$, donde se usa la expresión que generaliza un número combinatorio para un número real $-\alpha$, de la forma

$$\binom{-\alpha}{n} := \prod_{i=1}^{n} \frac{(-\alpha - i + 1)}{i}.$$

Notemos que $\binom{-\alpha}{n}(-1)^n > 0$ para cualquier $n \in \mathbb{N}$, ya que, si n es par, tendremos un número par de valores negativos procedentes de $\binom{-\alpha}{n}$, y viceversa. Por esto, aplicando el lema 1.1.8, tenemos que este k es un núcleo en $\mathbb{R}^d$. □

1.2. Funciones semidefinidas positivas y simétricas con valores reales

Hasta ahora, tan solo hemos dado algunos resultados útiles para la construcción de núcleos no triviales. A continuación, definimos lo que nosotros entenderemos por función semidefinida positiva, y ligaremos este concepto con la idea de núcleo. Veremos que, con gran generalidad, y limitándonos al caso de núcleos con valores reales, ambos conceptos son equivalentes.

Definición 1.2.1. Sea X un conjunto no vacío, y sea $k : X \times X \to \mathbb{R}$ una función tal que

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j k(x_i, x_j) \geq 0 \tag{1.2.1}$$

para cualesquiera $n \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, y $x_1, \ldots, x_n \in X$. Entonces, decimos que k es **semidefinida positiva**. Si, para $x_1, \ldots, x_n \in X$ mutuamente distintos, la igualdad en (1.2.1) solo es cierta para $\alpha_1 = \cdots = \alpha_n = 0$, diremos que k es **definida positiva**. Además, k será **simétrica** si $k(x, x') = k(x', x)$ para todo $x, x' \in X$.

Es inmediato ver que, si k es un núcleo con valores reales, siempre será simétrico, por la definición de núcleo. Además, llamando $\phi : X \to H$ a la aplicación característica de k,

$$\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \alpha_k \alpha_m k(x_k, x_m) = \langle \sum_{k=1}^{\infty} \alpha_k \phi(x_k), \sum_{m=1}^{\infty} \alpha_m \phi(x_m) \rangle_H \geq 0 \tag{1.2.2}$$

para todo $x_1, \ldots, x_n \in X$, $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. En resumidas cuentas, si k toma valores en $\mathbb{R}$, siempre será simétrico y semidefinido positivo.

Definición 1.2.2. Sea k un núcleo en X, sean $x_1, \ldots, x_n \in X$, y sea $\mathcal{M}_n(\mathbb{K})$ el conjunto de matrices cuadradas $n \times n$ con entradas en $\mathbb{K}$. La matriz $K \in \mathcal{M}_n(\mathbb{K})$ dada por $K := (k(x_i, x_j))_{i,j \in \{1,\ldots,n\}}$ es la **matriz de Gram del núcleo** k **para** $\{x_1, \ldots, x_n\} \subset X$.

Ahora, daremos un teorema y un corolario claves que relacionan las funciones reales semidefinidas positivas dadas en (1.2.1) y los núcleos con valores reales de una forma biunívoca.

Teorema 1.2.3. *Sea X no vacío. Cualquier función que tome valores reales de la forma $k : X \times X \to \mathbb{R}$ es un núcleo si y solo si es simétrica y semidefinida positiva.*

Demostración. Ya hemos probado que si k es un núcleo con valores reales, entonces tiene que ser simétrica y semidefinida positiva. Veamos el recíproco.

Consideramos una función simétrica y semidefnida positiva $k : X \times X \to \mathbb{R}$, y definimos el siguiente conjunto:

$$H_{\text{pre}} := \{\sum_{k=1}^{n} \alpha_k k(\cdot, x_k) : n \in \mathbb{N}, \alpha_k \in \mathbb{R}, x_k \in X, 1 \leq k \leq n\},$$

de forma que, tomando $f, g \in H_{pre}$, escribimos $f = \sum_{k=1}^{n} \alpha_k k(\cdot, x_k)$, $g = \sum_{j=1}^{m} \beta_j k(\cdot, x'_j)$, con $n, m \in \mathbb{N}$, para ciertos $x_1, \ldots, x_n, x'_1, \ldots, x'_m \in X$, y $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m \in \mathbb{R}$. Definimos el siguiente producto interno en H_{pre}

$$\langle f, g \rangle := \sum_{k=1}^{n} \sum_{j=1}^{m} \alpha_k \beta_j k(x'_j, x_k), \tag{1.2.3}$$

el cual no depende de la representación que se dé a f o a g, pues

$$\langle f, g \rangle = \sum_{k=1}^{n} \alpha_k (\sum_{j=1}^{m} \beta_j k(x'_j, x_k)) = \sum_{k=1}^{n} \alpha_k g(x_k),$$

donde, en la última igualdad, se ha utilizado que k es simétrica. Como se ve, la expresión anterior no depende de la representación para g. Análogamente,

$$\langle f, g \rangle = \sum_{j=1}^{m} \beta_j (\sum_{k=1}^{n} \alpha_k k(x'_j, x_k)) = \sum_{j=1}^{m} \beta_j f(x'_j).$$

La simetría y la bilinealidad de (1.2.3) son inmediatas de la forma de k (recuérdese la definición del conjunto H_{pre}). Sea el $f \in H_{\text{pre}}$ arbitrario de antes, entonces

$$\langle f, f\rangle = \sum_{k=1}^{n}\sum_{j=1}^{m}\alpha_k\alpha_j k(x_j, x_k) \geq 0, \qquad f \in H_{\mathrm{pre}},$$

donde hemos usado que, por hipótesis, k es semidefinida positiva.
Por la desigualdad de Cauchy-Schwarz sobre el semiproducto escalar que tenemos definido en H_{pre}, tenemos que, para todos $f, g \in H_{pre}$

$$|\langle f, g\rangle|^2 \leq \langle f, f\rangle \cdot \langle g, g\rangle. \tag{1.2.4}$$

Entonces, tomando la función $f(x) := \sum_{i=1}^{n}\alpha_i k(x, x_i) \in H_{pre}$ con $\langle f, f\rangle = 0$, es inmediato probar, usando (1.2.4), que $f = 0$, ya que, aplicando el semiproducto escalar de H_{pre}

$$|f(x)|^2 = |\sum_{i=1}^{n}\alpha_i k(x, x_i)|^2 = |\langle f, k(\cdot, x)\rangle|^2 \leq \langle k(\cdot, x), k(\cdot, x)\rangle \cdot \langle f, f\rangle = 0. \tag{1.2.5}$$

Es decir, $f(x) = 0$ para todo $x \in X$. Esto demuestra que $\langle \cdot, \cdot\rangle$ es un producto escalar en H_{pre}.
Consideramos ahora la compleción de H_{pre} en un espacio de Hilbert H a través de la inmersión isométrica $I : H_{pre} \to H$ [4]. Dada la condición de isometría de I se tiene, para todos $x, x' \in X$, lo siguiente

$$\langle Ik(\cdot, x'), Ik(\cdot, x)\rangle_H = \langle k(\cdot, x'), k(\cdot, x)\rangle_{H_{pre}} = k(x, x'). \tag{1.2.6}$$

En la última igualdad, se ha utilizado la definición del producto escalar dado en H_{pre}. De esta forma, para todo $x \in X$,

$$x \mapsto Ik(\cdot, x) \tag{1.2.7}$$

nos da la aplicación característica buscada en el espacio de Hilbert H. En otras palabras, acabamos de probar que k es un núcleo en X, con la aplicación característica (1.2.7). □

Como corolario, daremos un resultado que confirma la regularidad por límites en la topología, que caracterizaremos con la correspondiente norma, definida en el espacio de Hilbert característico.

Corolario 1.2.4. *Sea $(k_n)_{n=1}^{\infty}$ una sucesión de núcleos reales en $X \neq \emptyset$ que converge puntualmente a una función $k : X \times X \to \mathbb{R}$, esto es, tal que para todos $x, x' \in X$*

$$\lim_{n\to\infty} k_n(x, x') = k(x, x'). \tag{1.2.8}$$

Entonces, k es un núcleo que toma valores reales.

Demostración. Para todo $n \in \mathbb{N}$, k_n es simétrica y semidefinida positiva, pues es un núcleo (teorema 1.2.3)). Y, como estas propiedades se mantienen por límites puntuales, k es también simétrica y semidefinida positiva con valores en $\mathbb{R}$. Entonces, k es un núcleo con valores reales. □

1.3. Correspondencia entre núcleos y RKHS

Tratamos ahora la relación de correspondencia biunívoca entre los núcleos ya definidos y los espacio de Hilbert con núcleo reproductor, sobre los que aún no hemos dicho nada. Los fundamentos de este apartado son imprescindibles para todo el desarrollo que sigue. Volvemos al caso de funciones con llegada en un cuerpo $\mathbb{K}$ que no es necesariamente $\mathbb{R}$. Antes de dar las primeras definiciones, la

[4] La inmersión isométrica usada es la expuesta en el teorema A.0.1 del apéndice, donde enunciamos la exitencia de una inyección isométrica $I : \hat{H} \to H$, con $\hat{H}$ un espacio con producto escalar arbitrario de partida, y H su espacio de Hilbert completado, de modo que $I(\hat{H})$ es denso en H en la topología fuerte definida en H.

terminología merece una aclaración. Hasta ahora, hemos hablado de núcleos como funciones a las que únicamente pedimos que se reproduzcan a través del producto interno, en un espacio de Hilbert, de imágenes de una aplicación característica. A continuación, empezaremos hablando de núcleos reproductores en espacios de Hilbert sin saber, de antemano, si son núcleos. Se pide al lector que interprete este nombre, provisionalmente, como mera terminología. En el transcurso de este apartado se verá resuelto este problema: hablar de núcleos y de núcleos reproductores es equivalente.

Definición 1.3.1. Sea $X \neq \emptyset$ y H un $\mathbb{K}$-espacio de Hilbert de funciones de la forma $f : X \to \mathbb{K}$. Entonces

i) La función $k : X \times X \to \mathbb{K}$ es un **núcleo reproductor** si $k(\cdot, x) \in H$ para todo $x \in X$, y además, la relación

$$f(x) = \langle f, k(\cdot, x)\rangle, \tag{1.3.1}$$

se cumple para todo $x \in X$ y para todo $f \in H$. A esta propiedad (1.3.1) se le llama **propiedad reproductora** del núcleo k.

ii) H será un **espacio de Hilbert con núcleo reproductor** (**RKHS**) en X si, para todo $x \in X$, el funcional delta de Dirac $\delta_x : H \to \mathbb{K}$, dado por $\delta_x(f) = f(x)$ para todo $f \in H$, es continuo.

En resumen, un RKHS es un subconjunto de $F(X)$, donde $F(X)$ es el conjunto de todas las funciones de X en $\mathbb{K}$, con X un conjunto no vacío arbitrario, y $\mathbb{K}$ el cuerpo sobre el que trabajamos (siempre $\mathbb{R}$ o $\mathbb{C}$), en el cual, para todo $x \in X$, existe una función $k_x : X \to H$, de la forma $k_x := k(\cdot, x) \in H$, que satisface $f(p) = \delta_p(f) = \langle f, k(\cdot, p)\rangle$, para todo $f \in H$ y todo $p \in X$. Esto es consecuencia del teorema de representación de Riesz-Fréchet, aplicado sobre los funcionales continuos δ_p.
Notemos algo importante. Los espacios $L^p(\mathbb{K}^d)$ no están formados por funciones, sino por clases de equivalencia de funciones determinadas por igualdades casi siempre en espacios de medida conocidos. Por tanto, estos espacios no pueden ser RKHS, tal y como hemos definido dichos espacios de Hilbert. No obstante, sí que hablaremos de RKHSs a partir de *ciertos* representantes de las clases de equivalencia que forman dichos espacios $L^p(\mathbb{K}^d)$, como veremos para los espacios de Paley-Wiener o de Hardy. En estos, los representantes de las clases de estos L^p que cumplen alguna buena propiedad, son precisamente los elementos que constituyen dicho RKHS.
Respecto al funcional de Dirac δ_x observemos que, con toda generalidad para nosotros, éste es un funcional definido sobre un espacio de Hilbert. Para ver la linealidad, basta tomar $f, g \in H$ y darse cuenta de que $\delta_x(\alpha f + \beta g) = (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) = \alpha\delta_x(f) + \beta\delta_x(g)$. La continuidad o no continuidad de dichos funcionales ya dependerá de si H satisface la condición reproductora de la definición de RKHS.
De la definición de la *propiedad reproductora* tenemos una característica común a todos los RKHS que es de especial importancia.

Lema 1.3.2. *Sea H un RKHS. La convergencia en norma en H implica la convergencia puntual de los correspondientes elementos de H. Dicho de otra forma, dada una sucesión $(f_n)_{n=1}^{\infty} \subset H$, si $\lim_{n\to\infty} ||f_n - f||_H = 0$, entonces $\lim_{n\to\infty} f_n(x) = f(x)$ para todo $x \in X$, es decir, $f_n \to f$ puntualmente.*

Demostración. Basta observar que

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \delta_x(f_n) = \delta_x(\lim_{n\to\infty} f_n) = \delta_x(f) = f(x), \ \ x \in X, \tag{1.3.2}$$

donde hemos usado, en la segunda igualdad, la continuidad del funcional de Dirac, para poder así utilizar la continuidad secuencial, por la cual la convergencia en H implica la convergencia de las imágenes en $\mathbb{K}$. □

A continuación, damos un lema y dos teoremas que nos relacionarán propiamente los RKHS y los núcleos reproductores, viendo cómo, sin salir del marco del análisis funcional clásico, podemos establecer una biyección entre el conjunto de núcleos reproductores y el conjunto de espacios de Hilbert con núcleo reproductor.
Observemos que, si k es un núcleo reproductor, parece fácil probar que la aplicación $\phi(x) = k(\cdot, x)$ es la aplicación característica de k sobre el espacio de Hilbert en que k cumple la propiedad reproductora.

Lema 1.3.3 (Un núcleo reproductor es un núcleo). *Sea H un espacio de Hilbert de funciones en X con núcleo reproductor $k : X \times X \to \mathbb{K}$. Entonces, H será un RKHS. Además, si definimos $\phi : X \to H$ como $\phi(x) = k(\cdot, x)$ para todo $x \in X$, entonces k será un núcleo con aplicación característica ϕ y espacio característico H. A la aplicación ϕ se le suele llamar* ***aplicación característica canónica****.*

Demostración. Por hipótesis, k satisface la propiedad reproductora (1.3.1), luego dado $f \in H$, el funcional de Dirac cumple, para todo $x \in X$, que

$$|\delta_x(f)| = |f(x)| = |\langle f, k(\cdot, x)\rangle| \leq ||k(\cdot, x)||_H ||f||_H, \tag{1.3.3}$$

donde hemos usado la propiedad reproductora (1.3.1) en la segunda igualdad, y la desigualdad de Cauchy-Schwarz en la tercera relación. Con esto, para todo $x \in X$, acabamos de probar que el funcional de Dirac $\delta_x : H \to \mathbb{K}$ es acotado, lo que equivale a decir que es continuo en H. Entonces, H es un RKHS.

En cuanto a que k sea núcleo, basta tomar $f := k(\cdot, x')$ para un $x' \in X$ fijo, usar la propiedad reproductora de k, y recordar la forma que tiene la aplicación ϕ del enunciado. Con todo ello, para cada $x \in X$ obtenemos

$$\langle \phi(x'), \phi(x)\rangle = \langle k(\cdot, x'), k(\cdot, x)\rangle = \langle f, k(\cdot, x)\rangle) = f(x) = k(x, x'). \tag{1.3.4}$$

De esta manera, acabamos de probar que k cumple la definición de núcleo con la aplicación caracterísitca ϕ y el espacio característico H. □

Damos ahora dos teoremas de gran relevancia que nos permitirá explotar en adelante la biyección entre los RKHS y los núcleos reproductores. Para ello, tenga presente el lector el comentario sobre notación hecho después del enunciado del teorema de representación de Riesz-Fréchet A.0.12 del primer apéndice, relativo al sentido que tienen los productos internos que vamos a dar en el dual topológico H' del espacio de Hilbert H de partida. Recordemos que la definición del dual topológico de un espacio de Hilbert H sería la siguiente: es el conjunto de funcionales lineales y acotados $f : H \to \mathbb{K}$, con $\mathbb{K}$ el cuerpo sobre el que se construye el espacio.

Teorema 1.3.4 (Todo RKHS admite un único núcleo reproductor). *Sea H un RKHS en X, con X no vacío. Entonces, la función $k : X \times X \to \mathbb{K}$ dada por*

$$k(x, x') = \langle \delta_x, \delta_{x'}\rangle_{H'}, \qquad x, x' \in X, \tag{1.3.5}$$

es el único núcleo reproductor de H. Además, si $(e_i)_{i\in I}$ es una base ortonormal de H, entonces, para todos $x, x' \in X$, tenemos

$$k(x, x') = \sum_{i \in I} e_i(x)\overline{e_i(x')}. \tag{1.3.6}$$

Demostración. Necesitamos ver, en primer lugar, que k es efectivamente un núcleo reproductor. Recurriendo al teorema de representación de Riesz-Fréchet A.0.12, aplicado sobre H y su dual H', consideramos la isometría sobreyectiva $I : H' \to H$ (de hecho, es una biyección lineal conjugada) que asigna a cada elemento $g \in H'$ el elemento *representante* $Ig \in H$, de forma que, para todo $f \in H$ y

$g \in H'$, $g(f) = \langle f, Ig \rangle$. Por tanto, aplicando este teorema de representación sobre δ_x y $\delta_{x'}$, ambas en H' por ser H RKHS, para todos $x, x' \in X$ se tiene

$$k(x, x') = \langle \delta_x, \delta_{x'} \rangle_{H'} = \langle I\delta_x, I\delta_{x'} \rangle_H = \delta_x(I\delta_{x'}) = I\delta_{x'}(x). \tag{1.3.7}$$

Por tanto, se tiene $k(\cdot, x') = I\delta_{x'}$ para todo $x' \in X$. Es inmediato, de lo anterior, que

$$f(x') = \delta_{x'}(f) = \langle f, I\delta_{x'} \rangle = \langle f, k(\cdot, x') \rangle, \tag{1.3.8}$$

lo que demuestra que k, tal y como lo hemos definido, cumple la propiedad reproductora. Veamos que es el único. Supongamos que $\hat{k}$ es un núcleo reproductor arbitrario de H. Consideramos una base ortonormal $(e_i)_{i\in I}$ de H. Como $\hat{k}(\cdot, x') \in H$ para todo $x' \in X$, podemos expresar este elemento con la base mencionada de la forma que sigue

$$\hat{k}(\cdot, x') = \sum_{i\in I} \langle \hat{k}(\cdot, x'), e_i \rangle e_i = \sum_{i\in I} \overline{e_i(x')} e_i, \tag{1.3.9}$$

donde, en la última igualdad, hemos usado la propiedad reproductora de $\hat{k}$. La convergencia de la suma se da respecto a la norma $||\cdot||_H$. Ahora bien, como H es un RKHS, el lema 1.3.2 garantiza la convergencia puntual, es decir, para cada $x \in X$,

$$\hat{k}(x, x') = \sum_{i\in I} \overline{e_i(x')} e_i(x). \tag{1.3.10}$$

Esto demuestra la igualdad (1.3.6), y como $\hat{k}$ y la base $(e_i)_{i\in I}$ fueron arbitrariamente elegidos, es claro que $\hat{k} = k$. Por lo tanto, para rematar la demostración, k es el único núcleo reproductor del RKHS H. □

En la práctica, este resultado es clave a la hora de calcular núcleos reproductores de RKHSs particulares, ya que basta conocer una base ortonormal del espacio para que, extendiéndolo en una suma sobre la familia de etiquetas I, tener dicho núcleo k.
Veamos ahora la implicación contraria, partiendo de un núcleo arbitrario, para llegar a un RKHS determinado. Antes de meternos con el enunciado y la demostración de este resultado, conocido como el teorema de Moore-Aronszajn, definimos lo que entendemos por sobreyección métrica.

Definición 1.3.5. Dados H y K dos espacios de Hilbert, una *sobreyección métrica* de H en K es una aplicación sobreyectiva $V : H \to K$ tal que $V(\mathring{B}_H) = \mathring{B}_K$, con $\mathring{B}_H$ y $\mathring{B}_K$ las bolas unidad abiertas en H y K, respectivamente.

Teorema 1.3.6 (Todo núcleo admite un único RKHS. **Teorema de Moore-Aronszajn**). *Sea X no vacío. Sea k un núcleo en X con espacio característico F sobre un cuerpo $\mathbb{K}$, y con una aplicación característica $\psi : X \to F$. Entonces, el espacio de funciones*

$$H := \{f : X \to \mathbb{K} : \exists w \in F \text{ con } f(x) = \langle w, \psi(x) \rangle_F \text{ para todo } x \in X\}, \tag{1.3.11}$$

junto con la norma

$$||f||_H := \inf\{||w||_F : w \in F \text{ con } f = \langle w, \psi(\cdot) \rangle_F\}, \tag{1.3.12}$$

forma el único RKHS para el cual k es núcleo reproductor, de forma que tanto (1.3.11) *como* (1.3.12) *son independientes de la elección arbitraria de F y ψ.*
Además, la aplicación entre espacios de Hilbert $A : F \to H$ dada por $Aw := \langle w, \psi(\cdot) \rangle_F$ para todo $w \in F$ es una sobreyección métrica.
Por último, el conjunto dado como

$$H_{pre} := \{\sum_{i=1}^{n} \alpha_i k(\cdot, x_i) : n \in \mathbb{N}, \alpha_1, \ldots, \alpha_n \in \mathbb{K}, x_1, \ldots, x_n \in X\}, \tag{1.3.13}$$

es denso en H, y dado $f = \sum_{i=1}^n \alpha_i k(\cdot, x_i) \in H_{pre}$, tenemos

$$||f||_H^2 = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_j, x_i). \tag{1.3.14}$$

Diremos que H es la ***compleción funcional*** *del espacio H_{pre} (no solo completamos el espacio para dar un espacio de Hilbert, sino que este espacio de Hilbert que lo completa posee un núcleo reproductor).*

Demostración. En primer lugar, probaremos que el espacio H definido es un espacio de Hilbert de funciones de X en el cuerpo $\mathbb{K}$. Para ello, notemos que A es sobreyectiva, pues para todo $f \in H$, por definición de H, existe $w \in F$ tal que $A(w) = f$, de forma que $f(\cdot) = \langle w, \psi(\cdot)\rangle_F$. Además, H es un espacio vectorial de funciones de X en $\mathbb{K}$, en virtud de la sesquilinealidad del producto interno del espacio de Hilbert original F. Dicho esto, podemos simplificar la expresión de lo propuesto como norma en H con la igualdad

$$||f||_H = \inf_{w \in A^{-1}(\{f\})} ||w||_F. \tag{1.3.15}$$

Nótese que no se ha dicho nada de la inyectividad o no inyectividad de A, con lo que la contraimagen $A^{-1}(\{f\})$ puede estar formada por más de un elemento de F.
Probamos la linealidad de la aplicación A. Sean $a, b \in \mathbb{K}$, $w_1, w_2 \in F$. Tomamos dos elementos de H, $f(\cdot) = \langle w_1, \psi(\cdot)\rangle_F$ y $g(\cdot) = \langle w_2, \psi(\cdot)\rangle_F$, que satisfacen $A(w_1) = f$ y $A(w_2) = g$. Entonces, por la sesquilinealidad de $\langle \cdot, \cdot \rangle_F$,

$$A(aw_1 + bw_2) = \langle aw_1 + bw_2, \psi(\cdot)\rangle_F = a\langle w_1, \psi(\cdot)\rangle_F + b\langle w_2, \psi(\cdot)\rangle_F = aA(w_1) + bA(w_2), \tag{1.3.16}$$

lo que prueba la linealidad de A.
Veamos que $||\cdot||_H$ es norma en H. La propiedad $||f||_H \geq 0$ para todo $f \in H$ se sigue de la definición de $||\cdot||_H$. Probamos las otras tres propiedades:

1. Sea $\lambda \in \mathbb{K}$, $f \in H$. ¿$||\lambda f||_H = |\lambda| ||f||_H$?
   Si $\lambda = 0$, $\lambda f = 0$. Es claro que $||\lambda f||_H = ||0||_H = 0$, pues $||\cdot||_H$ es el ínfimo de los $||w||_F \in \mathbb{R}^+$ tales que $A(w) = 0$, y por linealidad de A, $w = 0$ es uno de tales elementos w en F. Por otra parte, es claro que $|\lambda| ||f||_H = 0$, lo que prueba el caso de $\lambda = 0$. Supongamos $\lambda \neq 0$. Queremos probar la igualdad de los dos siguientes objetos:

   $$\begin{aligned} ||\lambda f||_H &= \inf_{w \in A^{-1}(\{\lambda f\})} ||w||_F, \\ |\lambda| \cdot ||f||_H &= |\lambda| \cdot \inf_{w' \in A^{-1}(\{f\})} ||w'||_F. \end{aligned} \tag{1.3.17}$$

   Denotamos $B := A^{-1}(\{\lambda f\})$ y $C = A^{-1}(\{f\})$. Probemos que $B = \lambda C$. Para la primera contención, tomemos $w \in B$. Entonces, $A(w) = \lambda f$, luego

   $$A(\frac{1}{\lambda} w) = \frac{1}{\lambda} A(w) = f, \tag{1.3.18}$$

   es decir, $w = \lambda(\frac{1}{\lambda} w)$ con $\frac{1}{\lambda} w \in C$. De esto se sigue que $w \in \lambda C$. Para la contención contraria, tomemos $w' \in \lambda C$. Es decir, $w' = \lambda w$ con $w \in C$. Entonces, $A(w') = \lambda A(w) = \lambda f$, luego $w' \in B$.
   Como $B = \lambda C$, se tiene la igualdad

   $$\inf_{w \in B} ||w||_F = \inf_{w' \in \lambda C} ||w'||_F. \tag{1.3.19}$$

   Recordemos que, dados $D \subset \mathbb{R}^+$ y $\alpha \geq 0$, se satisface $\inf(\alpha D) = \alpha(\inf D)$. Por ello,

   $$\inf_{w' \in \lambda C} ||w'||_F = \inf_{w'' \in C} |\lambda| \cdot ||w''||_F = |\lambda| \cdot \inf_{w'' \in C} ||w''||_F. \tag{1.3.20}$$

   Por lo tanto, reescribiendo B y C en términos de A,

   $$\inf_{w \in A^{-1}(\{\lambda f\})} ||w||_F = |\lambda| \cdot \inf_{w \in A^{-1}(\{f\})} ||w||_F \implies ||\lambda f||_H = |\lambda| \cdot ||f||_H. \tag{1.3.21}$$

2. Dado $f \in H$ tal que $||f||_H = 0$, ¿$f \equiv 0$?
   Se tiene que $\inf_{w \in A^{-1}(\{f\})} ||w||_F = 0$. Para todo $n \in \mathbb{N}$, existe un $w_n \in A^{-1}(\{f\})$ tal que $Aw_n = f$ y
$$||w_n||_F < \frac{1}{n}. \tag{1.3.22}$$
   Para un $n \in \mathbb{N}$ arbitrario, como $Aw_n = f$, para todo $x \in X$ tenemos que
$$f(x) = \langle w_n, \psi(x) \rangle_F. \tag{1.3.23}$$
   Entonces, por la desigualdad de Cauchy-Schwarz,
$$|f(x)| \leq ||w_n||_F \cdot ||\psi(x)||_F < \frac{1}{n} ||\psi(x)||_F, \tag{1.3.24}$$
   donde el último término tiende a cero para $n \to \infty$. Como esta convergencia a 0 no depende del x escogido, tenemos que $f \equiv 0$.

3. Dados $f, g \in H$, ¿se satisface la desigualdad triangular $||f + g||_H \leq ||f||_H + ||g||_H$?

   Recordemos que $||f||_H = \text{ínf}_{w \in A^{-1}(\{f\})} ||w||_F$. Sea $\varepsilon > 0$. Sabemos que existe $w_1 \in A^{-1}(\{f\})$ tal que
$$||w_1||_F < ||f||_H + \frac{\varepsilon}{2}. \tag{1.3.25}$$
   Análogamente, existe $w_2 \in A^{-1}(\{g\})$ tal que
$$||w_2||_F < ||g||_H + \frac{\varepsilon}{2}. \tag{1.3.26}$$
   Como $A(w_1 + w_2) = A(w_1) + A(w_2) = f + g$, se sigue que
$$||f+g||_H = \underset{w \in A^{-1}(\{f+g\})}{\text{ínf}} ||w||_F \leq ||w_1 + w_2||_F \leq ||w_1||_F + ||w_2||_F < ||f||_H + ||g||_H + \varepsilon. \tag{1.3.27}$$
   Como $\varepsilon > 0$ es arbitrario, se deduce que $||f + g||_H \leq ||f||_H + ||g||_H$, con lo que se obtiene la desigualdad triangular para $|| \cdot ||_H$.

Veamos ahora que H es una espacio de Hilbert con la norma $|| \cdot ||_H$. Para ello, veremos que el núcleo de A es cerrado, lo que nos permitiría escindir el espacio H en dicho núcleo y su complementario ortogonal. Una vez probemos eso, restringiremos A a dicho complementario, para así dar una definición alternativa de la norma de H que nos muestre con claridad que, efectivamente, H es de Hilbert. Pues bien, para ver que $\ker A$ es cerrado, tomemos $(w_n) \subset \ker A := \{w \in F : A(w) = 0\}$ una sucesión convergente a un elemento $w \in F$, con el objetivo de probar que w están en $\ker A$. Por la continuidad de la aplicación $O : F \to \mathbb{K}$ tal que $O(\tilde{w}) = \langle \tilde{w}, \psi(x) \rangle$ para todo $\tilde{w} \in F$, tenemos que $\langle w, \psi(x) \rangle = \text{lím}_{n \to \infty} \langle w_n, \psi(x) \rangle = 0$ para todo $x \in X$. Entonces, por definición de A, $A(w) = 0$, luego $w \in \ker A$, con lo que $\ker A$ es un subespacio cerrado de F. Así pues, como ya mencionamos previamente, podemos tomar el complementario ortogonal del núcleo en F, al que llamaremos $\hat{H}$, de forma que podemos descomponer el espacio F como $F = \hat{H} \oplus \ker A$.

Hecho esto, pasamos a considerar la restricción de A al complementario ortogonal del núcleo $\ker A$, de la forma $A|_{\hat{H}} : \hat{H} \to H$, la cual es inyectiva por la construcción realizada. Probamos ahora que $A|_{\hat{H}} : \hat{H} \to H$ también es sobreyectiva, y que, además, es una isometría. Con eso y con la igualdad que vamos a dar, será inmediato ver que H es un espacio de Hibert con la norma definida. Sea pues $f \in H$, y sea $w \in F$ el elemento tal que $f(\cdot) = \langle w, \psi(\cdot) \rangle$. Descomponemos w en la única suma $w = w_0 + \hat{w}$ de forma que $w_0 \in \ker A$ y $\hat{w} \in \hat{H}$. Entonces, $f = A(w_0 + \hat{w}) = A(\hat{w}) = A|_{\hat{H}}(\hat{w})$, lo que nos dice que, como buscábamos, $A|_{\hat{H}}$ es sobreyectiva. Además, por definición de la norma dada en H, tenemos que

$$||f||_H^2 = \inf_{\substack{w_0 \in kerA,\ \hat{w} \in \hat{H} \\ w_0 + \hat{w} \in A^{-1}(\{f\})}} ||w_0 + \hat{w}||_F^2 = \inf_{\substack{w_0 \in kerA,\ \hat{w} \in \hat{H} \\ w_0 + \hat{w} \in A^{-1}(\{f\})}} ||w_0||_F^2 + ||\hat{w}||_F^2, \tag{1.3.28}$$

donde, por definición de la norma dada en H y dado que $w_0 \in \ker A$, el término de $||w_0||_F$ en esta expresión del ínfimo de una suma será necesariamente cero (nótese que $0 \in \ker A$). También, como $A|_{\hat{H}}$ es una biyección sabemos que, para $\hat{w}$, el elemento $f \in H$ será el único tal que $A|_{\hat{H}}(\hat{w}) = f$, luego restringiendo la norma de F a $\hat{H}$, $||\hat{w}||_F = ||\hat{w}||_{\hat{H}} = ||A^{-1}|_{\hat{H}} f||_{\hat{H}}$. Además, en la segunda igualdad de (1.3.28) hemos usado la ortonormalidad de los elementos $w_0, \hat{w} \in F$. Por tanto, acabamos de probar que $||f||_H = ||A^{-1}|_{\hat{H}} f||_{\hat{H}}$, esto es, que $A|_{\hat{H}}$ es un isomorfismo isométrico de $\hat{H}$ en H. Así pues, como $\hat{H}$ es el complementario ortogonal de un subespacio de F, entonces $\hat{H}$ es cerrado en F, y como F es un espacio de Hilbert, entonces $\hat{H}$ es un espacio de Hilbert. Por tanto, como $\hat{H}$ es de Hilbert, y $A|_{\hat{H}} : \hat{H} \to H$ es un isomorfismo isométrico, entonces H es un espacio de Hilbert.
De hecho, si cogemos la bola unidad abierta $B_1 = \mathring{B}_F$, para todo $w \in B_1$ se tiene que el elemento $f = A(w) \in H$ satisface $||f||_H < 1$, dada la definición de A y de $|| \cdot ||_H$, con lo que $AB_1 \subset B_2$, donde $B_2 = \mathring{B}_H$ es la bola unidad abierta de H. Esto demuestra, como también habíamos enunciado, que A es una sobreyección métrica.
A continuación, veamos que k es el núcleo reproductor de H, de lo que se deriva fácilmente que H es un RKHS, con el núcleo k como núcleo reproductor. Por la definición más primitiva que hemos dado de núcleo en un conjunto no vacío X dada su aplicación característica ψ, tenemos que, para todo $x \in X$, se cumple la igualdad $k(\cdot, x) = \langle \psi(\cdot), \psi(x) \rangle$, lo cual satisface estrictamente la definición de los elementos de H, porque $\psi(x) \in F$ para todo $x \in X$. Entonces, $k(\cdot, x) \in H$ para todo $x \in X$. Dada la construcción de la sobreyección métrica A, es inmediato que $\langle w, \psi(\cdot) \rangle = 0$ para cualquier $w \in \ker A$, por lo que $\psi(x) \in \ker A^\perp = \hat{H}$. Por lo tanto, podemos escribir $k(\cdot, x) = A|_{\hat{H}}(\psi(x))$, para todo $x \in X$, con el sentido de que $\psi(x) \in F$. Así pues, recodando que $A|_{\hat{H}}$ es una isometría biyectiva y usando la definición del espacio H, para todo $x \in X$ se tiene que

$$f(x) = \langle (A|_{\hat{H}})^{-1} f, \psi(x) \rangle_F = \langle f, A|_{\hat{H}} \psi(x) \rangle_H = \langle f, k(\cdot, x) \rangle, \;\; x \in X, f \in H. \tag{1.3.29}$$

Luego k satisface la propiedad reproductora en el espacio H. Esto es, H es un espacio de Hilbert con núcleo reproductor k, luego, por el lema 1.3.3, acabamos de probar que H es un RKHS con núcleo reproductor k. Nótese que aún no hemos probado que este RKHS sea el único ligado con el núcleo reproductor k.
Veamos ahora los resultados enunciados sobre el espacio H_{pre}. Veamos, de hecho, que las afirmaciones hechas sobre este espacio son ciertas para cualquier RKHS $\tilde{H}$ con núcleo reproductor k. En primer lugar, es claro que H_{pre} tiene estructura de espacio vectorial con la suma y el producto por escalares dados en el espacio $\tilde{H}$. De hecho, como $k(\cdot, x) \in \tilde{H}$ para todo $x \in X$, por dicha estrucutra vectorial es claro que $H_{pre} \subset \tilde{H}$, esto es, que H_{pre} es subespacio vectorial de $\tilde{H}$. Para ver que H_{pre} es denso en $\tilde{H}$, razonamos por reducción al absurdo. Supongamos que H_{pre} no es denso en H. Por el corolario A.0.10 del apéndice, como H_{pre} es subespacio vectorial de H, esto implica que $H_{pre}^\perp \neq \{0\}$. Entonces, debe existir $f \in H_{pre}^\perp$ para el que exista al menos un $x \in X$ con $f(x) \neq 0$. Usando la propiedad reproductora de k en $\tilde{H}$ y la condición de ortogonalidad sobre $k(\cdot, x) \in H_{pre}$, tenemos

$$0 = \langle f, k(\cdot, x) \rangle = f(x) \neq 0. \tag{1.3.30}$$

lo que es absurdo, por lo que H_{pre} es denso en $\tilde{H}$.
Para acabar con las afirmaciones sobre H_{pre}, sea ahora $f := \sum_{k=1}^n \alpha_k k(\cdot, x_k) \in H_{\mathrm{pre}}$ para $\alpha_k \in \mathbb{K}$, $x_k \in X$, $k = 1, \ldots, n$. La propiedad reproductora ya probada sobre k implica que la norma en $\tilde{H}$ de f toma la siguiente forma

$$||f||_{\tilde{H}}^2 = \sum_{k=1}^n \sum_{i=1}^n \alpha_k \overline{\alpha_i} \langle k(\cdot, x_k), k(\cdot, x_i) \rangle_{\tilde{H}} = \sum_{k=1}^n \sum_{i=1}^n \alpha_k \overline{\alpha_i} k(x_i, x_k), \tag{1.3.31}$$

lo que prueba (1.3.14), y con ello todo lo propuesto sobre el subespacio H_{pre}.
Para acabar con esta demostración, nos falta probar la unidad del RKHS asociado al núcleo k. Sean pues H_1 y H_2 dos RKHSs de k. Veamos que son iguales. Acabamos de probar que H_{pre} es denso en cualquier RKHS con núcleo reproductor asociado k, luego lo será en H_1 y en H_2. De hecho, por el razonamiento que acabamos de seguir, las norma en H_1 y H_2 de cualquier elemento $f \in H_{pre}$ coinciden. Por la mencionada densidad, dado $f \in H_1$, debe existir una sucesión $(f_n) \subset H_{pre}$ tal que converja en norma de H_1 a f, i.e., tal que $\lim_{n\to\infty} ||f_n - f||_{H_1} = 0$. Por otra parte, $H_{pre} \subset H_2$, luego $(f_n) \subset H_2$. Como acabamos de decir, las normas en H_1 y H_2 coinciden en H_{pre}, y como esta sucesión (f_n) converge en norma de H_1, entonces es de Cauchy en norma de H_1, y por la igualdad de normas en H_1 y H_2 en H_{pre}, entonces (f_n) es Cauchy también en H_2. Como H_2 es espacio de Hilbert, entonces existe $g \in H_2$ que es límite en norma de H_2 de dicha sucesión, i.e., $\lim_{n\to\infty} ||f_n - g||_{H_2} = 0$. Ya vimos en 1.3.2 que la convergencia en norma en un RKHS implica convergencia puntual, luego para todo $x \in X$

$$f(x) = \lim_{n\to\infty} f_n(x) = g(x). \tag{1.3.32}$$

Esto es, $f \equiv g$ en todo X. Por tanto, $f \in H_2$, y como f era un elemento arbitrario de H_1, entonces $H_1 \subset H_2$. Razonando análogamente se llega a la contención contraria, y además, si $\lim_{n\to\infty} ||f_n - f||_{H_1} = 0$ y $\lim_{n\to\infty} ||f_n - f||_{H_2} = 0$, entonces tenemos, recodando que $(f_n) \subset H_{pre}$:

$$||f||_{H_1} = \lim_{n\to\infty} ||f_n||_{H_1} = \lim_{n\to\infty} ||f_n||_{H_{pre}} = \lim_{n\to\infty} ||f_n||_{H_2} = ||f||_{H_2}, \tag{1.3.33}$$

luego H_1 y H_2 están contenidos isométricamente el uno en el otro. En otra palabras, $H_1 = H_2$ con iguales normas. Entonces, el RKHS asociado al núcleo k es único, lo cual era lo único que nos faltaba por probar. □

Por lo que acabamos de ver, dado un núcleo, existirá un solo RKHS asociado a él. Por tanto, por el teorema 1.3.4 cada RKHS tiene un único núcleo reproductor, que a su vez es núcleo; y este núcleo, por el teorema 1.3.6, admite un único RKHS. En otras palabras, acabamos de probar que existe una biyección entre el conjunto de RKHSs y el de núcleos en un conjunto arbitario no vacío X.
Además, podemos establecer también, lo cual nos será útil para la teoría que sigue y para la comprensión de estos espacios, una biyección entre los núcleos reproductores en un conjunto no vacío X, de $X \times X$ en $\mathbb{R}$, y las aplicaciones que van de X a $l^2(I)$, con $\mathbb{R}$ como cuerpo subyacente. En nuestras consideraciones también entra un conjunto arbitrario I.

Proposición 1.3.7. *Sea k un función que va de $X \times X$ en $\mathbb{R}$. Entonces, k es un núcleo reproductor en X, i.e., será una función semidefinida positiva y simétrica en $X \times X$, si y solo si existe al menos una aplicación T de X a algún espacio $l^2(I)$ tal que*

$$k(x, x') = \langle T(x), T(y)\rangle_{l^2(I)} = \sum_{i\in I} (T(x))_i (T(y))_i. \tag{1.3.34}$$

Demostración. Sea H el RKHS de funciones de X en $\mathbb{R}$ con núcleo reproductor k. Sea la aplicación $\psi_K : X \to H$ dada por $\psi_k(x) = k(\cdot, x)$ para todo $x \in X$, i.e., la aplicación característica canónica de k.
Como H es un espacio de Hilbert, H es isométrico a algún espacio $l^2(I)$ para algún conjunto I. Denotamos por $\varphi : H \to l^2(I)$ a una isometría entre H y $l^2(I)$. Entonces, la composición $T := \varphi \circ \psi_K$ satisface las condiciones que queremos, pues $T : X \to l^2(I)$, y además, para todo $x, x' \in X$, como k es núcleo reproductor y $k(\cdot, x) \in H$, se tiene que

$$\begin{aligned} k(x, x') &= \langle k(\cdot, x'), k(\cdot, x)\rangle_H = \langle \psi_K(x'), \psi_K(x)\rangle_H = \\ &= \langle \varphi \circ \psi_K(x'), \varphi \circ \psi_K(x)\rangle_{l^2(I)} = \langle T(x'), T(x)\rangle_{l^2(I)}, \end{aligned} \tag{1.3.35}$$

donde hemos usado, en la tercera igualdad, el hecho de que $\varphi : H \to l^2(I)$ es una isometría, así como la definición de T en la cuarta igualdad. Esto prueba una de las implicaciones.

Recíprocamente, consideremos que existe una aplicación $T : X \to l^2(I)$, para algún conjunto I, que satisface (1.3.34). Definimos la aplicación $k : X \times X \to \mathbb{K}$ dada por $k(x, x') := \langle T(x'), T(x)\rangle_{l^2(I)}$, la cual es semidefinida positiva y simétrica, por ser $\langle\cdot,\cdot\rangle_{l^2(I)}$ un producto escalar. Entonces, k es un núcleo, luego es núcleo reproductor. □

Ahora, aportamos una serie de ejemplos rápidos que muestran esta naturaleza reproductiva en numerosos espacios de Hilbert muy conocidos, antes de entrar con un estudio y unos ejemplos menos triviales.

(i) Cualquier espacio normado de dimensión finita H es isométricamente isomorfo a un $\mathbb{K}^n$, con $\mathbb{K}$ el cuerpo sobre el que se define el espacio. Como estos espacios $\mathbb{K}^n$ son de Hilbert, entonces H también. Además, si H está formado por funciones definidas en un conjunto X no vacío con valores en $\mathbb{K}$, entonces es un RKHS. La razón de esto es que el funcional delta de Dirac, $\delta_x : H \to \mathbb{K}$ con $x \in X$, es una transformación lineal de un espacio normado de dimensión finita (por lo tanto, de Hilbert) en un espacio normado. Esto implica que δ_x es continua, y con ello tenemos que H es RKHS.

(ii) Sea $X = \{x_1, \ldots, x_n\}$ un conjunto arbitrario de n elementos. Sea $A = (a_{ij})_{i,j=1,\ldots,n} \in \mathcal{M}_n^*(\mathbb{R})$ un elemento del conjunto de matrices cuadradas de orden n simétricas y definidas positivas (estos desarrollos se dan, para el cuerpo $\mathbb{C}$, con matrices hermíticas en el lugar de simétricas). Escribimos $B = (b_{ij})_{i,j=1,\ldots,n} = A^{-1}$ la inversa de A. Sea $H_A(X)$ el espacio de funciones con valores complejos en X, en el que definimos el producto interno

$$\langle f, g\rangle_{H_A(E)} = \sum_{i,j=1}^{n} f(x_i) b_{ij} \overline{g(x_j)} = \overline{(g(x_1), \ldots, g(x_n))} B (f(x_1), \ldots, f(x_n))^t. \tag{1.3.36}$$

Pues bien, este espacio $H_A(X)$ es un RKHS con núcleo $k(x_i, x_j) = a_{ij}$. Para probarlo, se observa que, para todo $f \in H_A(X)$

$$\langle f, k(\cdot, x_j)\rangle_{H_a(X)} = \sum_{i,k=1}^{n} f(x_i) b_{ki} \overline{k(x_k, x_j)} = \sum_{i,k=1}^{n} f(x_i) b_{ki} a_{jk}, \tag{1.3.37}$$

donde hemos usado la definición de la función k en la última igualdad, y que los valores que toma k son reales. Es inmediato darse cuenta de que el producto de los elementos de B y de A de la última expresión nos da la delta de Kronecker δ_{ij}, con la construcción seguida. Así pues, $H_A(X)$ es un espacio de Hilbert con núcleo reproductor k.

(iii) El espacio de Hilbert estándar $l^2(\mathbb{N})$ es un RKHS con núcleo reproductor dada por la conocida delta de Kronecker, i.e.,

$$k(m, n) = \delta_{m,n}, \tag{1.3.38}$$

para cualesquiera $m, n \in \mathbb{N}$. Nótese que, dado un elemento $x = \{x(n)\}_{n=1}^{\infty} \in l^2(\mathbb{N})$, la propiedad de núcleo reproductor se visualiza fácilmente con la igualdad:

$$x(m) = \sum_{n=1}^{\infty} x(n)\delta_{m,n} = \langle x, \delta_{m,(\cdot)}\rangle, \quad m \in \mathbb{N}. \tag{1.3.39}$$

(iv) Más adelante nos detendremos en estudios algo menos triviales, como es el caso de los polinomios trigonométricos y los polinomios ortogonales, y otros casos más complejos y extensos, como son los espacios de Paley-Wiener, el de Bergmann, y el de Hardy.

1.4. Medibilidad, continuidad e integrabilidad en RKHSs

Entendida la base que relaciona núcleos y RKHSs, nos adentramos en la teoría, más elaborada, de estos espacios. Los núcleos reproductores satisfacen buenas propiedades de medibilidad, continuidad, e integrabilidad bajo condiciones no muy restrictivas, como veremos. Estudiamos la extensión de estas propiedades a las funciones que conforman dichos RKHSs. En adelante fijamos la notación de H como RHKS, $k : X \times X \to \mathbb{K}$ como su núcleo reproductor sobre el conjunto $X \neq \emptyset$ y sobre el cuerpo $\mathbb{K}$. Denotaremos también $\phi : X \to H_0$ y $\langle \cdot, \cdot \rangle$ como la aplicación característica y el producto escalar en H (nótese que estamos cogiendo una aplicación característica arbitraria del núcleo k, la cual no tiene por qué tener llegada en H, sino en un espacio de Hilbert H_0 indefinido).
Empezamos observando que, por la desigualdad de Cauchy-Schwarz sobre el producto interno dado en H y su norma asociada, se tiene, para todo $x, x' \in X$

$$|k(x,x')|^2 = |\langle k(\cdot,x'), k(\cdot,x)\rangle_H|^2 \leq ||k(\cdot,x')||^2||k(\cdot,x)||^2 = k(x',x') \cdot k(x,x), \tag{1.4.1}$$

donde se ha usado que $k(x,x) = \langle \phi(x), \phi(x)\rangle_{H_0} \geq 0$, para $x \in X$. Así pues, en caso de que la función núcleo sea acotada superiormente, es inmediato probar que $\sup_{x,x' \in X} |k(x,x')| = \sup_{x\in X} |k(x,x)|$. Esto es así porque, denotando $\alpha = \sup_{x,x'\in X} |k(x,x')|$ y $\beta = \sup_{x\in X} |k(x,x)|$, si bien es trivial que $\alpha \geq \beta$, para ver la desigualdad contraria basta usar (1.4.1) de tal forma que, para todos $x, x' \in X$

$$|k(x,x')| \leq \sqrt{k(x',x') \cdot k(x,x)} \leq \sqrt{\beta^2} = \beta.$$

Así pues, podemos definir la acotación de un núcleo como sigue.

Definición 1.4.1. Sea $k : X \times X \to \mathbb{K}$ un núcleo asociado a un RKHS H. Diremos que k está acotado si y solo si

$$\sup_{x\in X} \sqrt{|k(x,x)|} < \infty. \tag{1.4.2}$$

A la expresión (1.4.2) la tomaremos como la norma infinito $||k||_\infty$ del núcleo k. También diremos que si k es un núcleo que satisface (1.4.2), entonces k está *acotado en el sentido de los núcleos.*

Por las propiedades heredadas en k del producto interno de H, se sigue que (1.4.2) es una norma en el conjunto de núcleos en X. De hecho, para todo $x \in X$, por la norma definida a partir de un producto escalar, $||\phi(x)||_{H_0} = \sqrt{\langle \phi(x), \phi(x)\rangle_{H_0}}$, y por la forma de reproducir el núcleo a partir de la aplicación característica, tenemos $\sqrt{\langle \phi(x), \phi(x)\rangle_{H_0}} = \sqrt{k(x,x)}$. Por lo tanto

$$||\phi(x)||_{H_0} = \sqrt{k(x,x)}, \qquad x \in X. \tag{1.4.3}$$

Esto significa que, por la definición 1.4.1, ϕ estará acotado respecto de la norma en H_0 si y solo si k es un núcleo acotado. Damos ahora un resultado esencial en la caracterización de los núcleos acotados. En este trabajo, dado un cuerpo $\mathbb{K}$, denotaremos por $L^\infty(X)$ al espacio de funciones $f : X \to \mathbb{K}$ acotadas en todo X cuando no tengamos en X ninguna medida definida a priori, y al espacio de funciones $f : X \to \mathbb{K}$ con supremo esencial finito con respecto a una medida definida en X.

Teorema 1.4.2 (RKHS de núcleo acotado). *Sea X un conjunto no vacío y k un núcleo en X con RKHS asociado H. Entonces, k es acotado en el sentido de los núcleos si y solo si toda $f \in H$ es acotada en la norma de H. De hecho, en el caso de tener esta acotación, la aplicación identidad $\mathrm{id} : H \to L^\infty(X)$ es continua, y se tiene la igualdad de normas $||\mathrm{id}|| = ||k||_\infty$, donde la primera norma es la definida en el espacio de aplicaciones continuas y lineales $\mathcal{L}(H, L^\infty(X))$, y la segunda es la norma infinito ya definida para núcleos.*

Demostración. Empezamos suponiendo que k es acotado. Sea f una función cualquiera de H[5]. Entonces, por la desigualdad de Cauchy-Schwarz, y para todo $x \in X$ tenemos, por la propiedad reproductora

[5]Recordemos que los espacios de Hilbert H considerados están formados por funciones $f : X \to \mathbb{K}$, en los que la norma en el conjunto de llegada es la norma estándar del cuerpo $\mathbb{K}$ en cuestión.

$$\begin{aligned} |f(x)| &= |\langle f, k(\cdot, x)\rangle| \leq ||f||_H ||k(\cdot, x)||_H \\ &= ||f||_H \sqrt{\langle k(\cdot, x), k(\cdot, x)\rangle} = ||f||_H \sqrt{k(x, x)} \leq ||f||_H ||k||_\infty, \end{aligned} \tag{1.4.4}$$

por lo que ya tenemos probada la acotación de f en todo X.
Por lo tanto, a través de (1.4.4), tiene sentido hablar de la norma $||f||_\infty$ de f como elemento de $L^\infty(X)$, y podemos afirmar que $||f||_\infty \leq ||k||_\infty ||f||_H$ para cualquier $f \in H$, lo que prueba que la aplicación $\mathrm{id} : H \to L^\infty(X)$ está bien definida, y que $||\mathrm{id}|| \leq ||k||_\infty$. Como la acotación de aplicaciones lineales entre espacios normados equivale a su continuidad, acabamos de probar la continuidad de id.
Supongamos ahora que todo $f \in H$ está acotado. Por tanto, es claro que la identidad $\mathrm{id} : H \to L^\infty(X)$ está bien definida, pues para todo $f \in H$ se tiene $||f||_\infty < \infty$. Nótese que, en particular, $k(\cdot, x) \in L^\infty(X)$, para todo $x \in X$, lo que implica la acotación de k en el sentido del núcleo. Veamos la continuidad de la aplicación id. Para ello, usaremos el teorema del grafo cerrado (ver resultado A.0.3 del apéndice), para lo que, como H y $L^\infty(X)$ son espacios de Hilbert (luego también de Banach respecto a la norma inducida por su producto interno), veamos que el grafo de ésta, al que denotamos por gra(id), es cerrado.
Sea $(f, g) \in \overline{\mathrm{gra}(\mathrm{id})}$. Entonces, existe una sucesión $(f_n) \subset H$ para la que $\lim_{n\to\infty} ||f_n - f||_H = 0$ y $\lim_{n\to\infty} ||\mathrm{id} f_n - g||_\infty = 0$.
Ya que, tanto la convergencia en norma del infinito, como la convergencia en norma de H, implican convergencia puntual, se deduce que para todo $x \in X$, $\mathrm{id}(f)(x) = g(x)$. Entonces, $(f, g) \in \mathrm{gra}(\mathrm{id})$, como queríamos, luego $\mathrm{id} : H \to L^\infty(X)$ es continua. De esta manera, por continuidad de id,

$$|k(x, x)| \leq ||k(\cdot, x)||_\infty \leq ||\mathrm{id}|| \cdot ||k(\cdot, x)||_H = ||\mathrm{id}|| \sqrt{k(x, x)}, \tag{1.4.5}$$

donde se ha usado que, por la propiedad reproductora, $||k(\cdot, x)|| = \sqrt{k(x, x)}$. Así pues, $||k||_\infty = \sup_{x\in X} \sqrt{k(x, x)} \leq ||\mathrm{id}||$, luego $||k||_\infty \leq ||\mathrm{id}||$.
En resumen, ya hemos probado las desigualdades en ambos sentidos, luego $||k||_\infty = ||\mathrm{id}||$, y habiendo probado también la equivalencia propuesta, hemos acabado la demostración.

□

Damos ahora una serie de resultados esenciales en la caracterización de la medibilidad e integrabilidad de los elementos de un RKHS a partir de su núcleo, y viceversa.

Lema 1.4.3 (RKHSs de núcleo medibles). *Sea X un espacio medible con medida μ y sea k un núcleo en X con RKHS asociado H. Entonces, todas las funciones $f \in H$ son medibles con la medida definida en X si y solo si $k(\cdot, x) : X \to \mathbb{R}$ es medible para todo $x \in X$.*

Demostración. Si empezamos suponiendo que cualquier $f \in H$ es medible respecto a μ, entonces $k(\cdot, x) \in H$ también será medible, respecto a la misma medida, para todo $x \in X$.

Para la otra implicación, supongamos que $k(\cdot, x)$ es medible para cualquier $x \in X$. Por lo tanto, todas las funciones $f : X \to \mathbb{K}$ del espacio H_{pre} definido en (1.3.13) serán medibles, por ser combinaciones lineales de funciones medibles. Sea f un elemento arbitrario de H. Por el teorema mencionado 1.3.6, dada la densidad de H_{pre} en H, existe una sucesión (f_n) en H_{pre} tal que $\lim_{n\to\infty} ||f_n - f||_H = 0$. Por el lema 1.3.2, usando que H es un RKHS, tenemos que la convergencia en norma de H implica la convergencia puntual de las funciones que pertenecen a H, luego se tiene que $\lim_{n\to\infty} f_n(x) = f(x)$ para todo $x \in X$. Así pues, por la regularidad de la medibildad por tales límites, tenemos que f es función medible respecto de la medida μ en X.

□

Antes de enunciar el siguiente resultado, recordemos que la medibilidad débil de una aplicación $g : X \to \mathbb{K}$, con X un espacio de medida no vacío, y $\mathbb{K}$ un cuerpo ($\mathbb{R}$ o $\mathbb{C}$), y donde g es un elemento de un espacio vectorial de funciones V, se define a través de la medibilidad de todas las aplicaciones $\langle w, g\rangle_{V',V}$ para todos los elementos w del dual topológico V'.

Lema 1.4.4 (Medibilidad de la aplicación característica canónica)**.** *Sea X un espacio medible, y sea k un núcleo en X tal que $k(\cdot, x) : X \to \mathbb{R}$ es medible para todo $x \in X$. Si el RKHS H asociado a k es separable, entonces tanto la aplicación característica canónica $\phi : X \to H$ como el propio núcleo $k : X \times X \to \mathbb{R}$ son medibles.*

Demostración. Sea $w \in H'$ un elemento del dual topológico de H. Entonces, para todo $x \in X$, por el teorema de representación de Fréchet-Riesz A.0.12, existe un único $f \in H$ tal que:

$$\langle w, \phi(x) \rangle_{H',H} = \langle f, \phi(x) \rangle_H = \langle f, k(\cdot, x) \rangle_H = f(x), \tag{1.4.6}$$

donde, en la segunda igualdad, hemos usado también la propiedad reproductora de k. Así pues, como por hipótesis $k(\cdot, x) : X \to \mathbb{R}$ es medible para todo $x \in X$, por el lema 1.4.3 tenemos que todos los elementos $f \in H$ son también medibles, con la σ-álgebra dada en el espacio de medida X y con la σ-álgebra de Borel $\mathbb{R}$, heredada por la topología usual inducida por la norma en éste último espacio. Por lo tanto, la igualdad anterior nos dice que la aplicación $\langle w, \phi(\cdot) \rangle_{H',H} : X \to \mathbb{R}$ es medible para todo $w \in H'$, i.e., $\phi : X \to \mathbb{R}$ es débilmente medible.
Además, como H es separable, entonces la imagen $\phi(X) \subset H$ también lo es, pues H es un espacio métrico. Por tanto, por el teorema de medibilidad de Petty B.0.6, $\phi : X \to H$ es medible.

En cuanto a la mediblidad de $k : X \times X \to \mathbb{R}$, consideramos las aplicaciones $h : X \times X \to H \times H$ y $\xi : H \times H \to \mathbb{R}$, con $h(x, x') = (\phi(x'), \phi(x))$ para todos $x, x' \in X$, y con ξ la aplicación producto escalar en H, $\xi(h, h') = \langle h, h' \rangle_H$. Obsérvese que h es medible, por ser ϕ medible (con lo que todas las componentes de la función son medibles, lo que caracteriza la medibilidad de una aplicación con llegada en un producto cartesiano de espacios de medida), y ξ es continua, con lo que $k = \xi \circ h : X \times X \to \mathbb{R}$ es medible. □

En lo que inmediatamente sigue, estudiamos la integrabilidad de las funciones que pertenecen a RKHSs a partir de la integrabilidad de su núcleo. Note el lector que, tal y como lo hemos definido, la función $x \to k(x, x)$ es positiva para todo $x \in X$, con X el conjunto arbitrario no vacío en que definimos las funciones en cuestión. Por tanto, su integral estará siempre bien definida, sea finita o infinita.
Antes, recordemos lo que es una medida σ-finita. Dado un conjunto X y dada una medida μ definida sobre una cierta σ-álgebra Δ de X, decimos que μ es finita si $\mu(X) \in [0, \infty)$, y diremos que es σ-finita si existe una familia numerable $\{A_n\}_n \subset \Delta$ tal que $X = \cup_n A_n$ y $\mu(A_n) \in [0, \infty)$ para todo $n \in \mathbb{N}$.

Teorema 1.4.5 (**Operadores integrales con núcleos I**)**.** *Sea X un espacio medible, μ una medida σ-finita en X, y H un RKHS separable en X con núcleo medible $k : X \times X \to \mathbb{R}$. Suponemos que existe un $p \in [1, \infty)$ tal que*

$$||k||_{L^p(\mu)} := \left(\int_X k^{\frac{p}{2}}(x, x) d\mu(x) \right)^{\frac{1}{p}} < \infty. \tag{1.4.7}$$

Entonces, H está formado por funciones $L^p(\mu)$-integrables y la aplicación inclusión $id : H \to L^p(\mu)$ es continua con $||id|| \leq ||k||_{L^p(\mu)}$. Además, el adjunto de esta aplicación id *es el operador $S_k : L_{p'}(\mu) \to H$ dado por*

$$S_k g(x) := \int_X k(x, x') g(x') d\mu(x'), \qquad g \in L_{p'}(\mu), x \in X, \tag{1.4.8}$$

donde p' es el conjugado de p, esto es, $\frac{1}{p} + \frac{1}{p'} = 1$ (si $p = 1$, entonces $p' = \infty$).
Finalmente, las siguientes afirmaciones se cumplen:

1. *H es denso en $L^p(\mu)$ si y solo si $S_k : L_{p'}(\mu) \to H$ es inyectiva.*

2. *$S_k : L_{p'}(\mu) \to H$ tiene una imagen densa en H si y solo si* $\mathrm{id} : H \to L^p(\mu)$ *es inyectiva.*

Demostración. Sea f cualquier elemento de H. Veamos primero que f es p-integrable, y que la función identidad $\mathrm{id} : H \to L^p(\mu)$ es continua. Recordando que, como ya vimos más arriba, $||k(\cdot, x)||_H = \sqrt{k(x,x)}$ para todo $x \in X$, se tiene

$$\begin{aligned}\int_X |f(x)|^p d\mu(x) &= \int_X |\langle f, k(\cdot,x)\rangle|^p d\mu(x) \leq \int_X [||f||_H \cdot ||k(\cdot,x)||_H]^p \, d\mu(x) \\ &= ||f||_H^p \int_X k^{\frac{p}{2}}(x,x) d\mu(x) < \infty,\end{aligned}$$

donde hemos usado la desigualdad de Cauchy-Schwarz en la primera de las desigualdades anteriores, además de la propiedad (1.4.7). Así, acabamos de probar que $||f||_{L^p(\mu)} \leq ||k||_{L^p(\mu)} \cdot ||f||_H$, i.e., que toda f en H es p-integrable con respecto a μ, y que la aplicación identidad $\mathrm{id} : H \to L^p(\mu)$ es acotada (luego continua), así como que $||\mathrm{id}|| \leq ||k||_{L^p(\mu)}$.
Dado ahora $p' \in [1, \infty]$ el conjugado de $p \in [1, \infty]$, consideremos un $g \in L_{p'}(\mu)$ arbitrario. Entonces:

$$\begin{aligned}\int_X |k(x,x')g(x')| d\mu(x') &\leq \sqrt{|k(x,x)|} \int_X \sqrt{k(x',x')} \cdot |g(x')| d\mu(x') \\ &\leq ||k||_\infty \left(\int_X |k(x',x')|^{\frac{p}{2}} d\mu(x')\right)^{\frac{1}{p}} \left(\int_X |g(x')|^{p'}\right)^{\frac{1}{p'}} \\ &= ||k||_\infty \cdot ||k||_{L^p(\mu)} \cdot ||g||_{L_{p'}(\mu)}, \qquad (1.4.9)\end{aligned}$$

donde hemos utilizado la desigualdad probada para el núcleo k, por la que

$$k(x,x') \leq \sqrt{k(x,x)}\sqrt{k(x',x')}, \qquad (1.4.10)$$

para todos $x, x' \in X$, así como la desigualdad de Hölder sobre $g \in L_{p'}(\mu)$ y sobre el núcleo k (ver expresión (1.4.7)). También hemos utilizado la definición de la norma infinito del núcleo, $||k||_\infty$.
En (1.4.9) hemos demostrado dos cosas:

1. La función $x' \in X \mapsto k(x,x')g(x')$ es μ-integrable para todo $x \in X$ fijo, esto es, pertenece a $L^1(\mu)$.

2. Obsérvese que, como ya vimos anteriormente, $\sqrt{k(x',x')} = ||\phi(x')||_H$ para todo $x' \in X$, y para ϕ la aplicación característica canónica de k. Por la segunda desigualdad de (1.4.9), tenemos que la función $\xi : x' \mapsto ||\phi(x')g(x')||_H$ pertenece a $L_1(\mu)$, ya que, como se vio implícitamente,

$$\int_X \sqrt{|k(x',x')|} \cdot |g(x')| d\mu(x') = \int_X ||\phi(x')||_H \cdot |g(x')| d\mu(x') = \int_X ||g(x')\phi(x')||_H \cdot d\mu(x'), \qquad (1.4.11)$$

lo cual es finito por lo dicho en (1.4.9). Así pues, por la caracterización de la integrabilidad de Bochner dada en B.0.8, la función $x' \mapsto g(x')\phi(x')$ es μ-integrable. Pongamos

$$\overline{g} := \int_X g(x')\phi(x') \cdot d\mu(x') \in H, \qquad (1.4.12)$$

y veamos que $\overline{g}$ coincide con $S_k g$ para el $g \in L_{p'}(\mu)$ arbitrario fijado. Consideramos, para un $x \in X$ fijo y arbitrario, la aplicación $\theta_x : H \to \mathbb{R}$ definida como $\theta_x(f) = \langle f, \phi(x)\rangle_H$, la cual es lineal y acotada. Aplicando la conmutatividad del resultado (B.0.12), tenemos

$$\begin{aligned}S_k g(x) &= \int_X k(x,x')g(x')d\mu(x') = \int_X \langle \phi(x'), \phi(x)\rangle_H g(x')d\mu(x') = \int_X \theta_x(\phi(x'))g(x')d\mu(x') \\ &= \theta_x\left(\int_X \phi(x')g(x')d\mu(x')\right) = \left\langle \int_X \phi(x')g(x')d\mu(x'), \phi(x)\right\rangle \\ &= \left\langle \int_X \phi(x')g(x')d\mu(x'), k(\cdot,x)\right\rangle.\end{aligned}$$

Por la propiedad reproductora del núcleo k, tenemos la igualdad $S_k g = \overline{g}$.
Ahora, para continuar, enunciamos un resultado de sobra conocido sobre la dualidad de los espacios $L^p(\mu)$.

Teorema 1.4.6. *Sea (X, μ) un espacio de medida, y sean $p \in [1, \infty)$ y su conjugado q. Si $g \in L^q(\mu)$, definimos $F_g : L^p(\mu) \to \mathbb{K}$ de la forma*

$$F_g(f) = \int fg d\mu. \tag{1.4.13}$$

Si $p \in (1, \infty)$, la aplicación $g \mapsto F_g$ define un isomorfismo isométrico de $L_q(\mu)$ en $L_p'(\mu)$. Si $p = 1$, y (X, μ) es σ-finito, entonces $g \mapsto F_g$ define también un isomorfismo isométrico de $L^\infty(\mu)$ en $L_1'(\mu)$.

Este resultado 1.4.6 nos da la expresión del isomorfismo isométrico por el que, dado el espacio de medida X con medida μ, relacionamos el dual de $L^p(\mu)$ y $L_{p'}(\mu)$, con p' conjugado de p. Nótese que, como μ es σ-finito en X, podemos tomar no solo $1 < p, p' < \infty$, sino también $p = 1, p' = \infty$. Tomamos dicho isomorfismo isométrico respecto de $\mathrm{id}(f) \in L^p(\mu)$, con $f \in H$, el cual lo podemos abreviar como $F_{\mathrm{id}(f)} := \langle \cdot, \mathrm{id}(f) \rangle_{L_{p'}(\mu), L^p(\mu)}$. Por lo tanto, dado $g \in L_{p'}(\mu)$[6], usando la propiedad reproductora de k y, una vez más, la conmutatividad que expresamos en B.0.12 dada la integrabilidad de Bochner, obtenemos

$$\begin{aligned} F_{\mathrm{id}(f)}(g) &= \langle g, F_{\mathrm{id}(f)} \rangle_{L_{p'}(\mu), L^p(\mu)} = \int_X g(x) \langle f, k(\cdot, x) \rangle_H d\mu(x) \\ &= \left\langle f, \int_X g(x) k(\cdot, x) d\mu(x) \right\rangle_H = \langle f, S_k g \rangle_H = \langle I S_k g, f \rangle_{H', H}, \end{aligned}$$

donde $I : H \to H'$ es el isomorfismo isométrico dado en el teorema de representación de Fréchet-Riesz A.0.12. Así pues, identificando H y H' a través de dicha aplicación I, tenemos que el adjunto de $\mathrm{id} : H \to L^p(\mu)$ es $\mathrm{id}' = S_k$, tal y como afirmamos en el enunciado del teorema[7].
Las últimas dos afirmaciones se siguen inmediatamente del hecho de que, dado un operador lineal y acotado (en este caso S_k e id), entonces su imagen es densa en el conjunto de llegada si y solo si su adjunto es inyectivo. Este resultado se recoge en A.0.17. □

A continuación, particularizamos al caso $p = 2$. La estructura del espacio de Hilbert $L^2(\mu)$ otorga al adjunto S_k algunas características adicionales de notable importancia. No obstante, para ello necesitamos dar una nociones de la teoría espectral, cuyo principal objetivo es el de buscar las condiciones que imponer a un operador arbitrario (nosotros los estudiaremos sobre espacios de Hilbert) de cara a "diagonalizar" estos operadores, análogamente a como se hace en Álgebra Lineal. Por supuesto, necesitamos sentar una serie de definiciones y de terminologías para poder entender qué significa "diagonalizar" en este contexto. Entienda el lector que, por no ser este el objetivo primero de este trabajo, no podemos detenernos en estas disquisiciones.
Denotamos por $\mathcal{L}(H_1, H_2)$ al espacio de aplicaciones continuas entre los espacios de Hilbert H_1 y H_2. Cuando $H_1 = H_2$, lo denotaremos directamente como $\mathcal{L}(H)$.
En este trabajo, la norma en $\mathcal{L}(H_1, H_2)$ (para H_1 y H_2 espacios de Hilbert) de un operador dado $A : H_1 \to H_2$ se define como $||A|| = \sup\{||Ah|| : h \in H_1,\ ||h|| \leq 1\}$. Esta aclaración es indispensable, entre otras cosas, para entender la caracterización más elemental de operadores compactos.
Consideraremos a H un $\mathbb{K}$-espacio de Hilbert, y a $T \in \mathcal{L}(H)$.
Por aclarar brevemente la terminología usada, $\lambda \in \mathbb{K}$ es *autovalor* de T si existe $x \in H$ (al que llamamos *autovector*), $x \neq 0$, tal que $Tx = \lambda x$. Claramente, para cada autovector existe un único

[6] La nomenclatura de producto escalar usada para la evaluación de $F_{id(f)}$ es un abuso de notación habitual en análisis funcional clásico en la evaluación de funcionales en un cierto espacio (esto es, en la evaluación de un elemento del dual topológico).

[7] Lo más correcto sería decir que S_k va de $L_{p'}(\mu)$ a H', pero a través de la identificación de H y H' con el isomorfismo I, simplificamos el enunciado diciendo, directamente, que S_k tiene llegada en H.

autovalor, lo cual no es recíproco. Para cada autovalor λ de T, llamamos *autoespacio de* λ al espacio $E(\lambda) := \ker(\lambda \mathrm{id}_H - T)$, y llamamos *multiplicidad geométrica* a la dimensión de este subespacio, $dim E(\lambda)$. Además, si T es autoadjunto, los subespacios $E(\lambda_1)$ y $E(\lambda_2)$ son ortogonales, y si T es (estrictamente) positivo entonces todos sus autovalores son también (estrictamente) positivos.
Antes de formular el teorema espectral, damos un lema que relaciona esta teoría con la teoría del apéndice A.

Lema 1.4.7. *Sean H_1 y H_2 dos espacios de Hilbert. Sean $S \in \mathcal{L}(H_1, H_2)$, y sus operadores autoadjuntos y positivos asociados, que definimos como:*

$$T_1 := S^*S : H_1 \to H_1, \qquad T_2 := SS^* : H_2 \to H_2. \tag{1.4.14}$$

Escribimos $E_1(\lambda) := \ker(\lambda id_H - T_1)$ y $E_2(\lambda) := \ker(\lambda id_H - T_2)$, para un autovalor no nulo λ de T_1. Entonces, $S : E_1(\lambda) \to E_2(\lambda)$ está bien definido y es inyectivo. Por lo tanto, λ también es autovalor de T_2. Además, se tiene que $(S^*S)|_{E_1(\lambda)} = \lambda id_{E_1(\lambda)}$.

De hecho, por la simetría del argumento, podemos intercambiar T_1 y T_2 para ver que S^*S y SS^* tienen los mismo autovalores *no nulos*, con las mismas multiplicidades geométricas.
A continuación, nos centramos en los autovalores de operadores compactos y autoadjuntos. En el siguiente enunciado, englobamos a las familias finitas e infinitamente numerables de autovalores. En el último caso, diremos que la familia $(\lambda_i)_{i\in I} \subset \mathbb{R}$ converge a cero si $\lim_{i\to\infty} \lambda_i = 0$. En general, diremos que dicha familia converge hacia cero aunque sea finita, en cuya caso ignore el lector la parte de convergencia.

Teorema 1.4.8 (Teorema espectral). *Sean H un $\mathbb{R}$-espacio de Hilbert y $T \in \mathcal{L}(H)$ un operador acotado que es compacto y autoadjunto. Entonces, existe un sistema ortonormal $(e_i)_{i\in I}$ a lo sumo infinito numerable, y una familia $(\lambda_i(T))_{i\in I}$ de valores reales que converge a cero, tales que*

1. $|\lambda_1| \geq |\lambda_2| \geq \cdots > 0.$
2. $Tx = \sum_{i\in I} \lambda_i(T)\langle x, e_i\rangle e_i$, *para todo* $x \in H$.
3. $\{\lambda_i(T) : i \in I\}$ *es el conjunto de autovalores no nulos de* T.

Por cómo hemos escrito la sucesión de autovalores λ_i, todos ellos son no nulos.
Nótese que si, además, T es un operador *positivo* de $\mathcal{L}(H)$, entonces los autovalores de 1.4.8 son positivos, y por ende tiene sentido definir, para todo $r \geq 0$, los operadores $T^r : H \to H$, de la forma:

$$T^r x = \sum_{i\in I} \lambda_i^r \langle x, e_i\rangle e_i, \tag{1.4.15}$$

para todo $x \in H$. Notemos también que, para todo $i \in I$, y para todo $r \geq 0$, $\lambda_i(T^r) = \lambda_i^r(T)$.
El siguiente lema tiene una demostración muy sencilla, a raíz de la construcción realizada.

Lema 1.4.9. *Sea H un $\mathbb{R}$-espacio de Hilbert, y sea $T \in \mathcal{L}(H)$ compacto, positivo, y autoadjunto. Entonces, con la notación especificada, $T^rT^s = T^{r+s}$, $T^1 = T$, y $T^2 = TT$.*

Con este resultado, ya podemos definir lo que son las potencias fraccionarias.

Definición 1.4.10. Con la notación del lema 1.4.9, y con $r \geq 0$, llamamos a T^r **potencia fraccionaria** de T de exponente r.

Con esta definición, podemos definir la **raíz cuadrada** de un operador $T \in \mathcal{L}(H)$ positivo, compacto y autoadjunto como el operador $T^{\frac{1}{2}}$ tal que $T^{\frac{1}{2}}T^{\frac{1}{2}} = T$. Escribimos $\sqrt{T} := T^{\frac{1}{2}}$.
Consideramos ahora un operador compacto $S \in \mathcal{L}(H_1, H_2)$ para dos $\mathbb{R}$-espacios de Hilbert H_1 y H_2. Entonces, el operador conocido $S^*S : H_1 \to H_1$ es compacto, positivo, y autoadjunto, y por lo tanto el teorema espectral 1.4.8 es aplicable sobre S^*S, lo que da sentido a la siguiente terminología.

Definición 1.4.11. Sean H_1, H_2 dos espacios de Hilbert. Sea S un operador compacto de $\mathcal{L}(H_1, H_2)$. Llamamos **números singulares** de S a los valores que siguen

$$s_i(S) := \begin{cases} \sqrt{\lambda_i(S^*S)} & si \quad i \in I \\ \\ 0 & si \quad i \in \mathbb{N} - I, \end{cases} \tag{1.4.16}$$

donde, por el desarrollo anterior, $\sqrt{\lambda_i(S^*S)} = \lambda_i(\sqrt{S^*S})$.

Como, en virtud de 1.4.7, S^*S y SS^* tienen los mismos autovalores no nulos en estas condiciones, con las mismas multiplicidades geométricas, entonces es claro que $s_i(S) = s_i(S^*)$ para todo $i \geq 1$. Si $T \in \mathcal{L}(H)$ es compacto, positivo, y autoadjunto, entonces $s_i(T) = \sqrt{\lambda_i(T^*T)} = \sqrt{\lambda_i(T^2)} = \lambda_i(T)$ para todo $i \geq 1$.
Es así que, por el teorema espectral, podemos afirmar que los números singulares de operadores compactos, positivos y autoadjuntos, si son infinitos, decaen al cero. Entremos ahora a refinar la *velocidad* con la que se da esta convergencia.

Definición 1.4.12. Sean H un espacio de Hilbert, y $T \in \mathcal{L}(H)$ un operador compacto de $\mathcal{L}(H_1, H_2)$. Diremos que T es **nuclear** o de la **clase de traza** si y solo si

$$||T||_{nuc} := \sum_{i=1}^{\infty} s_i(T) < \infty. \tag{1.4.17}$$

Por las consideraciones anteriores, tenemos que un operador compacto, positivo, y autoadjunto $T \in \mathcal{L}(H)$ es nuclear si y solo si

$$\sum_{i \in I} |\lambda_i(T)| < \infty, \tag{1.4.18}$$

de tal forma que esta última suma coincide con $||T||_{nuc}$. Para poder caracterizar la integrabilidad en RKHSs con $p = 2$, tiene interés el siguiente concepto.

Definición 1.4.13. Sean H_1, H_2 dos espacios de Hilbert. Diremos que una aplicación acotada $S \in \mathcal{L}(H_1, H_2)$ es **un operador de Hilbert-Schmidt** si y solo si

$$||S||_{HS} := \left(\sum_{i \in I} ||Se_i||_{H_2}^2 \right)^{\frac{1}{2}} < \infty, \tag{1.4.19}$$

donde $(e_i)_{i \in I}$ es una base ortonormal de H_1.

A raíz de la definición dada, damos el siguiente resultado.

Proposición 1.4.14. *Un operador Hilbert-Schmidt $T \in \mathcal{L}(H_1, H_2)$, con H_1 y H_2 dos espacios de Hilbert, es compacto, y $|| \cdot ||_{HS}$ define una norma, llamada **norma de Hilbert-Schmidt**, que es independiente de la base ortonormal. De hecho, tenemos que*

$$||S||_{HS} = \left(\sum_{i=1}^{\infty} s_i^2(S) \right)^{\frac{1}{2}}. \tag{1.4.20}$$

Demostración. La prueba de la igualdad (1.4.20) para la norma es laboriosa y no es relevante para nuestros propósitos, por lo que la omitimos. Ahora, si usamos que $s_i^2(S) = \lambda_i(S^*S)$ y $s_i(S^*S) = \lambda_i(\sqrt{S^*SS^*S}) = \lambda_i(S^*S)$ (i.e., que $s_i^2(S) = s_i(S^*S)$), de dicha igualdad se deduce que $||S||_{HS}^2 = ||S^*S||_{nuc}$. En particular, S es Hilbert-Schmidt si y solo si S^*S es nuclear. El hecho de que $|| \cdot ||_{HS}$

sea norma se sigue inmediatamente de que $||\cdot||_{H_2}$ es norma. La independencia con respecto a la base es implícita al estar expresando $||S||_{HS}$ en términos de los autovalores de S^*S.
Probemos ahora la primera implicación enunciada. Denotamos por $(e_i)_{i=1}^{\infty}$ y por $(f_j)_{j=1}^{\infty}$ a unas bases ortonormales de H_1 y H_2, respectivamente. Siguiendo el procedimiento estándar de aproximación de operadores compactos por operadores de rango finito, definimos los llamados operadores N-truncados de T como sigue

$$T_N(x) = \sum_{i,j=1}^{N} \langle f_j, Te_i\rangle_{H_2} \langle e_i, x\rangle_{H_1} f_j. \tag{1.4.21}$$

Pues bien, como los T_N son operadores de rango finito de $\mathcal{L}(H_1, H_2)$ para todo $N \in \mathbb{N}$, y por lo tanto compactos, entonces, si probamos la convergencia en norma en $\mathcal{L}(H_1, H_2)$ de T_N a T para $N \to \infty$, ya tendríamos que T es compacto.
Para todo $N \in \mathbb{N}$, y dado un $x \in H_1$, $x \neq 0$, se tiene que

$$(T_N - T)(x) = (T_N - T)\left(\sum_{i=1}^{\infty} \langle x, e_i\rangle_{H_1} e_i\right) = \sum_{i=N+1}^{\infty} \langle x, e_i\rangle_{H_1} T(e_i). \tag{1.4.22}$$

Tomando la norma de H_2, se sigue que

$$\begin{aligned} ||\,(T_N - T)\left(\sum_{i=1}^{\infty} \langle x, e_i\rangle_{H_1} e_i\right)||_{H_2} &= ||\sum_{i=N+1}^{\infty} \langle x, e_i\rangle_{H_1} T(e_i)||_{H_2} \\ &\leq \sum_{i=N+1}^{\infty} |\langle x, e_i\rangle_{H_1}|\, ||T(e_i)||_{H_2}, \end{aligned} \tag{1.4.23}$$

donde se ha utilizado la desigualdad triangular. Además, $(|\langle x, e_i\rangle_{H_1}|)_{i=1}^{\infty}$ y $(||T(e_i)||_{H_2})_{i=1}^{\infty}$ son sucesiones que pertenecen a $l^2(\mathbb{R})$, por la identidad de Parseval y por ser T un operador de HS, respectivamente. Entonces, podemos acotar la expresión anterior como sigue

$$||\,(T_N - T)\left(\sum_{i=1}^{\infty} \langle x, e_i\rangle_{H_1} e_i\right)||_{H_2} \leq \left(\sum_{i=N+1}^{\infty} |\langle x, e_i\rangle_{H_1}|^2\right)^{1/2} \left(\sum_{i=N+1}^{\infty} ||T(e_i)||_{H_2}^2\right)^{1/2}. \tag{1.4.24}$$

Notemos que, por la propia identidad de Parseval, se tiene que

$$\sum_{i=1}^{\infty} |\langle x, e_i\rangle_{H_1}|^2 = ||x||_{H_1}^2. \tag{1.4.25}$$

Entonces, de lo anterior se sigue la siguiente acotación

$$||\,(T_N - T)(x)||_{H_2} \leq ||x||_{H_1} \left(\sum_{i=N+1}^{\infty} ||T(e_i)||_{H_2}^2\right)^{1/2}. \tag{1.4.26}$$

Si ahora pasamos $||x||_{H_1}$ al otro lado, y tenemos en cuenta que la serie del último factor es el resto de una serie convergente, se sigue la convergencia deseada en $\mathcal{L}(H_1, H_2)$, es decir,

$$||T_N - T|| \to 0, \quad N \to \infty. \tag{1.4.27}$$

Entonces, T es compacto. □

Con las ideas que acabamos de aportar, tratamos ahora el caso de operadores integrales en RKHSs en $L^2(\mu)$, con μ una medida que refinamos en el enunciado.

Teorema 1.4.15 (Operadores integrales con núcleos II). *Sea X un espacio medible con una medida σ-finita, y H un RKHS separable en X con núcleo medible $k : X \times X \to \mathbb{R}$ que cumple $||k||_{L^2(\mu)} < \infty$ (de la expresión (1.4.7) con $p = 2$). Entonces, $S_k : L^2(\mu) \to H$ (de la expresión (1.4.8) con $p = 2$) es un operador de Hilbert-Schmidt con*

$$||S_k||_{HS} = ||k||_{L^2(\mu)}. \tag{1.4.28}$$

Además, el operador integral $T_k = S_k^* S_k : L^2(\mu) \to L^2(\mu)$ es compacto, positivo, autoadjunto, y nuclear con $||T_k||_{nuc} = ||S_k||_{HS} = ||k||_{L^2(\mu)}$.

Demostración. Empecemos recordando el resultado más general probado en el teorema 1.4.5 para todo $p \in [1, \infty)$, aunque particularizando ahora para $p = 2$: como estamos asumiendo que $||k||_{L^2(\mu)} < \infty$, entonces nuestro actual RKHS separable H sobre X está contenido en $L^2(\mu)$. Además, como el real conjugado de $p = 2$ es $p' = 2$, el adjunto del operador inclusión id : $H \to L^2(\mu)$ viene dado por $S_k : L^2(\mu) \to H$, tal y como lo expresamos en el mencionado teorema (en particular, con la expresión (1.4.8)). Además, por la construcción seguida, $S_k^* =$ id. Recordemos también que, tal y como se comentó anteriormente, consideramos S_k directamente con llegada en H vía el isomorfismo I del teorema de Riesz-Fréchet.
Veamos que $S_k^* : H \to L^2(\mu)$ es un operador de Hilbert-Schmidt. Por lo dicho anteriormente, esto implicaría que S_k también lo es. Sea pues $(e_i)_{i=1}^{\infty}$ una base ortonormal de H (la cogemos infinita numerable por ser H separable). Aplicaremos el teorema de la convergencia monótona para intercambiar los signos de suma e integral, así como que $S_k^*(e_i(x)) = \mathrm{id}(e_i(x)) = e_i(x)$. Además, utilizaremos el teorema 1.3.4. Por este, para todo $x, x' \in X$ sabemos que

$$k(x, x') = \sum_{i \in \mathbb{N}} e_i(x)\overline{e_i(x')}. \tag{1.4.29}$$

Así pues, tenemos que

$$\begin{aligned} \sum_{i=1}^{\infty} ||S_k^* e_i||^2_{L^2(\mu)} &= \int_X \sum_{i=1}^{\infty} |S_k^* e_i(x)|^2 d\mu(x) \\ &= \int_X \sum_{i=1}^{\infty} |e_i(x)|^2 d\mu(x) = \int_X k(x,x) d\mu(x) = ||k||^2_{L^2(\mu)} < \infty, \end{aligned} \tag{1.4.30}$$

siendo la última expresión, como ya vimos en el teorema 1.4.8, la norma al cuadrado del núcleo k sobre $L^p(\mu)$ para $p = 2$, la cual es finita por hipótesis.
Así pues, acabamos de probar que S_k^*, y por lo tanto S_k, es un operador de Hilbert-Schmidt con la norma dada sobre $L^2(\mu)$. Implícitamente, también hemos probado que $||S_k||_{HS} = ||k||_{L^2(\mu)}$.
Por otro lado, por la proposición 1.4.14, como S_k es de Hilbert-Schmidt, entonces es compacto. También, por el lema A.0.16, tenemos que $T_k = S_k^* S_k$ es positivo y autoadjunto.
Solo nos falta ver que T_k es compacto y nuclear, y que $||T_k||_{\text{nuc}}$ coincide con la norma de Hilbert-Schmidt de S_k. En cuanto a la compacidad de T_k, fijémonos que $T_k = S_k^* S_k$ es la composición de dos operadores compactos y acotados, de la forma $T_k : L^2(\mu) \to L^2(\mu)$. Entonces, $T_k = S_k^* S_k$ es compacto.
Veamos ahora las dos últimas propiedades propuestas para T_k. Como ya hemos visto en el breve análisis de la teoría espectral hecho, observando con cuidado las definiciones, afirmamos que S_k es de Hilbert-Schmidt si y solo si $S_k^* S_k$ es nuclear. Ya sabemos que S_k es de Hilbert-Schmidt, luego T_k es un operador nuclear. En cuanto a la igualdad de normas, también vimos en el mismo análisis que $||S_k||^2_{HS} = ||S_k^* S_k||_{\text{nuc}} = ||T_k||_{\text{nuc}}$. □

Nuestro siguiente objetivo es el de estudiar la continuidad para las funciones núcleo que estamos tratando. En adelante, diremos que k es **separadamente continua** si $k(\cdot, x) : X \to \mathbb{R}$ es continua.

Nótese que en esta definición va implícita la simetría de todo núcleo que tome valores reales.
Denotaremos por $C(X)$ al conjunto de funciones $f : X \to \mathbb{R}$ continuas, y por $C_a(X)$ al espacio de funciones $f : X \to \mathbb{R}$ continuas y acotadas en X.
Caracterizamos, con el siguiente lema, los RKHS contenidos en $C(X)$. Para hablar de continuidad de los elementos de H, necesitamos fijar en adelante al conjunto X no vacío como un espacio topológico.

Lema 1.4.16 (RKHS de funciones continuas)**.** *Sea X un espacio topológico y k un núcleo en X asociado a un RKHS H. Entonces, las dos siguiente afirmaciones son equivalentes:*

1. *Dada $f \in H$, f es continua y acotada.*

2. *k es acotado y separadamente continuo.*

Además, en el caso de que se cumplan, la inclusión $\mathrm{id} : H \to C_a(X)$ *es continua y se tiene la igualdad* $||\mathrm{id}|| = ||k||_\infty$.

Demostración. Empezamos suponiendo que k es un núcleo acotado y sepadamente continuo. Entonces, el espacio H_{pre} de (1.3.13) está formado por funciones continuas y acotadas, dado que k es separadamente continuo y acotado en $X \times X$. Por la densidad de dicho conjunto H_{pre}, dado $f \in H$, existe una sucesión $\{f_n\}_{n=1}^{\infty} \subset H_{\text{pre}}$ tal que $\lim_{n\to\infty} ||f_n - f||_H = 0$. Por el lema 1.4.2, esta convergencia implica la convergencia en $L^\infty(X)$, que no es otra que la convergencia uniforme. Así pues, $\lim_{n\to\infty} ||f_n - f||_\infty = 0$. Como f es límite uniforme de funciones continuas ya acotadas, entonces f también es continua y acotada. Además, tanto la continuidad de la inclusión $\mathrm{id} : H \to C_a(X)$ como la igualdad $||\mathrm{id}|| = ||k||_\infty$, se siguen inmediatamente del mentado lema 1.4.2, puesto que $\mathrm{id} : H \to C_a(X)$ es una aplicación restringida de $\mathrm{id} : H \to L^\infty(X)$, ya que $C_a(X) \subset L^\infty(X)$.
Recíprocamente, supongamos que H está formado por funciones continuas y acotadas, i.e., que $H \subset C_a(X)$. En particular, $k(\cdot, x) : X \to \mathbb{K}$ es continua, luego k es separadamente continua. Por su parte, la acotación de k se sigue, otra vez, del lema 1.4.2, ya que H está contenido en $C_a(X) \subset L^\infty(X)$. □

A continuación, caracterizamos la continuidad del núcleo k en diferentes términos. Para ello, recurrimos a una nueva noción que nos permitirá hablar de semimétrica, y hasta de métrica, en términos de k.
Para ello, sea k un núcleo en X con aplicación característica $\phi : X \to H$. Entonces, definimos la **distancia nuclear** d_k en X, asociada a k, como

$$d_k(x, x') := ||\phi(x) - \phi(x')||_H, \quad x, x' \in X. \tag{1.4.31}$$

Se deduce, dado que estamos definiendo d_k en términos de una norma (la de H asociada a su producto interno), que d_k es siempre una semimétrica. Además, si ϕ *es una aplicación inyectiva*, se sigue también que d_k es una métrica. Para ver esto, sean $x, x' \in X$, entonces

$$d_k(x, x') = 0 \implies \phi(x) = \phi(x') \implies x = x', \tag{1.4.32}$$

donde se ha usado, como era de esperar, que $|| \cdot ||_H$ es una norma. Además, un sencillo cálculo sobre productos internos en H nos lleva a

$$d_k(x, x') = \sqrt{k(x,x) - 2k(x,x') + k(x',x')}, \quad x, x' \in X, \tag{1.4.33}$$

con lo que d_k, tal y como está construido, *no depende de la aplicación característica escogida ϕ*. Esta semimétrica nos permite tratar la continuidad de k. Nos volvemos a centrar en el caso de núcleos con valores reales.

Lema 1.4.17 (**Caracterización de núcleos continuos**)**.** *Sea (X, τ) un espacio topológico y k un núcleo en X, para el que escogemos espacio y aplicación característicos H y ϕ arbitrarios. Entonces, las siguientes afirmaciones son equivalentes:*

a) k *es continuo.*

b) k *es separadamente continuo y* $x \mapsto k(x, x)$ *es continua.*

c) $\mathrm{id} : (X, \tau) \to (X, d_k)$ *es continua.*

d) $\phi : (X, \tau) \to H$ *es continua.*

Demostración. La primera implicación a) $\Longrightarrow$ b) es trivial.
En cuanto a b) $\Longrightarrow$ c), empecemos viendo que $d_k(\cdot, x) : (X, \tau) \to \mathbb{R}$ es continua, para un $x \in X$ fijo. Por la igualdad (1.4.33), esto es inmediato si suponemos que $x \mapsto k(x, x)$ es continua y que k es separadamente continua. Entonces, el conjunto $\{x' \in X : d_k(x', x) < \varepsilon\}$ es abierto con respecto a la topología τ, con lo que $\mathrm{id} : (X, \tau) \to (X, d_k)$ es continua, tal y como queríamos probar.
Para ver c) $\Longrightarrow$ d), basta observar que $\phi : (X, d_k) \to H$ es continua, por definición de d_k, con lo que la composición de esta aplicación con $\mathrm{id} : (X, \tau) \to (X, d_k)$, continua por hipótesis, nos da la continuidad de ϕ sobre la topología τ de X.
Por último, para ver d) $\Longrightarrow$ a), fijemos $x, y \in X$, así como $x', y' \in X$. Entonces, por definición de núcleo y por la desigualdad triangular, se tiene

$$\begin{aligned} |k(x, x') - k(y, y')| &\leq |\langle \phi(x'), \phi(x) - \phi(y) \rangle| + |\langle \phi(x') - \phi(y'), \phi(y) \rangle| \\ &\leq ||\phi(x')|| \cdot ||\phi(x) - \phi(y)|| + ||\phi(y)|| \cdot ||\phi(x') - \phi(y')||, \end{aligned}$$

donde también se ha usado la desigualdad de Cauchy-Schwarz, y la bilinealidad del producto interno. Entonces, la continuidad de ϕ implica la de k, tomando las pertinentes acotaciones en entornos de los puntos involucrados. □

1.5. Teorema de representación de Mercer

De cara a formalizar la representación de un núcleo, o de los elementos del RKHS con dicho núcleo como núcleo reproductor, en términos de una serie y en un contexto de núcleos continuos y espacios caracterísitos compactos con una cierta medida, daremos la versión adecuada del teorema de Mercer. En todo este apartado, nos limitamos a núcleos con valores reales.
A modo de preliminares, consideremos un espacio de medida (X, μ) con medida σ-finita μ, y k un núcleo continuo en X con $||k||_{L^2(\mu)} < \infty$ (recordemos, por el lema 1.4.3, que $k(\cdot, x) : X \to \mathbb{R}$ es medible en X para todo $x \in X$ si y solo si cualquier $f \in H$ es medible en X). Suponemos también que el RKHS asociado al núcleo k es separable. Así pues, por los teoremas 1.4.5 y 1.4.15, tenemos la factorización dada en el diagrama que sigue

$$\begin{tikzcd} L^2(\mu) \arrow[rr, "T_k"] \arrow[dr, "S_k"'] & & L^2(\mu) \\ & H \arrow[ur, "S_k^*"'] & \end{tikzcd}$$

Además, se cumplen todas las hipótesis del teorema 1.4.15, luego sabemos que $T_k = S_k^* S_k$ es positivo, autoadjunto y compacto, por lo que, por el teorema espectral dado en 1.4.8, existe una familia, a lo sumo infinita numerable, y ortonormal $(e_i)_{i \in I} \subset L^2(\mu)$, y una familia $(\lambda_i)_{i \in I} \subset \mathbb{R}$ convergente a cero si es infinita, que cumple que:

i) $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_i| \geq \cdots > 0$.

ii) $T_k f = \sum_{i \in I} \lambda_i \langle f, e_i \rangle e_i$, para cualquier elemento $f \in L^2(\mu)$.

iii) $\{\lambda_i : i \in I\}$ es el conjunto de autovalores no nulos de T_k.

A continuación, definimos $\tilde{e}_i := \lambda_i^{-1} S_k e_i$, que es un elemento de H (véase el diagrama anterior). Si aplicamos el operador S_k^*, el cual no es más que la inclusión $\mathrm{id} : H \to L^2(\mu)$ ($(S_k^*)^* = \mathrm{id}$), tenemos la igualdad

$$S_k^* \tilde{e}_i = \tilde{e}_i = \lambda_i^{-1} S_k^* S_k e_i = \lambda_i^{-1} T_k e_i. \tag{1.5.1}$$

Ahora, usamos de nuevo el teorema espectral sobre el operador compacto, positivo y autoadjunto T_k, que va de $L^2(\mu)$ en $L^2(\mu)$, de forma que

$$\lambda_i^{-1} T_k e_i = \lambda_i^{-1} [\sum_{j \in I} \lambda_j \langle e_i, e_j \rangle e_j] = \lambda_i^{-1} [\sum_{j \in I} \lambda_j \delta_{ij} e_j] = e_i. \tag{1.5.2}$$

Entonces, tenemos la igualdad $e_i = \lambda_i^{-1} T_k e_i = \tilde{e}_i = \lambda_i^{-1} S_k e_i$, donde la última igualdad se sigue de la definición de $\tilde{e}_i$. De esta forma, tiene sentido considerar e_i como un elemento de H y se tiene que $S_k e_i = \lambda_i e_i$ para todo $i \in I$.
De lo anterior se sigue la siguiente cadena de igualdades

$$\lambda_i \lambda_j \langle e_i, e_j \rangle_H = \langle S_k e_i, S_k e_j \rangle_{L^2(\mu)} = \langle e_i, S_k^* S_k e_j \rangle_{L^2(\mu)} = \langle e_i, T_k e_j \rangle_{L^2(\mu)} = \lambda_j \langle e_i, e_j \rangle_{L^2(\mu)}. \tag{1.5.3}$$

Es decir, la familia $(\sqrt{\lambda_i} e_i)_{i \in I}$ es un conjunto ortonormal de H, por serlo $(e_i)_{i \in I}$ en $L^2(\mu)$. Dicho esto, el teorema de representación que vamos a dar en este contexto nos dice que esta familia no solo es ortonormal, sino que además es una base. Para formalizar esto, necesitamos primero el teorema de Mercer original, dado en términos de núcleos simétricos y continuos, para después generalizarlo para los elementos de un RKHS a partir de la reproductividad de su núcleo.
Recordemos que $S_k : L^2(\mu) \to H$ es el operador adjunto de la inclusión de nuestro RKHS de partida H en $L^2(\mu)$, así como que $T_k = S_k^* S_k$ es autoadjunto, compacto y positivo.
Empezamos enunciando el teorema de Tonelli-Fubini sobre funciones definidas sobre un procucto de dos espacios de medidas σ-finitas, como resultado conocido y necesario para la demostración del teorema de Mercer para espacios métricos y compactos 1.5.2.

Teorema 1.5.1 (Teorema de Tonelli-Fubini)**.** *Sean (X, A, μ) y (Y, C, ν) dos espacios de medida con medidas σ-finitas, y sea $f : X \times Y \to [0, \infty)$ una función medible sobre el producto de espacios de medida, tal y como se formaliza en el apéndice B. Entonces*

$$\int_{X \times Y} f d(\mu \otimes \nu) = \int_Y \int_X f(x,y) d\mu(x) d\nu(y) = \int_X \int_Y f(x,y) d\nu(y) d\mu(x), \tag{1.5.4}$$

donde las integrales internas

$$\int_X f(x,y) d\mu(x), \quad \int_Y f(x,y) d\nu(y), \tag{1.5.5}$$

son medibles en Y y en X, respectivamente.
Además, en el caso de que $f \in \mathcal{L}^1(\mu \otimes \nu)$, tenemos las mismas igualdades, con la salvedad de que las integrales internas anteriores solo están definidas casi siempre en Y y X, respectivamente.
Este resultado es generalizable para funciones no necesariamente positivas $f : X \times Y \to \mathbb{R}$ a través del estudio de sus partes positiva y negativa, $f = f_+ - f_-$.

Dado un espacio de medida $(\Omega, \mathcal{F}, \mu)$, con Ω el espacio sobre el que trabajamos, $\mathcal{F}$ una σ-álgebra sobre Ω, y μ la medida, el *soporte de la medida* μ, que denotamos por $\mathrm{sop}(\mu)$, es el menor cerrado C en la topología de X tal que $\mu(\Omega \setminus C) = 0$.

Teorema 1.5.2 (Teorema de Mercer para espacios métricos y compactos)**.** *Sea X un espacio métrico y compacto, y sea $k : X \times X \to \mathbb{R}$ un núcleo continuo. Además, sea μ una medida finita de Borel en* $\mathrm{sop}(\mu) = X$*. Entonces, las familias $(e_i)_{i \in I}$ y $(\lambda_i)_{i \in I}$ dadas como arriba satisfacen*

$$k(x, x') = \sum_{i \in I} \lambda_i e_i(x) e_i(x'), \qquad \text{para todo } (x, x') \in X \times X, \tag{1.5.6}$$

donde la convergencia se da en $L^2(\tilde{\mu})$*, y donde* $\tilde{\mu}$ *denota la medida del espacio de medida* $(X\times X, \tilde{\mu}, \mathcal{G})$*, con* $\mathcal{G}$ *la* σ*-álgebra engendrada por los productos* $A\times B$*, para todo* A *y* B *en la* σ*-álgebra original del espacio de medida dado por* μ *y* $X = \operatorname{sop}(\mu)$*.*

Demostración. Se tiene la medida finita μ de Borel en el espacio métrico X, con la que construimos el subsiguiente espacio producto $X\times X$. Denotamos

$$\phi_i(x,x') = e_i(x)e_i(x'), \qquad x, x' \in X. \tag{1.5.7}$$

También, escribimos

$$S(x,y) = \sum_i \langle k, \phi_i\rangle_{L^2(\tilde{\mu})}\phi_i(x,y). \tag{1.5.8}$$

Observemos que aún no hemos dicho nada sobre la convergencia de la serie denotada por S.
Nótese que $\{\phi_i\}_i$ es un sistema ortonormal en el espacio $L^2(\tilde{\mu})$. Para ello, nótese que

$$\langle \phi_i, \phi_j\rangle_{L^2(\tilde{\mu})} = \int_{X\times X} \phi_i(x,y)\phi_j(x,y)d\tilde{\mu}(x,y) = \int_{X\times X} e_i(x)e_i(y)e_j(x)e_j(y)d\tilde{\mu}(x,y). \tag{1.5.9}$$

Para poder escindir esta integral vía el teorema de Tonelli-Fubini 1.5.1, notemos que $x\in X\mapsto e_i(x)e_j(x)$ está en $L^1(\mu)$. Esto es consecuencia de la desigualdad de Hölder y de la pertenencia de los e_i en $L^2(\mu)$, ya que

$$\int_X |e_i(x)e_j(x)|d\mu(x) \leq \left[\int_X (e_i(x))^2\, d\mu(x)\right]^{1/2} \cdot \left[\int_X (e_j(x))^2\, d\mu(x)\right]^{1/2} < \infty. \tag{1.5.10}$$

Por lo tanto, $x\in X\mapsto e_i(x)e_j(x)$ es integrable con respecto a μ en X. Entonces, el integrando dado por

$$\phi_i(x,y)\phi_j(x,y) = e_i(x)e_j(x)e_i(y)e_j(y), \tag{1.5.11}$$

es producto tensorial de funciones en $L^1(\mu)$, con lo que estamos en condiciones de aplicar 1.5.1 en (1.5.9), de forma que

$$\langle \phi_i, \phi_j\rangle_{L^2(\tilde{\mu})} = \left(\int_X e_i(x)e_j(x)d\mu(x)\right)\left(\int_X e_i(y)e_j(y)d\mu(y)\right) = \delta_{ij}^2 = \delta_{ij}, \tag{1.5.12}$$

donde δ_{ij} denota la delta de Kronecker, y donde se ha utilizado la ortonormalidad de partida de la base dada en $L^2(\mu)$.
La completitud de este sistema se puede probar viendo que si $h\in L^2(\tilde{\mu})$ es un elemento ortogonal a todos los ϕ_i con respecto al producto interno de $L^2(\tilde{\mu})$, entonces necesariamente $h=0$. Sea pues tal $h\in L^2(\tilde{\mu})$. Entonces:

$$0 = \langle h, \phi_i\rangle_{L^2(\tilde{\mu})} = \int_{X\times X} h(x,y)\phi_i(x,y)d\tilde{\mu}(x,y) = \int_{X\times X} h(x,y)e_i(x)e_i(y)d\tilde{\mu}(x,y), \tag{1.5.13}$$

donde, aplicando el teorema de Tonelli-Fubini, aunque ahora pidiendo solo medibilidad sobre el integrando $h(x,y)e_i(x)e_i(y)$, se tiene

$$\int_X e_i(x)\left(\int_X e_i(y)h(x,y)d\mu(y)\right)d\mu(x) = 0, \qquad i\in I. \tag{1.5.14}$$

De esto, utilizando el teorema de anulación sobre la medida μ en X, tenemos que:

$$\int_X e_i(y)h(x,y)d\mu(y) = 0, \qquad i\in I, \quad \text{para casi todo } x\in X, \tag{1.5.15}$$

y aplicándolo otra vez, $h(x,y) = 0$, para casi todos $x, y \in X$, donde la locución *casi todo* hace referencia a la medida original μ. Entonces, $\{\phi_i\}_i$ forma una base ortonormal en $L^2(\tilde{\mu})$, con lo que la serie $S(x,y)$ dada en (1.5.8) converge en $L^2(\tilde{\mu})$, dada la completitud del sistema $\{\phi_i\}$ en $L^2(\tilde{\mu})$. Notemos que

$$\langle k, \phi_i \rangle_{L^2(\tilde{\mu})} = \int_{X\times X} k(x,y) e_i(x) e_i(y) d\tilde{\mu}(x,y), \tag{1.5.16}$$

sobre lo que aplicamos el teorema de Fubini, por idéntico razonamiento al expuesto sobre la función h antes dada,

$$\begin{aligned}\langle k, \phi_i \rangle_{L^2(\tilde{\mu})} &= \int_X e_i(x) \left(\int_x e_i(y) k(x,y) d\mu(y) \right) d\mu(x) \\ &= \int_X e_i(x) T_k e_i(x) d\mu(x) = \langle T_k e_i, e_i \rangle_{L^2(\mu)} = \lambda_i \delta_{ii} = \lambda_i.\end{aligned}$$

En la segunda igualdad hemos utilizado la definición de $T_k : L^2(\mu) \to L^2(\mu)$, como la composición de un operador integral con el núcleo k en el integrando, y la inclusión de H en $L^2(\mu)$, con autovalores λ_i sobre los autovectores $e_i \in L^2(\mu)$. Por lo tanto,

$$S(x,y) = \sum_i \lambda_i \phi_i(x,y). \tag{1.5.17}$$

De nuevo, por el teorema de Tonelli-Fubini, dadas $f, g \in L^2(\mu)$, tenemos que $F(x,y) = f(x)g(y) \in L^2(\tilde{\mu})$. Además,

$$\langle S, F \rangle_{L^2(\tilde{\mu})} = \sum_i \lambda_i \langle \phi_i, F \rangle. \tag{1.5.18}$$

Ahora, utilizando de nuevo la estructura de la aplicación T_k, se tiene que:

$$\langle k, F \rangle_{L^2(\tilde{\mu})} = \int_{X\times X} k(x,y) f(x) g(y) d\tilde{\mu}(x,y) = \int_X g(y) \left(\int_X k(x,y) f(x) d\mu(x) \right) d\mu(y) = \langle T_k f, g \rangle_{L^2(\mu)}, \tag{1.5.19}$$

sobre lo que, utilizando la condición de completitud de la base $\{\phi_i\}$ de $L^2(\tilde{\mu})$, así como el hecho de que T_k es un operador autoadjunto, tenemos que

$$\begin{aligned}\langle k, F \rangle_{L^2(\tilde{\mu})} &= \sum_i \langle T_k f, e_i \rangle_{L^2(\mu)} \langle e_i, g \rangle_{L^2(\mu)} = \sum_i \langle f, T_k e_i \rangle_{L^2(\mu)} \langle e_i, g \rangle_{L^2(\mu)} \\ &= \sum_i \lambda_i \langle f, e_i \rangle_{L^2(\mu)} \langle e_i, g \rangle_{L^2(\mu)}.\end{aligned}$$

Por el teorema de Fubini sobre el producto tensorial de funciones de cuadrado integrable que sigue, obtenemos

$$\langle f, e_i \rangle_{L^2(\mu)} \langle e_i, g \rangle_{L^2(\mu)} = \left(\int_X f(x) e_i(x) d\mu(x) \right) \left(\int_X f(y) e_i(y) d\mu(y) \right) = \int_{X\times X} F(x,y) \phi_i(x,y) d\tilde{\mu}(x,y), \tag{1.5.20}$$

usando, en la última igualdad, la definición de F y de ϕ_i. Así, como acabamos de probar que

$$\langle f, e_i \rangle_{L^2(\mu)} \langle e_i, g \rangle_{L^2(\mu)} = \langle \phi_i, F \rangle_{L^2(\tilde{\mu})}, \tag{1.5.21}$$

entonces

$$\langle k, F \rangle_{L^2(\tilde{\mu})} = \sum_i \lambda_i \langle \phi_i, F \rangle_{L^2(\tilde{\mu})}. \tag{1.5.22}$$

La última expresión se corresponde, por lo ya calculado, con $\langle S, F \rangle_{L^2(\tilde{\mu})}$, de forma que tenemos la igualdad $\langle k, F \rangle_{L^2(\tilde{\mu})} = \langle S, F \rangle_{L^2(\tilde{\mu})}$, lo que implica que $k - S$ es ortogonal a toda función de la forma de F, en particular a todos los elementos de la base ortonormal $\{\phi_i\}_i$ de $L^2(\tilde{\mu})$. Entonces, $k(x,x') = S(x,x')$ para $\tilde{\mu}$-casi todo $(x,x') \in X \times X$. Así, llegamos a la igualdad que buscábamos, de la forma

$$k(x,x') = \sum_{i\in I} \lambda_i e_i(x) e_i(x'), \qquad \text{para todo } (x,x') \in X \times X. \tag{1.5.23}$$

Notemos que la convergencia se da en $L^2(\tilde{\mu})$, pues el sistema $\{\phi_i\}$ es completo en $L^2(\tilde{\mu})$. □

Nótese que tal y como viene dado el núcleo en (1.5.6), y dado el lema 1.1.2 del inicio de este trabajo, podemos afirmar que la aplicación $\phi : X \to l^2$, dada por $\phi(x) := (\sqrt{\lambda_i} e_i(x))_{i\in I}$ para todo $x \in X$, es una aplicación característica del núcleo k.
Recordemos que nuestro objetivo en este momento es probar que $(e_i)_{i\in I}$ es una base ortonormal de un RKHS H sobre un espacio métrico y compacto X arbitrario. Antes de dar el resultado que nos lo garantice, necesitamos el siguiente corolario.

Corolario 1.5.3. *Con las hipótesis y la notación del teorema 1.5.2, la serie dada por $\sum_{i\in I} a_i \sqrt{\lambda_i} e_i(x)$ converge absoluta y uniformemente para todo $(a_i) \in l^2(I)$, donde I es un conjunto a lo sumo infinito numerable.*

Demostración. Sea $x \in X$. Entonces, usando la desigualdad de Hölder en $l^2(I)$, y el teorema de Mercer 1.5.2, tenemos las siguientes desigualdades:

$$\sum_{i\in I} |a_i \sqrt{\lambda_i} e_i(x)| \leq \left(\sum_{i\in I} a_i^2\right)^{\frac{1}{2}} \left(\sum_{i\in I} \lambda_i e_i^2(x)\right)^{\frac{1}{2}} = ||(a_i)_{i\in I}||_{l^2} \cdot \sqrt{k(x,x)}, \tag{1.5.24}$$

donde, en la última igualdad, hemos usado la igualdad del teorema 1.5.2, para el mismo argumento x en ambas entradas de k.
Notemos que k es un núcleo continuo que toma valores reales, y que parte de un espacio compacto. Entonces, es necesariamente acotado, de lo cual se sigue que $||k||_\infty = \sup_{x\in X} \sqrt{|k(x,x)|} < M$, para $M > 0$ constante. Por esto, es claro que la serie propuesta converge absolutamente.
En cuanto a la convergencia uniforme, el razonamiento es el siguiente: usamos otra vez la desigualdad de Hölder en $l^2(I)$ sobre (a_i) y sobre $\sqrt{\lambda_i} e_i(x)$, para llegar a

$$\sum_{i\in I} |a_i \sqrt{\lambda_i} e_i(x)| \leq \left(\sum_{i\in I} a_i^2\right)^{\frac{1}{2}} \left(\sum_{i\in I} \lambda_i e_i^2(x)\right)^{\frac{1}{2}}, \quad x \in X. \tag{1.5.25}$$

Notemos que id : $H \to L^\infty(X)$ es continua, ya que k es acotado. Entonces, la convergencia en H implica convergencia en $L^\infty(X)$, es decir, implica convergencia uniforme. Por tanto, como la convergencia de la serie

$$\sum_{i\in I} \lambda_i e_i^2(x), \tag{1.5.26}$$

se da en H (por la identificación de los espacios H y $L^2(\mu)$ en términos de estos elementos de bases ortonormales de $L^2(\mu)$), entonces dicha serie converge uniformemente.
Además, como k es continua, entonces todos los elementos de su RKHS asociado son continuos, en particular los e_i. Luego la serie (1.5.26) es el límite uniforme de una sucesión de funciones continuas en un compacto X, con lo cual dicha serie es continua. Entonces, por ser X compacto y esta serie continua, la podemos acotar por un $Q > 0$, con lo que la convergencia uniforme de la serie propuesta en este corolario se sigue de la convergencia de la serie que se deriva de la norma $||\{a_i\}||_{l^2(I)}$. □

Ahora, damos el resultado de representación de RKHSs con núcleos continuos sobre espacios métricos compactos X, que da sentido a todo el desarrollo de este apartado.

Teorema 1.5.4 (Teorema de representación de Mercer sobre RKHS). *Con las hipótesis del teorema 1.5.2, definimos*

$$H := \left\{ \sum_{i \in I} a_i \sqrt{\lambda_i} e_i : (a_i) \in l^2(I) \right\}. \tag{1.5.27}$$

Para $f := \sum_{i \in I} a_i \sqrt{\lambda_i} e_i \in H$ *y* $g := \sum_{i \in I} b_i \sqrt{\lambda_i} e_i \in H$*, definimos también*

$$\langle f, g \rangle_H = \sum_{i \in I} a_i b_i. \tag{1.5.28}$$

Entonces H*, equipado con el producto interno* $\langle \cdot, \cdot \rangle_H$*, es el RKHS con el núcleo* k*. Además, el operador* $T_k^{\frac{1}{2}} : L^2(\mu) \to H$ *es un isomorfismo isométrico.*

Demostración. El hecho de que $\langle \cdot, \cdot \rangle_H$ sea un producto interno es claro por la forma que esta operación toma. Además, H es un espacio de funciones del espacio métrico y compacto X a $\mathbb{K}$, las cuales están bien definidas, en términos de convergencia uniforme y absoluta, en virtud del corolario 1.5.3. Para ver que H es efectivamente de Hilbert, basta tener en cuenta que los valores λ_i y las funciones e_i vienen dadas, y lo que cambia para cada elemento de H son los coeficientes $(a_i)_{i \in I} \in l^2(I)$. De esta forma, es la condición de espacio de Hilbert de este último espacio $l^2(I)$ de la que se deriva que H lo sea también. Denotando directamente por $|| \cdot ||$ a la norma H inducida por su producto escalar, tenemos, dados $f := \sum_{i \in I} a_i \sqrt{\lambda_i} e_i \in H$ y $g := \sum_{i \in I} b_i \sqrt{\lambda_i} e_i \in H$, lo que sigue

$$||f - g||^2 = || \sum_{i \in I} (a_i - b_i) \sqrt{\lambda_i} e_i ||^2 = \sum_{i \in I} (a_i - b_i)^2 = ||(a_i)_{i \in I} - (b_i)_{i \in I}||^2_{l^2(I)}, \tag{1.5.29}$$

donde se ha usado la definición del producto interno de H. De esta forma, como $l^2(I)$ es de Hilbert, entonces H también lo es.
Fijemos ahora $x \in X$. El desarrollo seguido en la demostración del teorema de Mercer 1.5.2 implica que la función que sigue converge en $L^2(\mu)$

$$k(\cdot, x) = \sum_{i \in I} \sqrt{\lambda_i} e_i(x) \sqrt{\lambda_i} e_i(\cdot). \tag{1.5.30}$$

Ya vimos anteriormente que $\phi : X \to l^2(I)$ dado por $\phi(x) = (\sqrt{\lambda_i} e_i(x))_{i \in I}$ está bien definida, ya que, en virtud del teorema 1.5.2,

$$||(\sqrt{\lambda_i} e_i(x))_{i \in I}||^2_{l^2(I)} = \sum_{i \in I} \lambda_i e_i^2(x) = k(x, x) < \infty. \tag{1.5.31}$$

Sea $x \in X$. La expresión (1.5.31) implica que $k(\cdot, x) \in H$, tomando $a_i = e_i(x)$ para todo $i \in I$. Sea ahora $f := \sum_{i \in I} b_i \sqrt{\lambda_i} e_i \in H$, entonces

$$\langle f, k(\cdot, x) \rangle_H = \sum_{i \in I} b_i \sqrt{\lambda_i} e_i(x) = f(x), \tag{1.5.32}$$

lo cual demuestra la propiedad reproductora de k en H, es decir, H es un RKHS con núcleo asociado k.
Centrémonos ahora en el operador $T_k^{\frac{1}{2}} : L^2(\mu) \to H$. Empezamos fijando un elemento $f \in L^2(\mu)$. Tal y como lo construimos desde un principio, la familia $(e_i)_{i \in I}$ es una familia ortonormal de elementos de $L^2(\mu)$, luego se tiene que:

$$f = \sum_{i \in I} \langle f, e_i \rangle_{L^2(\mu)} e_i, \tag{1.5.33}$$

con la convergencia en el espacio $L^2(\mu)$. Recordemos que $(e_i)_{i\in I}$ no solo es una familia ortonormal en $L^2(\mu)$, sino que además cada e_i es un autovector de la aplicación $T_k = S_k^* S_k : L^2(\mu) \to L^2(\mu)$, con autovalores λ_i. Por lo tanto, la imagen por T_k de f es la que sigue

$$T_k^{\frac{1}{2}} f = \sum_{i\in I} \langle f, e_i \rangle_{L^2(\mu)} \sqrt{\lambda_i} e_i, \tag{1.5.34}$$

donde, otra vez, la convergencia se da con respecto a la norma en $L^2(\mu)$. Por la identidad de Parseval sobre la familia ortonormal $(e_i)_{i\in I} \subset L^2(\mu)$ para el elemento $f \in L^2(\mu)$, tenemos que $(\langle f, e_i \rangle_{L^2(\mu)})_{i\in I} \in l^2(I)$, luego por construcción del RKHS H se tiene que $T_k^{\frac{1}{2}} f \in H$ para todo $f \in L^2(\mu)$. Esto es, la aplicación $T_k^{\frac{1}{2}}$, tal y como la hemos dado, está bien definida.
Ahora, usamos el corolario 1.5.3 para afirmar que la convergencia de la serie dada en (1.5.34) es absoluta y uniforme en X. Además, por este mismo corolario, y por cómo hemos definido el producto interno en H, se sigue que

$$||T_k^{\frac{1}{2}} f||_H^2 = \sum_{i\in I} |\langle f, e_i \rangle_{L^2(\mu)}|^2 = ||(\langle f, e_i \rangle_{L^2(\mu)})_{i\in I}||_{l^2(I)}^2. \tag{1.5.35}$$

En la última igualdad, hemos usado la identidad de Parseval. Acabamos de demostrar que $T_k : L^2(\mu) \to H$ es isométrico. Aún nos falta demostrar que es un isomorfismo.
Para probar la sobreyectividad, fijemos un elemento $f \in H$. Por definición de H, existe un elemento $(a_i)_{i\in I} \in l^2(I)$ tal que:

$$f(x) = \sum_{i\in I} a_i \sqrt{\lambda_i} e_i(x), \tag{1.5.36}$$

para todo $x \in X$. Tomemos el elemento $g := \sum_{i\in I} a_i e_i \in L^2(\mu)$, cuya convergencia se da en el espacio $L^2(\mu)$. Entonces, dada la familia ortonormal $(e_i)_{i\in I}$, tenemos $a_i = \langle g, e_i \rangle_{L^2(\mu)}$, para todo $i \in I$. Además, la convergencia en (1.5.34) es, en particular, puntual (pues ya hemos demostrado que es uniforme), luego para un $x \in X$ arbitrario, se tiene

$$T_k^{\frac{1}{2}} g(x) = \sum_{i\in I} \langle g, e_i \rangle_{L^2(\mu)} \sqrt{\lambda_i} e_i(x) = \sum_{i\in I} a_i \sqrt{\lambda_i} e_i(x) = f(x). \tag{1.5.37}$$

Por último, la inyectividad de $T_k^{\frac{1}{2}}$ se sigue de la construcción hecha para probar la sobreyectividad, ya que, dado un $f \in H$ cualquiera de la forma (1.5.36), tomábamos un $f' \in L^2(\mu)$ tal que $\langle f', e_i \rangle = a_i$, para todo $i \in I$. Entonces, dado cualquier otro $g \in H$ que cumpla estas últimas igualdades tendrá los mismos coeficientes a_i, con lo que le corresponderá la misma clase f' en $L^2(\mu)$. $\square$

Capítulo 2

Teoría del muestreo. Grandes ejemplos de RKHS

La *teoría del muestreo* trata de la reconstrucción de funciones a partir de sus valores (muestras) en una secuencia apropiada de puntos, mediante expansiones (sumas o integrales) en que estos valores están involucrados. Téngase en cuenta que esto no es siempre posible. Pensemos, por ejemplo, en una función continua en la recta real $\mathbb{R}$. Sin ninguna otra imposición, no podemos garantizar que f esté completamente determinada por una sucesión numerable de valores $\{f(t_n)\}_{n=1}^{\infty}$, dados sobre una sucesión $(t_n)_{n=1}^{\infty}$. Necesitamos más restricciones sobre f o sobre la sucesión. O dicho de otra forma, las funciones f objeto de nuestro estudio deben pertenecer a cierto tipo de espacios. Uno de ellos, y este es en el que nosotros nos detendremos, es el espacio de Hilbert con núcleo reproductor.

2.1. Teorema del muestreo

Consideremos un RKHS H sobre un cuerpo $\mathbb{K}$, dado por funciones $f : X \to \mathbb{K}$, con X un conjunto no vacío arbitrario. Recordemos que la definición de estos espacios nos decía que el funcional evaluación $f \in H \mapsto f(x) \in \mathbb{K}$ es acotado para cualquier $x \in X$, lo que, vía el teorema de representación de Riesz, traducíamos en que, para cualquier $x \in X$, existe una única función $k_x \in H$ tal que $f(x) = \langle f, k_x \rangle_H$, con $\langle \cdot, . \rangle_H$ el producto interno sobre H. Nos restringimos a espacios de Hilbert separables, sobre los que, como es bien sabido, tendremos bases ortonormales numerables.
El caso que a nosotros nos interesa es aquel en que la sucesión $\{k_{t_n}\}_{n=1}^{\infty}$ forme una base para un RKHS H, de modo que podemos obtener una fórmula para un muestreo en H que permita 'reconstruir' funciones $f \in H$ a partir de un número a lo sumo infinito numerable de valores de X. Esta tendría la forma:

$$f(t) = \sum_{n=1}^{\infty} \langle f, k_t \rangle_H S_n(t), \quad t \in X, \tag{2.1.1}$$

con $S_n(t)$ funciones que están por definir, y donde se ha usado que la convergencia de sucesiones en un RKHS H con su norma implica la convergencia puntual de dichas sucesiones como funciones definidas en X.
Solo en términos de bases ortonormales, enunciamos el teorema del muestreo en RKHSs.

Teorema 2.1.1 (Teorema de muestreo en RKHS)**.** *Sea H un RKHS separable de funciones definidas en un conjunto X con un núcleo reproductor $k : X \times X \to \mathbb{K}$, con $\mathbb{K}$ el cuerpo sobre el que está construido H. Supongamos que existe una sucesión $\{t_n\}_{n=1}^{\infty} \subset X$ tal que $\{k(\cdot, t_n)\}_{n=1}^{\infty}$ es una base ortogonal en H. Entonces, cualquier $f \in H$ se puede representar como:*

$$f(t) = \sum_{n=1}^{\infty} f(t_n) \frac{k(t, t_n)}{k(t_n, t_n)}, \quad t \in X, \tag{2.1.2}$$

con convergencia uniforme en subconjuntos de X donde la función $t \mapsto k(t,t)$ es acotada[1].

Demostración. El resultado se sigue directamente de la expansión de $f \in H$ en la base ortonormal $\{k(\cdot,t_n)/\sqrt{k(t_n,t_n)}\}_{n=1}^{\infty}$, i.e., dado $f \in H$, tenemos que

$$f = \sum_{n=1}^{\infty} \langle f, \frac{k(\cdot,t_n)}{\sqrt{k(t_n,t_n)}} \rangle_H \frac{k(\cdot,t_n)}{\sqrt{k(t_n,t_n)}} = \sum_{n=1}^{\infty} f(t_n) \frac{k(\cdot,t_n)}{k(t_n,t_n)}, \quad \text{en } H. \tag{2.1.3}$$

Por el teorema 1.4.2, en todas aquellas regiones de X en que k sea acotado tendremos que la inclusión

$$\mathrm{id} : H \to L^{\infty}(X) \tag{2.1.4}$$

es continua, y por lo tanto secuencialmente continua. Como la convergencia de la serie (2.1.3) se da en la norma de H, y como la aplicación id anterior es secuencialmente continua, de la convergencia en H de dicha serie se sigue la convergencia uniforme de esta misma serie sobre las regiones de X en que k sea un núcleo acotado. □

2.2. Algunos ejemplos. Aplicación del teorema del muestreo

Damos ahora un par de ejemplos de RKHSs no tan triviales como los estudiados hasta ahora, en que la aplicación de este teorema 2.1.1 es directa.

2.2.1. Polinomios trigonométricos

Empezamos definiendo lo que es un polinomio trigonométrico de grado $\leq N$, para un $N \in \mathbb{N}$. Se denomina *polinomio trigonométrico* a toda función $P : \mathbb{R} \to \mathbb{C}$ de la forma:

$$P(t) = \sum_{k=-N}^{N} a_k e^{ikt}, \tag{2.2.1}$$

con $a_k \in \mathbb{C}$, $a_k \neq 0$, $N \in \mathbb{N}$. Denotamos por $\mathcal{H}_N$ al conjunto de tales polinomios trigonométricos de grado $\leq N$.
Definimos sobre este espacio el producto interno estándar restringido del espacio $L^2[-\pi,\pi]$. Como $\mathcal{H}_N$ es un subespacio de dimensión finita en $L^2[-\pi,\pi]$, entonces es cerrado en este espacio. Tal y como está definido el espacio, y con este mismo producto interno, tenemos que $\mathcal{H}_N$ tiene una base ortonormal formada por exponenciales complejas, de la forma $\{e^{ikt}/\sqrt{2\pi}\}_{k=-N}^{N}$. Como $\mathcal{H}_N$ es un espacio normado con dimensión finita, ya vimos en un ejemplo anterior que entonces $\mathcal{H}_N$ es RKHS. Entonces, podemos aplicar la expresión (1.3.6) del teorema 1.3.4, por el cual la siguiente función conforma el núcleo reproductor de $\mathcal{H}_N$:

$$k_N(t,s) = \frac{1}{2\pi} \sum_{k=-N}^{N} e^{ik(t-s)} = \frac{1}{2\pi} D_N(t-s), \tag{2.2.2}$$

donde D_N es el llamado **núcleo de Dirichlet** N-ésimo, definido como:

$$D_N(t) := \sum_{k=-N}^{N} e^{ikt} = \begin{cases} \frac{\sin[(N+\frac{1}{2})t]}{\sin(\frac{t}{2})} & \text{si} \quad t \in \mathbb{R} \setminus 2\pi\mathbb{Z}, \\ \\ 2N+1 & \text{si} \quad t \in 2\pi\mathbb{Z}. \end{cases} \tag{2.2.3}$$

[1]Ya probamos que la acotación de la función $t \mapsto k(t,t)$ equivale a la acotación general del núcleo k.

Consideramos la sucesión finita de puntos $s_n = \frac{2\pi n}{2N+1} \in [-\pi, \pi]$, $-N \leq n \leq N$, y con ella la sucesión de funciones en $\mathcal{H}_N$ de la forma $\{k_N(\cdot, s_n)\}_{n=-N}^{N}$. Esta última, aunque no esté normalizada, forma una base ortogonal en $\mathcal{H}_N$, pues:

$$\langle k_N(\cdot, s_n), k_N(\cdot, s_m)\rangle_{L^2[-\pi,\pi]} = k_N(s_m, s_n) = \frac{1}{2\pi} \frac{\sin(\pi(m-n))}{\sin(\frac{\pi(m-n)}{2N+1})} = \frac{2N+1}{2\pi}\delta_{mn}. \tag{2.2.4}$$

Aplicamos directamente el teorema de muestreo sobre el RKHS $\mathcal{H}_N$ y sobre este conjunto ortogonal $\{k(\cdot, s_n)\}_{n=-N}^{N}$, para obtener una representación de cada uno de los elementos P de $\mathcal{H}_N$ como sigue

$$P(t) = \frac{1}{2N+1} \sum_{n=-N}^{N} P\left(\frac{2\pi n}{2N+1}\right) \frac{\sin(\frac{2N+1}{2})(t - \frac{2\pi n}{2N+1})}{\sin(\frac{1}{2}(t - \frac{2\pi n}{2N+1}))}, \quad t \in [-\pi, \pi). \tag{2.2.5}$$

2.2.2. Polinomios ortogonales

Este es un ejemplo clásico de núcleo reproductor. Dada una sucesión infinita de polinomios ortogonales $\{P_n\}_{n=0}^{\infty}$ sobre un intervalo I de la recta real $\mathbb{R}$, respecto al producto interno definido sobre la integración en medida de Lebesgue $\langle P, Q\rangle = \int_I P(x)Q(x)\alpha(x)dx$, con un cierto peso $\alpha : I \to \mathbb{R}$, $\alpha(x) > 0$ para todo $x \in \mathbb{R}$, podemos construir espacios de Hilbert a través de los subespacios de $L^2(I)$ generados por subfamilias finitas de estos polinomios, $\{P_n\}_{n=0}^{N}$, para algún $N \in \mathbb{N}$.
Particularizamos para la familia de los conocidos polinomios de Legendre. Ésta, como es bien sabido de la teoría de polinomios ortogonales, es la única familia, salvo constante, de polinomios ortogonales sobre el intervalo $[-1, 1]$ (esto es, sobre el espacio $L^2[-1, 1]$). Estos pueden ser definidos por la fórmula de Rodrigues, esto es

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} \left[(t^2 - 1)^n\right], \ \ n \geq 0, t \in [-1, 1]. \tag{2.2.6}$$

De nuevo, como el subespacio generado por N de estos polinomios conforma un espacio normado de dimensión finita, entonces éste es un RKHS, y sobre él podemos aplicar un argumento análogo al usado para polinomios trigonométricos.
Estos polinomios forman una base ortogonal en $L^2[-1, 1]$, con norma $||P_n||_{L^2[-1,1]} = (n + \frac{1}{2})^{-1/2}$.
Tomemos una familia finita de estos polinomios, en orden creciente de grado hasta $N \in \mathbb{N}$, de la forma $\{P_0, \ldots, P_N\}$. Enunciamos la fórmula de Christoffel-Darboux.

Proposición 2.2.1. *Sean $N \in \mathbb{N}$ y $\{Q_n\}_{n=0}^{N}$ una familia finita de polinomios ortonormales en un intervalo $I \subset \mathbb{R}$. Entonces, dados $x, y \in I$, tenemos que*

$$\sum_{n=0}^{N} Q_n(x)Q_n(y) = \begin{cases} \frac{k_n}{k_{n+1}} \frac{Q_{n+1}(x)Q_n(y) - Q_n(x)Q_{n+1}(y)}{x-y} & si \quad x \neq y, \\ \\ \frac{k_n}{k_{n+1}}\{Q'_{n+1}(x)Q_n(x) - Q'_n(x)Q_{n+1}(x)\} & si \quad x = y, \end{cases} \tag{2.2.7}$$

donde k_n denota al coeficiente de mayor orden en el polinomio Q_n, y Q'_n denota a la primera derivada del polinomio Q_n.

Aplicamos esta proposición 2.2.1 sobre nuestra familia finita de polinomios de Legendre, teniendo en cuenta que no están normalizados (basta incluir el inverso de la norma de cada P_n en la fórmula). Así, para $t \neq s$, tenemos que

$$\sum_{n=0}^{N} (n + \frac{1}{2})P_n(t)P_n(s) = \frac{2N+1}{2} \frac{P_{N+1}(t)P_N(s) - P_N(t)P_{N+1}(s)}{t-s}. \tag{2.2.8}$$

Y, análogamente, si $t = s$:

$$\sum_{n=0}^{N}(n+\frac{1}{2})P_n(t)P_n(s) = \frac{2N+1}{2}\left[P'_{N+1}(t)P_N(t) - P'_N(t)P_{N+1}(t)\right]. \tag{2.2.9}$$

Estas dos sumas, que cubren todos los casos de $t, s \in I$, conforman un núcleo reproductor, que denotamos $k_N(t,s)$, del subespacio cerrado generado por esta familia finita de polinomios de Legendre hasta grado N, en virtud de la expresión (1.3.6) del teorema 1.3.4.

2.2.3. Espacio de Paley-Wiener

Antes de empezar el estudio de este espacio, note el lector que, en el apéndice B, generalizamos la teoría de la Transformada de Fourier, dando los resultados más importantes que, en adelante, usaremos sin aclaración, así como la notación utilizada.
Empezamos definiendo lo que son las funciones de banda limitada o *band-limited functions* en un intervalo $[-R, R] \subset \mathbb{R}$, para algún $R > 0$ arbitrario. Estos elementos definen un subconjunto de $L^2(\mathbb{R})$. Recordemos que un RKHS no puede estar formado por clases de equivalencia, pero esta consideración de tomar elementos de $L^2(\mathbb{R})$ adquiere sentido con la expresión (2.2.19), que demostraremos. Los elementos del RKHS que estamos a punto de estudiar son representantes de clases en $L^2(\mathbb{R})$ que satisfacen determinadas propiedades, como ocurrirá también con el espacio de Bergman, cambiando $\mathbb{R}$ por un abierto en el plano complejo, y pidiendo, además, holomorfía. En el espacio que ahora nos ocupa, por ejemplo, esta propiedad sería la de que su transformada de Fourier tenga soporte contenido en $[-\pi, \pi]$.
Utilizaremos la caracterización de soporte para funciones localmente integrables, por la cual el soporte de una función $f \in L^2_{\text{loc}}(\mathbb{R})$ viene dado como el complementario del conjunto:

$$\{x \in \mathbb{R} : \text{existe } r > 0 \text{ tal que } f = 0 \text{ c.s. en } B(x,r)\}. \tag{2.2.10}$$

Definición 2.2.2. Sean $R > 0$ y $f \in L^2(\mathbb{R})$. Diremos que f es una *función de banda limitada* en un intervalo $[-R, R]$ si su transformada de Fourier $\hat{f}$ se anula fuera de $[-R, R]$, i.e., $\text{sop}(\hat{f}) \subset [-R, R]$.

Es habitual tomar la R de la definición anterior como la constante π. Nosotros, en adelante, tomaremos esa convención, salvo cuando se diga lo contrario.
En la literatura matemática convencional, el espacio de funciones de banda limitada en $[-\pi, \pi]$ es conocido como el **espacio de Paley-Wiener**, y se denota por PW_π. Esto es, definimos el espacio de Paley-Wiener como sigue:

$$PW\pi := \{f \in L^2(\mathbb{R}) : \text{sop}(\hat{f}) \subset [-\pi, \pi]\}. \tag{2.2.11}$$

Primero daremos unos resultados básicos para entender la naturaleza de este espacio, para después relacionarlo con nuestro estudio de RKHS (como veremos, PW_π es un RKHS). A continuación, generalizaremos esta teoría al espacio de funciones enteras en $\mathbb{C}^n$.

Proposición 2.2.3. *El espacio de Paley-Wiener PW_π es un subespacio cerrado de $L^2(\mathbb{R})$.*

Demostración. Denotamos por $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ a la aplicación transformada de Fourier. En primer lugar, es claro que PW_π forma un subepacio en $L^2(\mathbb{R})$, pues si $f, g \in PW_\pi$, y $a, b \in \mathbb{C}$, entonces $af + bg \in L^2(\mathbb{R})$. Además, por la linealidad de la aplicación transformada de Fourier, es claro que $\text{sop}(\mathcal{F}(af + bg)) \subset [-\pi, \pi]$.
Consideramos el subespacio de $L^2(\mathbb{R})$ dado por

$$S := \{f \in L^2(\mathbb{R}) : \text{sop}(f) \subset [-\pi, \pi]\}. \tag{2.2.12}$$

En virtud del teorema de Plancherel B.0.13, ésta nos da una isometría de $L^2(\mathbb{R})$ en $L^2(\mathbb{R})$, luego en particular una aplicación continua entre estos espacios de Hilbert. Si probamos que S es cerrado

en $L^2(\mathbb{R})$, entonces, como $PW_\pi = \mathcal{F}^{-1}(S)$, ya tendríamos probado que PW_π es cerrado en $L^2(\mathbb{R})$, por ser contraimagen de un cerrado por la aplicación continua $\mathcal{F}$. Veamos pues que S es cerrado en $L^2(\mathbb{R})$.
Sea $\{f_n\}_{n=1}^{\infty} \subset S$ tal que $||f_n - f||_{L^2(\mathbb{R})} \to 0$ para $n \to \infty$, para algún $f \in L^2(\mathbb{R})$. Es decir:

$$\int_{\mathbb{R}} |f_n(x) - f(x)|^2 dx \to 0, \quad n \to \infty. \tag{2.2.13}$$

Basta pues probar que $\mathrm{sop}(f) \subset [-\pi, \pi]$. Razonemos por reducción al absurdo, para lo que usaremos la caracterización de soporte dada en (2.2.10). Supongamos que $\mathrm{sop}(f)$ no está contenido en $[-\pi, \pi]$. Tomamos pues, sin pérdida de generalidad, un $x > \pi$ tal que $x \in \mathrm{sop}(f)$. Tomamos un $r > 0$ tal que $x - r > \pi$, y para el que $f \neq 0$ c.s. en $B(x, r) = (x - r, x + r)$. Entonces, haciendo uso de que $\mathrm{sop}(f_n) \subset [-\pi, \pi]$ para todo $n \in \mathbb{N}$, tenemos que

$$\int_{\mathbb{R}} |f_n(z) - f(z)|^2 dz \geq \int_{x-r}^{x+r} |f_n(z) - f(z)|^2 dz = \int_{x-r}^{x+r} |f(z)|^2 dz \neq 0, \tag{2.2.14}$$

lo que invalida la convergencia dada en (2.2.13). □

Es bien sabido de la forma en que hemos definido la transformada de Fourier que $\left\{\frac{1}{\sqrt{2\pi}} e^{-inw}\right\}_{n=1}^{\infty}$ es una base ortonormal del espacio $L^2[-\pi, \pi]$. Para probar tal afirmación, formalizamos el enunciado del correspondiente lema, y lo demostramos.

Lema 2.2.4. *La familia* $\{\frac{1}{\sqrt{2\pi}} e_n(x)\}_{n\in\mathbb{Z}} = \{\frac{1}{\sqrt{2\pi}} e^{-inx}\}_{n\in\mathbb{Z}}$ *de funciones de cuadrado integrable en* $L^2[-\pi, \pi]$ *forma una base ortonormal en dicho espacio.*

Demostración. El hecho de que esta familia de funciones es ortonormal es trivial por su forma y por el producto interno definido en $L^2_{\mathbb{C}}[-\pi, \pi]$.
Consideramos el conjunto:

$$H := \{\sum_{k=-n}^{n} a_k e_k : a_k \in \mathbb{C}, n \geq 0\}. \tag{2.2.15}$$

Es claro que H es una subálgebra del conjunto de funciones continuas de $[-\pi, \pi]$ en $\mathbb{C}$, que denotamos como $\mathcal{C}_{\mathbb{C}}[-\pi, \pi]$. Notemos que, por la definición dada, $f(-\pi) = f(\pi)$ para todo $f \in H$. Veamos que la adherencia de H es

$$G := \{f \in \mathcal{C}_{\mathbb{C}}[-\pi, \pi] : f(-\pi) = f(\pi)\}. \tag{2.2.16}$$

Por el comentario que acabamos de hacer, $H \subset G$. Sea B la bola abierta unidad en $\mathbb{C}$, y denotamos por $\mathbb{T}$ a su frontera, i.e., a la circunferencia $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. Dada $f \in H$, definimos una función $F : \mathbb{T} \to \mathbb{C}$ como $F(e^{it}) = f(t)$. F es continua, pues la aplicación $\psi : [-\pi, \pi) \to \mathbb{T}$ definida como $\psi(t) = e^{it}$ es continua, teniendo en cuenta que $\lim_{t\to-\pi^+} \psi(t) = \lim_{t\to\pi^-} \psi(t)$. Así, como $F \circ \psi = f$, tenemos la continuidad de F. La razón de esta continuidad se sigue directamente para entornos en $\mathbb{C}$ cuya contraimagen en $\mathbb{T}$, vía F, no incluye el -1. El caso problemático se solventa con la relación dada de límites laterales de ψ en π y $-\pi$, y a través de la continuidad conocida tanto de f como de ψ con esa prolongación.
Pasamos a enunciar una afirmación que, como veremos, es consecuencia directa del teorema de Stone-Weierstrass.

Afirmación 2.2.5. Si $g : \mathbb{T} \to \mathbb{C}$ es una función continua, entonces existe una sucesión $\{p_n(z, \overline{z})\}_{n=1}^{\infty}$ de polinomios en z y $\overline{z}$ tal que $p_n(z, \overline{z}) \to g(z)$ uniformemente en $\mathbb{T}$.

Demostración. Denotamos por $\mathcal{C}(\mathbb{T},\mathbb{C})$ al espacio de funciones continuas de $\mathbb{T}$ en $\mathbb{C}$, así como $P := \{p_n(z,\overline{z})\}_{n=1}^{\infty}$. El conjunto $\mathbb{T}$ es compacto en $\mathbb{C}$ (lo que topológicamente es equivalente a decir que, a través del pertinente homeomorfismo, es compacto en $\mathbb{R}^2$). Así, P es una familia de elementos de $\mathcal{C}(\mathbb{T},\mathbb{C})$ que satisface las condiciones del teorema de Stone-Weierstrass, sobre las tres operaciones naturales en esta subálgebra (producto por escalares en $\mathbb{C}$, y suma y producto de aplicaciones de $\mathcal{C}(\mathbb{T},\mathbb{C})$):

1. Claramente, P contiene a todas las aplicaciones de $\mathbb{T}$ en $\mathbb{C}$ constantes.
2. Dados $x, y \in \mathbb{T}$, $x \neq y$, basta tomar el polinomio identidad $p(z,\overline{z}) = z$ para tener un elemento de P que sea separante de estos dos elementos x, y.
3. Si $p \in P$, la aplicación conjugada de p es la misma p con sus argumentos z y $\overline{z}$ intercambiados, lo que sigue estando dentro de P.

Por lo tanto, en virtud del mencionado teorema de Stone-Weierstrass, la subálgebra P es densa en $\mathcal{C}(\mathbb{T},\mathbb{C})$ bajo la topología inducida por la norma $||\cdot||_\infty$, es decir, bajo convergencia uniforme. Dicho de otra forma, dada $g : \mathbb{T} \to \mathbb{C}$ continua, existe una sucesión de polinomios $\{p_n(z,\overline{z})\}_{n=1}^{\infty}$ en z y $\overline{z}$ tal que $p_n(z,\overline{z}) \to g(z)$ uniformemente en $\mathbb{T}$. □

Haciendo uso de la afirmación anterior, existe una sucesión de polinomios $\{p_n(z,\overline{z})\}_{n=1}^{\infty}$, tal que $p_n(z,\overline{z}) \to F(z)$ uniformemente en $\mathbb{T}$. Luego dada la aplicación continua ψ anterior, y la definición de la aplicación F, tenemos que $p_n(e^{it}, e^{-it}) \to f(t)$ uniformemente en $[-\pi,\pi]$. Además, $p_n(e^{it}, e^{-it})$ está en la subálgebra H.
Por otra parte, sabemos que el subespacio G es denso en $L^2_{\mathbb{C}}[-\pi,\pi]$, tal y como se demostró en la asignatura de Introducción a los Espacios de Funciones.
Por lo tanto, el subespacio generado por la familia de funciones $\{\frac{1}{\sqrt{2\pi}}e_n\}_{n\in\mathbb{Z}}$ es denso en $L^2_{\mathbb{C}}[-\pi,\pi]$, esto es[2]:

$$\vee\{\frac{1}{\sqrt{2\pi}}e_n : n \in \mathbb{Z}\} = L^2_{\mathbb{C}}[-\pi,\pi]. \tag{2.2.17}$$

Por lo tanto, $\{\frac{1}{\sqrt{2\pi}}e_n : n \in \mathbb{Z}\}$ es una base de $L^2_{\mathbb{C}}[-\pi,\pi]$. □

Además, como la transformada de Fourier $\mathcal{F}$ es un homeomorfismo de $L^2(\mathbb{R})$ en $L^2(\mathbb{R})$, entonces también lo es $\mathcal{F}^{-1}$. Por ello, las contraimágenes de los elementos $\frac{1}{\sqrt{2\pi}}e^{inx}$, para todo $n \in \mathbb{Z}$, constituyen una nueva base ortonormal en PW_π (pues, recordemos, $PW_\pi = \mathcal{F}^{-1}(L^2[-\pi,\pi])$, con la identificación dada anteriormente).
Dado un elemento $\frac{1}{\sqrt{2\pi}}e_n(t) = \frac{1}{\sqrt{2\pi}}e^{-int}$ arbitrario de dicha base en $L^2[-\pi,\pi]$, su transformada de Fourier inversa es como sigue

$$\mathcal{F}^{-1}(\frac{1}{\sqrt{2\pi}}e_n)(x) = \frac{1}{(\sqrt{2\pi})^2}\int_{-\pi}^{\pi} e^{it(x-n)}dt = \frac{1}{2\pi}\cdot\left[\frac{e^{i\pi(x-n)} - e^{-i\pi(x-n)}}{i(x-n)}\right] = \frac{\sin(\pi(x-n))}{\pi(x-n)}. \tag{2.2.18}$$

Esto suscita el siguiente resultado.

Proposición 2.2.6. *La familia* $\left\{\frac{\sin(\pi(x-n))}{\pi(x-n)}\right\}_{n\in\mathbb{Z}}$ *conforma una base ortonormal del espacio de Paley-Wiener* PW_π.

En este caso de estudio del espacio PW_π sobre $\mathbb{R}$ es sencillo constatar su condición de RKHS, así como dar su núcleo. Dado $f \in PW_\pi$, haciendo uso de la transformada de Fourier inversa en $[-\pi,\pi]$, tenemos que

$$f(t) = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}\hat{f}(w)e^{iwt}dw = \langle\hat{f}, \frac{e^{-iwt}}{\sqrt{2\pi}}\rangle_{L^2[-\pi,\pi]}. \tag{2.2.19}$$

[2] $\vee\{\frac{1}{\sqrt{2\pi}}e_n : n \in \mathbb{Z}\}$ denota al subespacio cerrado más pequeño que contiene a la familia $\{\frac{1}{\sqrt{2\pi}}e_n : n \in \mathbb{Z}\}$.

Esta expresión (2.2.19) era de la que hablábamos en la introducción para justificar la relación entre el espacio de Paley-Wiener y las clases de equivalencia de $L^2(\mathbb{R})$. Cuando interpretamos este espacio como RKHS, precisamente nos estamos quedando con los representantes que satisfacen esta expresión.
Por la condición de isometría biyectiva del operador transformada de Fourier, aplicado sobre elementos de PW_π, es conocido que $||f||_{L^2(\mathbb{R})} = ||\hat{f}||_{L^2(\mathbb{R})} = ||\hat{f}||_{L^2[-\pi,\pi]}$ (teniendo en cuenta que, por tener a la función f en PW_π, tenemos $\text{sop}(\hat{f}) \subset [-\pi, \pi]$). Así, para todo $t \in \mathbb{R}$, tenemos, por la desigualdad de Cauchy-Schwarz en $L^2[-\pi, \pi]$, que

$$|f(t)| \leq ||\hat{f}||_{L^2[-\pi,\pi]} ||\frac{e^{-iwt}}{\sqrt{2\pi}}||_{L^2[-\pi,\pi]} = ||f||_{L^2(\mathbb{R})}, \quad f \in PW_\pi. \tag{2.2.20}$$

En otras palabras, acabamos de probar que los funcionales evaluación son acotados en PW_π, i.e., son continuos en PW_π. Por lo tanto, PW_π es un RKHS.
Para deducir su núcleo reproductor, haremos un cálculo directo evaluando un elemento arbitrario en PW_π. Sea $f \in PW_\pi$. Entonces, se tiene que, a través de la transformada inversa de Fourier

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \hat{f}(w) e^{iwt} dw = \langle \hat{f}, \frac{e^{-iwt}}{\sqrt{2\pi}} \rangle_{L^2[-\pi,\pi]} = \langle f, \mathcal{F}^{-1}(\frac{1}{\sqrt{2\pi}} e^{-iwt}) \rangle_{L^2(\mathbb{R})}, \tag{2.2.21}$$

donde se ha usado que $\mathcal{F}^{-1}$ es una isometría biyectiva de $L^2(\mathbb{R})$ en $L^2(\mathbb{R})$. Esto demuestra que la contraimagen de $\mathcal{F}^{-1}(\frac{1}{\sqrt{2\pi}} e^{-iwt})$ nos da un núcleo reproductor de PW_π, por satisfacer, en todo $t \in \mathbb{R}$, y para toda $f \in PW_\pi$, la propiedad reproductora.
Haciendo un cálculo muy parecido al hecho en (2.2.18), tenemos que, fijando $t \in \mathbb{R}$, y tomando $s \in \mathbb{R}$, $s \neq t$,

$$\mathcal{F}^{-1}(\frac{1}{\sqrt{2\pi}} e^{-iws})(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{iw(t-s)} dw = \frac{1}{2\pi} \left[\frac{e^{iw(t-s)}}{i(t-s)} \right]_{-\pi}^{\pi} \tag{2.2.22}$$

$$= \frac{1}{2\pi} \left[\frac{e^{i\pi(t-s)} - e^{-i\pi(t-s)}}{i(t-s)} \right] = \frac{\sin(\pi(t-s))}{\pi(t-s)}. \tag{2.2.23}$$

Extendemos la función con continuidad por el 1 si $t = s$. Expresamos dicho núcleo reproductor como

$$k_\pi(t, s) = \text{sinc}(\pi(t-s)), \quad t, s \in \mathbb{R}, \tag{2.2.24}$$

donde, para todo $x \in \mathbb{R}$, definimos

$$\text{sinc}(x) = \begin{cases} \frac{\sin(x)}{x} & si \quad x \neq 0 \\ 1 & si \quad x = 0. \end{cases} \tag{2.2.25}$$

De esta forma, reescribiendo dicha propiedad reproductora, ahora con el núcleo reproductor conocido explícitamente, se tiene que

$$f(s) = \langle \hat{f}, \frac{e^{-iws}}{\sqrt{2\pi}} \rangle_{L^2[-\pi,\pi]} = \langle f, \text{sinc}(\pi(\cdot - s)) \rangle_{L^2(\mathbb{R})}, \quad s \in \mathbb{R}. \tag{2.2.26}$$

Así, observando por (2.2.24) que $k_\pi(t, t) = 1$ para cualquier $t \in \mathbb{R}$, estamos en condiciones de enunciar el teorema de muestreo de Shannon, el cual es un caso particular del teorema del muestreo dado anteriormente. La demostración es la misma que la del teorema general del muestreo 2.1.1, con una diferencia: partimos de una serie convergente en $L^2[-\pi, \pi]$, y utilizamos $\mathcal{F}^{-1}$ para pasar de este espacio a PW_π. El resto del razonamiento es análogo.

Teorema 2.2.7 (Teorema de muestreo de Shannon). *Sea $f \in PW_\pi$, i.e., una función de banda limitada en $[-\pi, \pi]$. Entonces, podemos representar f en términos de sus muestras $\{f(n)\}_{n\in\mathbb{Z}}$ a través de la igualdad:*

$$f(t) = \sum_{n=-\infty}^{\infty} f(n)\frac{\sin(\pi(t-n))}{\pi(t-n)}, \quad t \in \mathbb{R}, \tag{2.2.27}$$

donde la convergencia de esta serie es uniforme en $\mathbb{R}$.

Demostración. Sea $f \in PW_\pi$. La expansión de su transformada de Fourier $\hat{f} \in L^2[-\pi, \pi]$ en la base ortonormal $\{e^{-inw}/\sqrt{2\pi}\}_{n=-\infty}^{\infty}$ de $L^2[-\pi, \pi]$ es:

$$\hat{f} = \sum_{n=-\infty}^{\infty} \langle \hat{f}, \frac{e^{-inw}}{\sqrt{2\pi}}\rangle \frac{e^{-inw}}{\sqrt{2\pi}} = \sum_{-\infty}^{\infty} f(n)\frac{e^{-inw}}{\sqrt{2\pi}}, \quad \text{en } L^2[-\pi, \pi]. \tag{2.2.28}$$

Aplicando la transformada inversa de Fourier $\mathcal{F}^{-1}$ sobre (2.2.28), y utilizando su linealidad:

$$f = \sum_{n=-\infty}^{\infty} f(n)\mathcal{F}^{-1}\left(\frac{e^{-inw}}{\sqrt{2\pi}}\chi_{[-\pi,\pi]}(w)\right) = \sum_{n=-\infty}^{\infty} f(n)\frac{\sin(\pi(\cdot - n))}{\pi(\cdot - n)}, \text{ en } L^2(\mathbb{R}). \tag{2.2.29}$$

En la última igualdad hemos utilizado la expresión de las transformadas inversas de $e^{-inw}/\sqrt{2\pi}$ deducidas en (2.2.23).
Para deducir de ello la convergencia uniforme, recordemos que $\mathcal{F}^{-1}$ es continua, luego la convergencia de (2.2.28) en $L^2[-\pi, \pi]$ implica la convergencia de (2.2.29) en $PW_\pi = \mathcal{F}^{-1}(L^2[-\pi, \pi])$, con la norma de $L^2(\mathbb{R})$ restringida a PW_π. Este último espacio es el RKHS que estamos tratando, es decir, tenemos convergencia de la suma (2.2.29) en un RKHS. Como este núcleo, $k_\pi(\cdot, \cdot)$, es acotado, por el teorema 1.4.2 tenemos que la inclusión id : $PW_\pi \to L^\infty(\mathbb{R})$ es continua. Luego la convergencia en PW_π dada en (2.2.29) implica la convergencia uniforme. □

Dado que entre un término y otro de la serie (2.2.27) hay una diferencia de una unidad, el espacio de Paley-Wiener tiene lo que se llama un *período de muestreo* $T_s = 1$, y no es para nada relevante en qué puntos exactamente son tomadas estas muestras que nos permitan expandir $f \in PW_\pi$ a través de (2.2.27).
Notemos que, por la expresión de (2.2.27), si dicha expansión se da sobre una base que, además de ortogonal, es ortonormal, entonces, recordando que $f(n) = \langle \hat{f}, e^{-inw}/\sqrt{2\pi}\rangle$ para todo $n \in \mathbb{N}$, la identidad de Parseval implica:

$$||f||^2 = \sum_{n=-\infty}^{\infty} |f(n)|^2, \quad f \in PW_\pi. \tag{2.2.30}$$

A este valor $E_f := ||f||^2$ se le llama, como es común ver en la literatura de la teoría del muestreo, *enegía de f*. También en términos de esta teoría, se suele decir que $f \in PW_\pi$ *está contenida en sus muestras* $\{f(n)\}_{n\in\mathbb{Z}}$.
Este último comentario da significado al siguiente diagrama conmutativo:

$$\begin{tikzcd}
f \in PW_\pi \arrow[r, "\mathcal{F}"] \arrow[d, "\mathcal{S}"] & \hat{f} \in L^2[-\pi, \pi] \arrow[d, "\mathcal{P}"] \\
\{f(n)\}_{n\in\mathbb{Z}} \in l^2(\mathbb{Z}) \arrow[r, "\mathcal{F}"] & \tilde{f} \in \tilde{L}_2[-\pi, \pi]
\end{tikzcd}$$

donde los operadores involucrados son los siguientes:

(i) $\mathcal{S}$ es la *aplicación de muestreo* con período $T_s = 1$.

(ii) $\mathcal{P}$ es una *aplicación de* 2π*-periodización* que extiende una función $\hat{f}$ de tener un dominio en $[-\pi, \pi]$, a tenerlo en todo $\mathbb{R}$ con período 2π.

(c) Las otras dos aplicaciones son transformadas de Fourier:

a) La aplicación $\mathcal{F}$ superior es una transformada de Fourier funcional en $L^2(\mathbb{R})$.

b) La aplicación $\mathcal{F}$ inferior es la transformada de Fourier en $l^2(\mathbb{Z})$, definida como:

$$\mathcal{F}(\{a_n\})(w) := \sum_{n\in\mathbb{Z}} a_n \frac{e^{-inw}}{\sqrt{2\pi}}. \tag{2.2.31}$$

No olvidemos que PW_π es un subespacio de $L^2(\mathbb{R})$, que de hecho es cerrado. Esto nos permite hablar de la aplicación proyección en PW_π en términos de una aplicación bien definida, continua, y lipschitziana (con constante de Lipschitz $K = 1$). En nuestro subespacio PW_π, esta proyección adquiere una forma muy bien conocida.

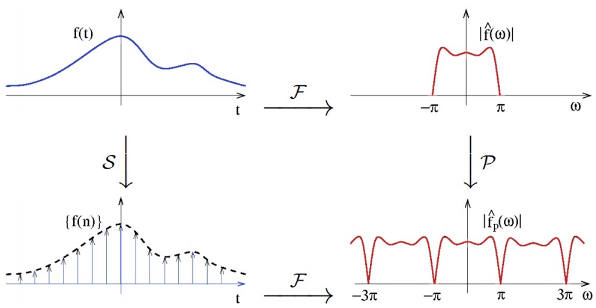

Figura 2.1: Interpretación gráfica del diagrama conmutativo entre los espacios PW_π, $L^2[-\pi,\pi]$, $\tilde{L}_2[-\pi,\pi]$, y $l^2(\mathbb{Z})$. Esta construcción nos permite descomponer las funciones de PW_π en los valores que tome en un conjunto, a lo sumo, infinito numerable de X, o bien en la forma que tome su transformada de Fourier en su soporte acotado (de hecho, compacto).

Proposición 2.2.8. *Sea $f \in PW_\pi$, y sea P_{PW_π} el operador proyección sobre PW_π dentro del espacio $L^2(\mathbb{R})$. Entonces, para todo $x \in \mathbb{R}$:*

$$P_{PW_\pi} f(x) = \langle f, \frac{\sin(\pi(. - x))}{\pi(. - s)} \rangle_{L^2(\mathbb{R})} = (f * \text{sinc})\,(s), \quad s \in \mathbb{R}, \tag{2.2.32}$$

donde $*$ denota al operador convolución.

Demostración. Se sigue inmediatamente de la definición de convolución en $\mathbb{R}$. □

Este último resultado deja bien claro el significado de convolucionar una función f de cuadrado integrable arbitraria con la función sinc: no es más que proyectar dicha f sobre un espacio en que el soporte de la transformada de Fourier, $\hat{f}$, está contenido en el intervalo $[-\pi, \pi]$.
Así, algunas de las consecuencias interesantes que este *muestreo* tiene en el espacio PW_π, son:

1. Por lo que hemos visto, en nuestro espacio PW_π tenemos un *período de muestreo* de $T_s = 1$. De hecho, este es el rasgo relevante en términos de localización, no así el conjunto infinito numerable de puntos en que localizamos las muestras con que recomponer los elementos $f \in PW_\pi$. Esto es así porque, dado $f \in PW_\pi$, esta puede ser reconstruido con la sucesión $\{f(n+a)\}_{n=-\infty}^{\infty}$,

para cualquier $a \in \mathbb{R}$ fijado. Este hecho procede de que el conjunto $\left\{e^{-i(n+a)w/\sqrt{2\pi}}\right\}_{n\in\mathbb{Z}}$ es también una base ortonormal de $L^2[-\pi,\pi]$, con una demostración análoga a la ya dada para $\left\{e^{-inw/\sqrt{2\pi}}\right\}_{n\in\mathbb{Z}}$. Repetiendo el mismo razonamiento, tenemos, vía $\mathcal{F}^{-1}$, que el conjunto

$$\left\{\frac{\sin(\pi(t-n-a))}{\pi(t-n-a)}\right\}_{n\in\mathbb{Z}}, \tag{2.2.33}$$

es también una base ortonormal de PW_π. Esto nos da una nueva fórmula de muestreo con la siguiente expresión:

$$f(t) = \sum_{n=-\infty}^{\infty} f(n+a)\frac{\sin(\pi(t-n-a))}{\pi(t-n-a)}, \quad t \in \mathbb{R}. \tag{2.2.34}$$

2. Por un sencillo cálculo sobre la fórmula de Shannon (2.2.27) (basta utilizar la identidad trigonométrica $\sin(a+b) = \sin(a)\cos(b) + \sin(b)\cos(a)$, $\sin(n\pi) = 0$, y $\cos(n\pi) = (-1)^n$), para todo $f \in PW_\pi$ se obtiene:

$$f(t) = \sum_{n\in\mathbb{Z}} f(n)\frac{(-1)^n \sin(\pi t)}{\pi(t-n)} = \sum_{n\in\mathbb{Z}} f(n)\frac{P(n)}{P'(n)(t-n)}, \tag{2.2.35}$$

lo cual adopta la forma de una serie de interpolación de tipo Lagrange, para $P(t) = \sin(\pi t)$.

3. La longitud que tome el intervalo en que contenemos el soporte de las transformadas de Fourier de los elementos de PW_π es fácilmente maleable. Para ver esto, consideremos una constante $a > 0$ totalmente arbitraria. Consideramos el espacio de Paley-Wiener $PW_{a\pi}$ de funciones de banda limitada en el intervalo $[-a\pi, a\pi]$, definido como:

$$PW_{a\pi} := \left\{f \in L^2(\mathbb{R}) : \mathrm{sop}(f) \subset [-a\pi, a\pi]\right\}. \tag{2.2.36}$$

En este caso, el período de muestreo es $T_s = 1/a$. Si tomamos $h \in PW_{a\pi}$, y definimos $f(t) := h(t/a)$, para todo $t \in \mathbb{R}$, por un sencillo cambio de variable tenemos que $\hat{f}(w) = a\hat{h}(aw)$, para todo $w \in [-\pi,\pi]$, por lo que $f \in PW_\pi$. Además, aplicando la fórmula de muestreo de Shannon (2.2.27) sobre $f \in PW_\pi$:

$$f(t) = h(t/a) = \sum_{n\in\mathbb{Z}} h(n/a)\frac{\sin(\pi(t-n))}{\pi(t-n)}, \quad t \in \mathbb{R}. \tag{2.2.37}$$

Aplicando ahora el cambio de variable $s = t/a$, para $h \in PW_{a\pi}$ arbitrario, de lo anterior obtenemos:

$$h(s) = \sum_{n\in\mathbb{Z}} h(n/a)\frac{\sin(\pi(as-n))}{\pi(as-n)}, \quad s \in \mathbb{R}, \tag{2.2.38}$$

de lo que se deduce un núcleo reproductor para el espacio de Paley-Wiener $PW_{a\pi}$ con período de muestreo $T_s = 1/a$ con la siguiente expresión, para $t, s \in \mathbb{R}$,

$$k_{a\pi}(t,s) = a\mathrm{sinc}(a(t-s)), \quad , \ t \neq s,$$

$$k_{a\pi}(t,s) = a, \ t = s, \tag{2.2.39}$$

donde se ha incluido el nuevo período de muestreo $T_s = 1/a$ sobre los nuevos elementos $e^{-inwa}/\sqrt{2\pi}$ de la base ortonormal de $L^2[-\pi,\pi]$, cuya antitransformada de Fourier es, precisamente, el $k_{a\pi}$ anterior.

Nótese que el efecto que la dilatación tiene sobre las expansiones de los elementos de $PW_{a\pi}$ es la de dilatar, a través de $a > 0$, el propio término $\mathrm{sinc}(a(t-s))$, así como contraer a través de $1/a$ el soporte de la transformada, en coherencia con el nuevo período de muestreo T_s dado.

4. Usualmente, y de hecho es así como se ha presentado, los dominios de las transformadas de Fourier de los elementos $f \in PW_\pi$ están centrados en 0. Esto ocurre siempre que estemos tratando con funciones de banda limitada con valores reales.
   Para ver esto, sea $f \in PW_\pi$, de tal forma que, por tomar valores reales, $|\hat{f}(w)|^2 = \hat{f}(w)\overline{\hat{f}}(w) = \hat{f}(w)\hat{f}(-w)$, para todo $w \in [-\pi, \pi]$. En otras palabras, $\hat{f}$ es una función par con respecto al 0. De momento, en ningún momento se ha utilizado la naturaleza del soporte de $\hat{f}$.
   Supongamos que f, además de estar en $L^2(\mathbb{R})$, tiene banda limitada en $[w_0 - \pi, w_0 + \pi]$, para algún $w_0 \in \mathbb{R}$ arbitrario. Tomamos $g \in L^2(\mathbb{R})$ de la forma $g(t) := e^{-iw_0t}f(t)$, con lo que $\hat{g}(w) = \hat{f}(w + w_0)$, y entonces g es de banda limitada en $[-\pi, \pi]$. Aplicando entonces el teorema de muestreo de Shannon (2.2.27) sobre g, obtenemos:
   $$g(t) = e^{-iw_0t}f(t) = \sum_{n\in\mathbb{Z}} e^{-iw_0n}f(n)\frac{\sin(\pi(t-n))}{\pi(t-n)}, \quad t \in \mathbb{R}, \tag{2.2.40}$$
   con lo que, despejando, la fórmula de muestreo para el f dado es como sigue
   $$f(t) = \sum_{n\in\mathbb{Z}} e^{iw_0(t-n)}f(n)\frac{\sin(\pi(t-n))}{\pi(t-n)}, \quad t \in \mathbb{R}. \tag{2.2.41}$$
   La lectura de este fenómeno es la siguiente: dado un $f \in L_w(\mathbb{R})$ con banda limitada contenida en un intervalo compacto de longitud 2π, independientemente de la localización de ese intervalo, siempre podremos expresar f a través de la fórmula de muestreo de Shannon en términos de funciones sinc evaluadas en $\{\frac{\pi(t-n)}{\pi(t-n)}\}_{n\in\mathbb{Z}}$, añadiendo un factor $e^{iw_0(t-n)}$ a cada uno de esos términos que dé cuenta de esa traslación. Si entendemos f como una señal, y el dominio de su transformada $\hat{f}$ como un dominio de frecuencias de dicha señal, la consecuencia es que, en términos de la intensidad de la señal (calculado a partir de módulos al cuadrado, para los que $|e^{iw_0(t-n)}|^2 = 1$), dicha señal, con banda localizada en $w_0 \in \mathbb{R}$, será equivalente a cualquier otra señal localizada en otro $w_1 \in \mathbb{R}$.

Espacio de Paley-Wiener como RKHS de funciones enteras

Extendemos brevemente algunas de las nociones del análisis del espacio de Paley-Wiener, hecho hasta ahora para funciones en $\mathbb{R}$, a funciones enteras.
Esta extensión es inmediata si tomamos una f en nuestro PW_π original y escribimos, para todo $z \in \mathbb{C}$:
$$f(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{f}(w)e^{izw}dw. \tag{2.2.42}$$
Su buena definición se sigue de la integrabilidad de la función $\hat{f}$ en cualquier intervalo contenido en su soporte $[-\pi, \pi]$, por ser f elemento de PW_π.
La continuidad de dicha extensión de f se sigue del teorema de continuidad de integrales paramétricas, pues $\hat{f}(w)e^{izw}$ es continua con respecto de w, es integrable con respecto de z, y puede ser mayorada, en términos de su módulo, por una función integrable con respecto de w (de hecho, por $|\hat{f}(w)|$).
Además, f es elemento del espacio de funciones enteras $\mathcal{H}(\mathbb{C})$, por la aplicación recurrente del teorema de Leibniz de derivación de integrales paramétricas. Esto es así por ser inmediato que el integrando en (2.2.42) es una función holomorfa en la variable compleja z, y por satisfacer éste las condiciones de Cauchy-Riemann (no es más que una función exponencial acompañada por un factor que no depende de z).
También, esta extensión de elementos f de PW_π es una función *con acotación exponencial de exponente, a lo sumo,* π, i.e., probaremos a continuación que dicha extensión satisface $|f(z)| \leq Ae^{\pi|z|}$, para todo $z \in \mathbb{C}$, y para alguna constante $A > 0$. Veámoslo: tomemos $z = x + iy \in \mathbb{C}$, con lo que $\mathrm{Re}(izw) = -wy$, y
$$|f(x+iy)| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |\hat{f}(w)|e^{-wy}dw \leq \frac{e^{\pi|z|}}{2\pi}\int_{-\pi}^{\pi} |\hat{f}(w)|dw, \tag{2.2.43}$$

donde se ha usado $|y| \leq |z|$, luego $e^{-\pi y} \leq e^{\pi|y|} \leq e^{\pi|z|}$. De este modo, aplicando la desigualdad de Hölder en $L^2(\mathbb{R})$, se tiene que

$$|f(x+iy)| \leq \frac{e^{\pi|z|}}{2\pi} \int_{-\pi}^{\pi} 1 \cdot |\hat{f}(w)| dw \leq \frac{e^{\pi|z|}}{2\pi} \left(\int_{-\pi}^{\pi} 1^2 dw \right)^{1/2} \left(\int_{-\pi}^{\pi} |\hat{f}(w)|^2 dw \right)^{1/2} \tag{2.2.44}$$

$$= e^{\pi|z|} \left[\frac{1}{2\pi} (2\pi)^{1/2} ||\hat{f}||_{L^2(\mathbb{R})} \right]. \tag{2.2.45}$$

El factor que acompaña a $e^{\pi|z|}$ es una constante en $\mathbb{C}$. Esto demuestra la acotación exponencial enunciada.
Es natural preguntarse si una función entera que satisfaga la acotación anterior, cuya restricción a $\mathbb{R}$ está en $L^2(\mathbb{R})$, es siempre de la forma (2.2.42). Esto nos da una caracterización completa del espacio de Paley-Wiener de funciones enteras.

Teorema 2.2.9 (**Teorema de Paley-Wiener**). *Sea f una función entera tal que $|f(z)| \leq Ce^{\pi|z|}$ para todo $z \in \mathbb{C}$, y que $f|_{\mathbb{R}} \in L^2(\mathbb{R})$. Entonces, existe una función $F \in L^2[-\pi,\pi]$ tal que*

$$f(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(w) e^{izw} dw, \quad z \in \mathbb{C}. \tag{2.2.46}$$

Demostración. Consideramos $f_\varepsilon(x) = f(x)e^{-\varepsilon|x|}$, para $\varepsilon > 0$, y $x \in \mathbb{R}$. Nótese que estamos definiendo f_ε a partir de la restricción de f a $\mathbb{R}$. Nuestro objetivo será probar que

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}} f_\varepsilon(x) e^{-itx} = 0, \quad t \in \mathbb{R}, |t| > \pi. \tag{2.2.47}$$

Supongamos que (2.2.47) se cumple. Además, por el teorema de la convergencia dominada sobre $|f_\varepsilon - f|^2$ restringido a $\mathbb{R}$, aplicable porque $|f_\varepsilon - f|^2 \leq 4|f|^2 \in L^1(\mathbb{R})$, tenemos que $||f_\varepsilon - f||_2 \to 0$ si $\varepsilon \to 0$. Entonces, por el teorema de Plancherel B.0.13, las transformadas de Fourier $\mathcal{F}(f_\varepsilon)$ convergen a $F = \mathcal{F}(f)$ en $L^2(\mathbb{R})$ (mejor dicho, a la tranformada de f restringida a $\mathbb{R}$). Esta convergencia se puede leer en los siguientes términos:

$$\int_{\mathbb{R}} f_\varepsilon(x) e^{-itx} dx \to \int_{\mathbb{R}} f(x) e^{-itx} dx = F(t), \quad \varepsilon \to 0. \tag{2.2.48}$$

De esta forma, (2.2.47) implicaría que F se anula fuera de $[-\pi,\pi]$. Por tanto, escribiendo f a través de la transformada de Fourier inversa de F, tenemos que

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} F(t) e^{itx} dt \implies f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(t) e^{itx} dt. \tag{2.2.49}$$

Si probamos que la integral paramétrica

$$z \in \mathbb{C} \mapsto g(z) := \frac{1}{2\pi} \int_{-\pi}^{\pi} F(t) e^{itz} dt, \tag{2.2.50}$$

es entera, por el principio de identidad y dado que f también es entera, entonces ya habríamos acabado, una vez probado (2.2.47). El hecho de que (2.2.50) es entera es consecuencia del teorema de diferenciabilidad en el sentido complejo de integrales paramétricas. Tal teorema se enuncia como sigue, simplificando sus hipótesis a nuestro interés.

Teorema 2.2.10 (Teorema de diferenciabilidad compleja de integrales paramétricas). *Sea $U \subset \mathbb{C}$ abierto, A un subespacio medible de $\mathbb{R}^n$, y $f : U \times A \to \mathbb{C}$. Supongamos que:*

1. *Para todo $z \in U$, $f_z : A \to \mathbb{C}$, dada por $f_z(x) = f(z,x)$ es medible Lebesgue.*

2. *Para todo $x \in A$, la función $f_x : U \to \mathbb{C}$ dada por $f_x(z) = f(z,x)$ es holomorfa en U.*

3. *Para todo $z_0 \in U$, existe un entorno V de z_0 contenido en U, y una función $h : A \to [0,\infty)$ integrable Lebesgue en A tal que*

$$|f(z,x)| \leq h(x), \text{ para todo } (z,x) \in V \times A.$$

Entonces, la función $F : U \to \mathbb{C}$ dada por $F(z) = \int_A f(z,x)dx$, para todo $z \in U$, es holomorfa en U, y además $F'(z) = \int_A \frac{\partial f}{\partial z}(z,x)dx$, para todo $z \in U$.

Así, tomando la F y la f de 2.2.10 como la g y la $\frac{1}{2\pi}F(t)e^{itx}$ de nuestro desarrollo, la holomorfía en $\mathbb{C}$ de esta última integral paramétrica se sigue de:

1. $t \in (-\pi,\pi) \mapsto \frac{1}{2\pi}F(t)e^{itz}$ continua para todo $z \in \mathbb{C}$, luego medible Lebesgue.

2. Fijado $t \in (-\pi,\pi)$, la función $z \in \mathbb{C} \mapsto F(t)e^{itz}$ es holomorfa, por serlo la función exponencial compleja.

3. Para todos $(z,t) \in \mathbb{C} \times (-\pi,\pi)$, se tiene $|F(t)e^{itz}| \leq |F(t)|e^{-tIm(z)}$, que es integrable en $(-\pi,\pi)$ por la desigualdad de Hölder, ya que

$$\int_{-\pi}^{\pi} |F(t)|e^{-tIm(z)}dt \leq \left(\int_{-\pi}^{\pi} |F(t)|^2 dt\right)^{1/2} \left(\int_{-\pi}^{\pi} e^{-2tIm(z)}dt\right)^{1/2}, \qquad (2.2.51)$$

con $F \in L^2(\mathbb{R})$, y $e^{-2tIm(z)} \in L^1(-\pi,\pi)$.

Entonces, ahora sí, g es entera. Veamos pues que (2.2.47) se satisface, y con ello habríamos acabado. Para cada real $a \in \mathbb{R}$, definimos el camino $\Gamma_a(s) = se^{ia}$, con $s \in [0,\infty)$. Sea

$$\Omega_a = \{w \in \mathbb{C} : \mathrm{Re}(we^{ia}) > \pi\}. \qquad (2.2.52)$$

Si $w \in \Omega_a$, definimos

$$\phi_a(w) = \int_{\Gamma_a} f(z)e^{-wz}dz = e^{ia}\int_0^{\infty} f(se^{ia})\exp(-wse^{ia})ds. \qquad (2.2.53)$$

El integrando se puede acotar como sigue, dado que $|f(z)| \leq C\exp(\pi|z|)$ y dada la definición de Ω_a,

$$\begin{aligned}|f(se^{ia})\exp(-wse^{ia})| &\leq C\exp(\pi s)\cdot\exp(-\mathrm{Re}(we^{ia})s)\\ &= C\exp\left[-s(\mathrm{Re}(we^{ia})-\pi)\right].\end{aligned}$$

Es claro que tanto las condiciones de medibilidad como de holomorfía del integrando se cumplen para aplicar 2.2.10. Solo faltaría encontrar, para afirmar que ϕ_a es holomorfa en Ω_a, alguna acotación del módulo del integrando por medio de una función integrable en $(0,\infty)$. Tomamos un $w_0 \in \Omega_a$. Podemos escoger un entorno V de w_0 en Ω_a tal que, para todo $w \in V$, $\mathrm{Re}(we^{ia}) > A$, para alguna constante A tal que $A > \pi$. Entonces, en V, el integrando se puede acotar como sigue

$$\begin{aligned}|f(se^{ia})\exp(-wse^{ia})| &\leq C\exp(\pi s)\cdot\exp(-\mathrm{Re}(we^{ia})s)\\ &\leq C\exp(-s(A-\pi)), \quad s \in (0,\infty),\end{aligned}$$

donde la última función es integrable en $(0,\infty)$. Así, ya hemos probado la holomorfía de ϕ_a en Ω_a, para todo $a \in \mathbb{R}$.

La gracia de esta construcción está en que, si cogemos $a = 0$ y $a = \pi$, podemos extender los semiplanos

Ω_a de holomorfía de estas dos ϕ_a, que serían $\{w \in \mathbb{C} : \mathrm{Re}(w) > \pi\}$ y $\{w \in \mathbb{C} : \mathrm{Re}(w) < -\pi\}$, a los nuevos semiplanos, respectivamente, $\{w \in \mathbb{C} : \mathrm{Re}(w) > 0\}$ y $\{w \in \mathbb{C} : \mathrm{Re}(w) < 0\}$. Entonces,

$$\phi_0(w) = \int_0^\infty f(x)e^{-wx}dx, \ \ \mathrm{Re}(w) > 0, \tag{2.2.54}$$

$$\phi_\pi(w) = -\int_{-\infty}^0 f(x)e^{-wx}dx, \ \ \mathrm{Re}(w) < 0. \tag{2.2.55}$$

La holomorfía de estas dos funciones se sigue de que estamos suponiendo que f, restringida a $\mathbb{R}$, están en $L^2(\mathbb{R})$, lo que permite acotar los integrandos por funciones integrables en $(0, \infty)$. Las condiciones de medibilidad y de holomorfía del integrando son inmediatas. Se tiene que

$$|f(s)e^{-ws}| = |f(s)| \cdot e^{-\mathrm{Re}(w)s}. \tag{2.2.56}$$

Si aplicamos la desigualdad de Hölder en la última expresión,

$$\int_{\mathbb{R}} |f(s) \cdot e^{-\mathrm{Re}(w)s}|ds \leq \left(\int_{\mathbb{R}} |f(s)|^2 ds\right)^{1/2} \cdot \left(\int_{\mathbb{R}} e^{-2\mathrm{Re}(w)s} ds\right)^{1/2}, \tag{2.2.57}$$

donde $f|_{\mathbb{R}} \in L^2(\mathbb{R})$, y si $\mathrm{Re}(w) > 0$ entonces el último factor también es finito. Esto confirma que la cota encontrada para el integrando de (2.2.54) es integrable en $(0, \infty)$. Con ello, hemos probado la holomorfía de ϕ_0 para el semiplano dada por $\mathrm{Re}(w) > 0$. Se procede análogamente con ϕ_π y el semiplano dada por $\mathrm{Re}(w) < 0$.
Ahora bien, si $t \in \mathbb{R}$,

$$\begin{aligned}\int_{-\infty}^\infty f_\varepsilon(x)e^{-itx}dx &= \int_{-\infty}^0 f(x)e^{-itx+\varepsilon x}dx + \int_0^\infty f(x)e^{-itx-\varepsilon x}dx \\ &= \int_{-\infty}^0 f(x)e^{-x(it-\varepsilon)}dx + \int_0^\infty f(x)e^{-x(it+\varepsilon)}dx \\ &= \phi_0(it+\varepsilon) - \phi_\pi(it-\varepsilon).\end{aligned}$$

Por tanto, para terminar basta con demostrar que

$$\phi_0(it+\varepsilon) - \phi_\pi(it-\varepsilon) \to 0, \ \ \varepsilon \to 0, \tag{2.2.58}$$

si $|t| > \pi$. Para ello, veamos que las funciones ϕ_a son prolongaciones analíticas las unas de las otras, con tal de que sus dominios tengan intersección no vacía.
Tomamos $c, d \in \mathbb{R}$ tales que $0 < d - c < \pi$, y definimos

$$\gamma = \frac{c+d}{2}, \ \ \eta = \cos\left(\frac{d-c}{2}\right) > 0. \tag{2.2.59}$$

Consideramos $w = |w|e^{-i\gamma}$, con lo que $\mathrm{Re}(we^{ic}) = |w|\cos(\gamma - c) = |w|\eta$. Análogamente, $\mathrm{Re}(we^{id}) = |w|\eta$.
Así, como $\mathrm{Re}(we^{id}) = |w|\eta = \mathrm{Re}(we^{ic})$, entonces $w \in \Omega_c \cap \Omega_d$ si $|w| > \pi/\eta$. Con esta condición probada, veamos que $\phi_c(w) = \phi_d(w)$ para estos $w = |w|e^{-i\gamma}$, y por el principio de identidad esto implica que estas funciones coinciden en la intersección de los semiplanos Ω_c y Ω_d.
Consideramos

$$\int_\Gamma f(z)e^{-wz}dz, \tag{2.2.60}$$

con $\Gamma(t) = re^{it}$, $t \in [a, b]$, que describe el arco que conecta las semirrectas de argumentos c y d. Como la holomorfía implica la existencia de primitiva, y la existencia de primitiva implica la anulación al integrar sobre curvas cerradas, tenemos que

$$\int_{[0,r]e^{ic}} f(z)e^{-wz}dz + \int_\Gamma f(z)e^{-wz}dz = \int_{[0,r]e^{id}} f(z)e^{-wz}dz. \tag{2.2.61}$$

Acotamos el segundo término del primer miembro. Como en Γ tenemos $z = re^{it}$, entonces

$$\mathrm{Re}(-wz) = -|w|r\cos(t-\gamma) \leq -|w|r\eta, \tag{2.2.62}$$

luego

$$|f(z)e^{-wz}| \leq Ce^{(\pi-|w|\eta)r}, \tag{2.2.63}$$

y de este modo

$$\left|\int_\Gamma f(z)e^{-wz}dz\right| \leq Ce^{(\pi-|w|\eta)r}r(d-c). \tag{2.2.64}$$

Llegados a este punto, como $|w| > \pi/\eta$ para que la intersección sea no vacía, entonces el coeficiente de r en la cota anterior es negativo. Por tanto, dicha integral converge a 0 si $r \to \infty$, es decir

$$\int_{\Omega_d} f(z)e^{-wz}dz = \int_{\Omega_c} f(z)e^{-wz}dz, \tag{2.2.65}$$

si $|w| > \pi/\eta$ y $w = |w|e^{-i\gamma}$. Así, por el principio de identidad, ϕ_c y ϕ_d coinciden en la intersección de sus semiplanos de definición.
Ya para acabar, notemos que $\pm\varepsilon + it \in \Omega_{-\pi/2}$ si $t > \pi$. De ello se sigue que ϕ_0 y ϕ_π coinciden por $\phi_{-\pi/2}$. Entonces, podemos sustituir, en (2.2.58), la función $\phi_{-\pi/2}$, de tal forma que nos quedaría

$$\int_{-\infty}^{\infty} f_\varepsilon(x)e^{-itx}dx = \phi_{-\pi/2}(\varepsilon+it) - \phi_{-\pi/2}(-\varepsilon+it), \tag{2.2.66}$$

luego, por continuidad, si hacemos $\varepsilon \to 0$, tenemos que el límite es 0.
Análogamente, para $t < -\pi$ se tiene que $\pm\varepsilon + it \in \Omega_{\pi/2}$, luego podemos sustituir en la expresión anterior ϕ_0 y ϕ_π por $\phi_{\pi/2}$, de forma que

$$\int_{-\infty}^{\infty} f_\varepsilon(x)e^{-itx}dx = \phi_{\pi/2}(\varepsilon+it) - \phi_{\pi/2}(-\varepsilon+it) \ \ \to \ \ 0, \ \ \varepsilon \to 0, \tag{2.2.67}$$

de nuevo por continuidad. Esto prueba el resultado. $\square$

Como ya hemos dicho, caracterizamos entonces PW_π para funciones enteras de la siguiente manera:

$$PW_\pi = \left\{ f \in H(\mathbb{C}) : |f(z)| \leq Ae^{\pi|z|}, f|_{\mathbb{R}} \in L^2(\mathbb{R}) \right\}. \tag{2.2.68}$$

Nótese además que, para todo $f \in PW_\pi$, y tal y como hemos definido las extensiones de estos elementos a partir de sus transformadas de Fourier con soporte contenido en $[-\pi,\pi]$, se tiene

$$f(z) = \left\langle \hat{f}, \frac{e^{-iw\bar{z}}}{2\pi} \right\rangle_{L^2[-\pi,\pi]} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{f}(w)e^{iwz} \tag{2.2.69}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\int_{\mathbb{R}} f(t)e^{-iwt}dt\right)e^{iwz}dw. \tag{2.2.70}$$

Para aplicar el teorema de Fubini sobre la última integral (2.2.70), necesitamos probar que $f(t)e^{iw(z-t)}$ es integrable en $[-\pi,\pi]\times\mathbb{R}$, para un $z \in \mathbb{C}$ fijo y arbitrario. Denotaremos $\Omega = [-\pi,\pi]\times\mathbb{R}$. Consideramos las funciones auxiliares $f_\varepsilon(t) = f(t)e^{-t^2/\varepsilon}$, para todo $\varepsilon > 0$. Estas funciones f_ε convergen puntualmente a f en todo $\mathbb{R}$, para $\varepsilon \to 0^+$. Escribimos $z = x+iy$. Así, como $|e^{iw(z-t)}| = |e^{-itw}|\cdot|e^{iwz}| = e^{\mathrm{Re}(iwz)} = e^{-wy}$, por la desigualdad de Hölder se sigue que

$$\int_\Omega |f_\varepsilon(t)e^{iw(z-t)}|dtdw \leq \left(\int_\Omega |f(t)|^2dtdw\right)^{1/2}\left(\int_\Omega e^{-2t^2/\varepsilon}e^{-2wy}dtdw\right)^{1/2}. \tag{2.2.71}$$

El primer factor resultante es finito porque $f \in L^2(\mathbb{R})$ y la medida de $[-\pi, \pi]$ es finita. El segundo factor es finito porque $e^{-2wy}e^{-2t^2/\varepsilon}$ es una función integrable en Ω. Esto demuestra que $g_\varepsilon(t, w) = f_\varepsilon(t)e^{iw(z-t)}$ está en $L^1(\Omega)$, para todo $\varepsilon > 0$.
También, $|g_\varepsilon(t, w)| \leq |f(t)|e^{-wy}e^{-t^2/\varepsilon}$ para todo $t \in \mathbb{R}$ y todo $w \in [-\pi, \pi]$. Esta última función, con los exponentes ligeramente cambiados, acabamos de probar que es integrable en Ω. Entonces, por el teorema de la convergencia dominada, para $\varepsilon \to 0^+$,

$$\int_\Omega g_\varepsilon(t, w)dtdw \;\; \to \;\; \int_\Omega f(t)e^{iw(z-t)}dtdw. \tag{2.2.72}$$

La última integral es finita en virtud del mentado teorema de Lebesgue. Acabamos de probar que $f(t)e^{iw(z-t)}$ está en $L^1(\Omega)$.
Ahora sí, estamos en condiciones de aplicar el teorema de Fubini en (2.2.70), con lo que obtenemos, para $z \neq t$,

$$\begin{aligned} f(z) &= \frac{1}{2\pi}\int_{\mathbb{R}} f(t)\left(\int_{-\pi}^{\pi} e^{iw(z-t)}dw\right)dt \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} f(t)\left(\frac{e^{iw(z-t)}}{i(z-t)}\right)\Big|_{-\pi}^{\pi}dt = \int_{\mathbb{R}} \frac{\sin(\pi(z-t))}{\pi(z-t)}f(t)dt = \langle f, \frac{\sin(\pi(t-\overline{z}))}{\pi(t-\overline{z})}\rangle_{L^2(\mathbb{R})}, \end{aligned}$$

donde se ha usado el hecho de que

$$\frac{\sin(\pi(t-\overline{z}))}{\pi(t-\overline{z})} = \frac{\sin(\pi(\overline{z}-t))}{\pi(\overline{z}-t)},$$

y que

$$\overline{\left(\frac{\sin(u)}{u}\right)} = \frac{\sin(\overline{u})}{\overline{u}}, \;\; \text{para todo } \; u \in \mathbb{C}.$$

Si $z = t$, basta repetir el argumento anterior y utilizar la prolongación continua de $\sin(\pi(z-t))/(\pi(z-t))$ a 1 que ya utilizamos en el caso real.
Por lo tanto, la función $k_\pi : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ dada por

$$k_\pi(z, w) = \frac{\sin(\pi(z-\overline{w}))}{\pi(z-\overline{w})}, \;\; z, w \in \mathbb{C}, \tag{2.2.73}$$

es el núcleo reproductor de nuestro nuevo espacio PW_π. Esto es, la extensión del espacio de Paley-Wiener a funciones enteras mantiene su propiedad reproductora, con un núcleo reproductor que es la extensión directa de $\mathbb{R}$ a $\mathbb{C}$.

2.3. Más ejemplos paradigmáticos de RKHS

2.3.1. Espacio de Bergman

En este apartado, aportamos un nuevo ejemplo de RKHS, de denso contenido histórico y matemático, que es el *espacio de Bergman*. El núcleo reproductor en este espacio fue originariamente introducido a razón del estudio de aplicaciones pseudo-conformes por medio de pares de funciones analíticas de dos variables complejas. No obstante, su utilidad para la teoría clásica de funciones con una variable compleja pronto fue descubierta por el matemático que da nombre al conjunto, Stefan Bergman. Además, la conexión establecida entre funciones de Green y aplicaciones características de estos núcleos, dio otro punto más a su estudio. De hecho, también se generalizó su uso, extendiendo la propiedad reproductora del mencionado núcleo sobre métricas hermíticas, en teoría de ecuaciones en derivadas parciales elípticas, dando un método alternativo con el que afrontar problemas con valor en la frontera.
En adelante, salvo que se diga lo contrario, trabajaremos en un abierto conexo G arbitrario en $\mathbb{C}$.

Definición 2.3.1. Sea G un abierto conexo del plano complejo $\mathbb{C}$. Denotamos por $L_a^2(G)$ al conjunto de funciones analíticas $f : G \to \mathbb{C}$ tales que

$$\int\int_G |f(x+iy)|^2 dxdy < \infty \tag{2.3.1}$$

A este conjunto $L_a^2(G)$ se le llama **espacio de Bergman en** G.

Téngase en cuenta que la integral de la definición (2.3.1) se puede entender como la medida de Lebesgue en dos dimensiones, sobre un abierto G sobre el que se dan las partes reales e imaginarias x e y. También es frecuente ver en la literatura que esta integral se denote como:

$$\int\int_G f, \quad \text{o} \quad \int_G f dA, \tag{2.3.2}$$

donde es importante notar que A es la medida de Lebesgue convencional en $\mathbb{C}$, y donde $L_a^2(G) \subset L^2(\mu)$, con $\mu = A|G$, i.e., con μ la medida dada por A y restringida al subespacio G. De esta forma, $L_a^2(G)$ tiene el producto interno y la norma naturales inducidos de $L^2(\mu)$. De esta forma, el producto interno que nosotros usamos en $L_a^2(G)$ viene dado por

$$\langle f, g \rangle = \int\int_G f(x+iy)\overline{g}(x+iy)dxdy. \tag{2.3.3}$$

Empezaremos probando que $L_a^2(G)$ es un espacio de Hilbert. Para ello, necesitamos dos resultados previos. Empezamos recordando la forma integral que se puede asignar a una función holomorfa en un abierto Ω de $\mathbb{C}$ como consecuencia directa de la integral de Cauchy en un disco. Denotamos por $B(a,r)$ a la bola abierta $\{z \in \mathbb{C} : |z-a| < r\}$, y por $\overline{B}(a,r)$ a la bola cerrada $\{z \in \mathbb{C} : |z-a| \leq r\}$, para $a \in \mathbb{C}$ y $r > 0$.
Dado $z_0 \in \mathbb{C}$, podemos tomar una bola $\overline{B}(z_0, r) \subset \Omega$, para algún $r > 0$, de tal forma que, para todo $z \in B(z_0, r)$ y denotando $S(z_0, r)$ la circunferencia o frontera de dicha bola:

$$f(z) = \frac{1}{2\pi i} \int_{S(z_0,r)} \frac{f(w)}{w-z} dw. \tag{2.3.4}$$

Así, tomando $z = z_0$, y aplicando la parametrización del camino de integración $w = re^{i\theta}$, con $\theta \in [-\pi, \pi]$, tenemos:

$$f(z_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(z_0 + re^{i\theta}) d\theta. \tag{2.3.5}$$

Lema 2.3.2. *Sean $r > 0$, y $a \in \mathbb{C}$. Si f es analítica en un entorno de $\overline{B}(a,r)$, entonces*

$$f(a) = \frac{1}{\pi r^2} \int\int_{B(a,r)} f(x+iy) dxdy. \tag{2.3.6}$$

Demostración. Notemos que, evaluando la función f en a y haciendo unas sencillas cuentas:

$$f(a) = \frac{2}{r^2} \int_0^r t f(a) dt. \tag{2.3.7}$$

Aplicamos ahora la expresión (2.3.5) sobre la función f evaluada en a, teniendo en cuenta que $0 < t \leq r$:

$$f(a) = \frac{2}{r^2} \frac{1}{2\pi} \int_0^r t \left[\int_{-\pi}^{\pi} f(a + te^{i\theta}) \right] dt. \tag{2.3.8}$$

Usando ahora, sobre la última expresión, el teorema de Fubini (nótese que estamos manejando funciones integrables en la medida de Lebesgue dada), además del cambio a coordenadas polares centradas en a, se tiene que

$$\frac{2}{r^2}\frac{1}{2\pi}\int_0^r t\left[\int_{-\pi}^{\pi} f(a+te^{i\theta})\right]dt = \frac{1}{\pi r^2}\int\int_{B(a,r)} f(x+iy)dxdy = \frac{1}{\pi r^2}\int\int_{B(a,r)} f, \tag{2.3.9}$$

con lo que tenemos la identidad buscada. □

Corolario 2.3.3. *Si $f \in L_a^2(G)$, $a \in G$, y $0 < r < \operatorname{dist}(a, \partial G)$, con ∂G la frontera de G, o bien simplemente $r > 0$ si G es no acotado, entonces*

$$|f(a)| \leq \frac{1}{r\sqrt{\pi}}||f||_2, \tag{2.3.10}$$

con $||\cdot||_2$ la norma inducida de la medida de Lebesgue natural en $\mathbb{C}$.

Demostración. Como, por hipótesis, $\overline{B}(a,r) \subset G$, la demostración se limita a usar el lema previo 2.3.2, y la desigualdad de Cauchy-Schwarz:

$$|f(a)| = \frac{1}{\pi r^2}\left|\int\int_{B(a,r)} f\cdot 1\right| \leq \frac{1}{\pi r^2}\left[\int\int_{B(a,r)} f^2\right]^{\frac{1}{2}}\left[\int\int_{B(a,r)} 1^2\right]^{\frac{1}{2}} \leq \frac{1}{\pi r^2}||f||_2 r\sqrt{\pi}, \tag{2.3.11}$$

que es justo lo que queríamos probar. □

Con lo dicho, entramos ahora a enunciar y demostrar que $L_a^2(G)$ es un espacio de Hilbert.

Teorema 2.3.4. *$L_a^2(G)$ es un espacio de Hilbert.*

Demostración. Dada la medida de Lebesgue μ restringida al abierto $G \subset \mathbb{C}$, como $L_a^2(G)$ es de esta forma un subespacio vectorial de $L^2(\mu)$, y $L^2(\mu)$ es un espacio de Hilbert, basta probar entonces que $L_a^2(G)$ es cerrado con la topología inducida por su norma. Por ser un espacio metrizable, cumple el primer axioma de numerabilidad, con lo que basta también probar que toda sucesión convergente en $L_a^2(G)$, converge a un elemento de $L_a^2(G)$. Sea pues $(f_n)_{n=1}^{\infty} \subset L_a^2(G)$ convergente a un elemento $f \in L^2(\mu)$, i.e., tal que $\int_G |f_n - f|^2 \to 0$ cuando $n \to \infty$. Veamos que $f \in L_a^2(G)$.
Sea pues $r > 0$ tal que $\overline{B}(a,r) \subset G$, y sea $0 < t < \operatorname{dist}(B(a,r), \partial G)$ (como antes, si G es no acotado, basta tomar $t > 0$). Por el corolario previo 2.3.3, existe una constante $C_t > 0$ tal que, para todos $n, m \in \mathbb{N}$, y para todo $z \in \mathbb{C}$ tal que $|z-a| \leq r$, tenemos que:

$$|f_n(z) - f_m(z)| \leq C_t||f_n - f_m||_2. \tag{2.3.12}$$

Por lo tanto, la sucesión $(f_n)_{n=1}^{\infty}$ es uniformemente de Cauchy en cualquier bola cerrada de G, ya que $||f_n - f_m||_2$ converge a cero por hipótesis. Por ello, $(f_n)_{n=1}^{\infty}$ converge uniformemente en los compactos de G.
Por el teorema de Weierstrass sobre la convergencia de sucesiones de funciones analíticas en un abierto de $\mathbb{C}$, debe existir una función analítica g en G tal que $f_n \to g$ uniformemente en dichos compactos. Y como $\int_G |f_n - f|^2 \to 0$ para $n \to \infty$, el hecho de que $L_a^2(G) \subset L^2(\mu)$, y el de que $L^2(\mu)$ sea completo, implican que, por un conocido resultado de Riesz en estas topologías de los espacios L^p, existe una subsucesión $(f_{n_k})_{k=1}^{\infty}$ tal que $f_{n_k}(z) \to f(z)$ casi siempre con respecto a μ. Entonces, $f = g$ casi siempre en respecto a μ en G. Así, como g es analítica en G, ya tenemos un representante de la clase de equivalencia de f en $L^2(\mu)$ que es función analítica en dicho abierto. □

Veamos ahora que $L_a^2(G)$ es un RKHS, para lo que mantenemos, como punto de referencia, el abierto conexo de $\mathbb{C}$. Apreciará el lector que, implícitamente, la afirmación de que $L_a^2(G)$ es un RKHS ya sido probada.

Teorema 2.3.5. *Sea $G \subset \mathbb{C}$ abierto y conexo. Entonces, $L_a^2(G)$ es un RKHS.*

Demostración. Ya hemos probado que $L_a^2(G)$ es un espacio de Hilbert.
Para ver que $L_a^2(G)$ es un RKHS, tendríamos que ver que los funcionales evaluación δ_w, en puntos $w \in G$, son continuos o, equivalentemente, acotados. Utilizando el corolario previo 2.3.3, sabemos que, para todo $f \in L_a^2(G)$, existe un $C > 0$ tal que

$$|\delta_w(f)| = |f(w)| \leq C||f||_2, \tag{2.3.13}$$

lo que prueba la acotación de δ_w. □

A continuación, hagamos una apreciación muy característica de este espacio. Recordemos que, por el corolario 2.3.3, es inmediato probar que $L_a^2(\mathbb{C}) = \{0\}$, ya que si podemos aplicar, sobre cualquier $a \in \mathbb{C}$, la desigualdad

$$|f(a)| \leq \frac{1}{r\sqrt{\pi}}||f||_2, \tag{2.3.14}$$

haciendo tender r a infinito, tenemos que $f = 0$.
Nos centramos en un tipo particular de espacio de Bergman. Consideramos, en adelante, el espacio de Bergman definido sobre el abierto conexo $G = B = B(0,1)$. Como estamos trabajando e integrando en B, estamos manejando variables complejas w, z tales que $|w| < 1$, y $|z| < 1$. Por lo tanto, utilizando la segunda desigualdad triangular,

$$|1 - \overline{w}z| \geq ||1| - |\overline{w}z|| > (1 - |z|) \implies \frac{1}{|1 - \overline{w}z|^2} < \frac{1}{(1 - |z|)^2}. \tag{2.3.15}$$

De esta forma, tomando módulos, podemos acotar la integral que sigue de la siguiente forma

$$\int\int_B \left|\frac{1}{[\pi(1 - z\overline{w})^2]}\right|^2 dw \leq \int\int_B \left|\frac{1}{[\pi(1 - |z|)^2]}\right|^2 dw < \infty. \tag{2.3.16}$$

Así, estamos en condiciones de probar la convergencia de la integral para $f \in L_a^2(B)$

$$f(z) = \int\int_B \frac{f(w)}{\pi(1 - z\overline{w})^2} dw. \tag{2.3.17}$$

Para ello, identificamos (2.3.17) con el producto interno definido en $L_a^2(B)$, y aplicando la desigualdad de Cauchy-Schwarz, tenemos que

$$|f(z)| = \left|\int\int_B \frac{f(w)}{\pi(1 - z\overline{w})^2} dw\right| = |\langle f, g_z\rangle| \leq ||f||_2 \cdot ||g_z||_2, \tag{2.3.18}$$

donde $g_z(w) = 1/[\pi(1 - \overline{w}z)^2]$, para todo $w \in B$, y para un $z \in B$ prefijado. Por lo tanto, como $f \in L_a^2(B)$, entonces $||f||_2 < \infty$; y por la desigualdad (2.3.16), tenemos $||g||_2 < \infty$. Esto prueba la convergencia de (2.3.17).
Con todo ello, ahora podemos calcular el núcleo reproductor de este espacio. Todo girará en torno a la expresión, probada para todo $f \in L_a^2(B)$, dada en (2.3.17). Vista ya la convergencia de la integral, veamos ahora que verdaderamente se satisface la igualdad (2.3.17) en todo $L_a^2(B)$. Lo enunciamos y probamos.

Proposición 2.3.6. *Sea $B = B(0,1) \subset \mathbb{C}$ la bola unidad en $\mathbb{C}$, centrado en un punto $a \in \mathbb{C}$. Sea $f \in L_a^2(B)$. Entonces, para todo $z \in B$, se tiene:*

$$f(z) = \int\int_B \frac{f(w)}{\pi(1 - z\overline{w})^2} d\mu. \tag{2.3.19}$$

Demostración. Aplicamos una parametrización del camino de integración $w = re^{i\theta}$, para $r \in [0,1]$, y $\theta \in [0, 2\pi)$:

$$\int\int_B \frac{f(w)}{(1-z\overline{w})^2} d\mu = \int_0^1 \int_0^{2\pi} \frac{f(re^{i\theta})}{(1-zre^{-i\theta})^2} r d\theta dr. \tag{2.3.20}$$

Utilizando las expresiones de la parametrización dada, se tiene que $d\theta = -idw/w$, $e^{-i\theta} = r/w$, luego

$$\begin{aligned}\int_0^1 \int_0^{2\pi} \frac{f(re^{i\theta})}{(1-zre^{-i\theta})^2} r d\theta dr &= \int_0^1 \int_{\partial B} \frac{f(w)}{(1-z\frac{r^2}{w})^2} r \left(\frac{-idw}{w}\right) dr \\ &= \int_0^1 \int_{\partial B} \frac{-irf(w)}{(\frac{w-zr^2}{w})^2} \frac{dw}{w} dr = \int_0^1 \int_{\partial B} \frac{-irwf(w)}{(w-zr^2)^2} dwdr.\end{aligned} \tag{2.3.21}$$

Notemos ahora que, como $|z| < 1$, y $r \leq 1$, entonces $|zr^2| < 1$. Entonces, la función

$$h(w) = \frac{-irwf(w)}{(w-zr^2)^2} \tag{2.3.22}$$

es holomorfa en $B \setminus \{zr^2\}$, con un polo doble en zr^2. Aplicando entonces el teorema de los residuos sobre la integral de h a lo largo del soporte de la curva que parametriza al contorno de B, i.e., a ∂B, tenemos:

$$\int_{\partial B} \frac{-irwf(w)}{(w-zr^2)^2} dwdr = 2\pi i \cdot \mathrm{Res}\left[\frac{-irwf(w)}{(w-zr^2)^2}, zr^2\right] = 2\pi i \cdot \mathrm{Res}\left[h(w), zr^2\right]. \tag{2.3.23}$$

Dado que zr^2 es un polo doble de h, calculamos el residuo como sigue:

$$\begin{aligned}\mathrm{Res}\left[h(w), zr^2\right] &= \lim_{w\to zr^2} \frac{d}{dw}\left((w-zr^2)^2 \frac{-irwf(w)}{(w-zr^2)^2}\right) \\ &= \lim_{w\to zr^2} \frac{d}{dw}(-irwf(w)) = \lim_{w\to zr^2}(-irf(w) - irwf'(w)) \\ &= -irf(zr^2) - ir^3 zf'(zr^2).\end{aligned}$$

Por lo tanto, la integral (2.3.21) se convierte en:

$$\int_0^1 \int_{\partial B} \frac{-irwf(w)}{(w-zr^2)^2} dwdr = \int_0^1 \pi(2rf(zr^2) + 2zr^3 f'(zr^2))dr = \int_0^1 \pi \frac{d}{dr}(r^2 f(zr^2))dr = \pi f(z). \tag{2.3.24}$$

Entonces, para todo $z \in B$,

$$f(z) = \frac{1}{\pi} \int_B \frac{f(w)}{(1-\overline{w}z)^2} d\mu. \tag{2.3.25}$$

□

Tenemos pues la forma analítica explícita del único núcleo reproductor del RKHS $L_a^2(B)$.

Definición 2.3.7. La función $k : B \times B \to \mathbb{C}$, dada por $k(z,w) = \frac{1}{\pi(1-\overline{w}z)^2}$ para todos $z, w \in B$, es el **núcleo de Bergman** sobre la bola unidad $B = B(0,1)$.

Notemos que $k(z,w)$ es una función simétrica conjugada (i.e., $B(z,w) = \overline{B(w,z)}$, para todos $w, z \in B$), holomorfa con respecto a z y antiholomorfa con respecto a w (i.e., su función conjugada $\overline{k}$ es holomorfa con respecto de w). Notemos la riqueza de la construcción seguida: el operador integral correspondiente a esta función $k(z,w)$, integrada sobre w en B, reproduce cualquier función holomorfa f en B, con tal de que f sea de cuadrado integrable en dicho abierto.

2.3.2. Espacio de Hardy ponderado sobre $B(0,R)$

Los espacios de Hardy fueron introducidos por Frigyes Riesz en 1923, aunque utilizados ya en 1915 por el propio Goldfrey Harold Hardy. La teoría de funciones analíticas de estos espacios fue ampliamente estudiada en el siglo XX, y sigue siendo estudiada hoy. Caracterizaremos los espacios de Hardy pidiendo determinadas propiedades a sus series de potencias centradas en el cero. Veremos que de ello se seguirá una acotación uniforme de todas las integrales sobre circunferencias de radio en un intervalo $(0,R)$, para $R>0$ el radio de convergencia de dichas series. Estos son pasos indispensables para su caracterización como RKHSs.

Definición 2.3.8. Sea $\beta=\{\beta_n\}_{n=0}^{\infty}\subset\mathbb{R}$ una sucesión tal que $\beta_n>0$ para todo $n\in\mathbb{N}_0$, y sea $r>0$ un número real positivo fijo tal que satisface $r=\liminf_{n\to\infty}(\beta_n)^{\frac{1}{n}}>0$. Definimos el **espacio de Hardy ponderado con** β como el conjunto de funciones

$$H^2_\beta:=\{f\in\mathcal{H}(B(0,r)):f(z)=\sum_{n=0}^{\infty}a_nz^n\ ,\ \ z\in B(0,r),\ \sum_{n=0}^{\infty}\beta_n^2|a_n|^2<\infty\},\tag{2.3.26}$$

donde $\mathcal{H}(B(0,r))$ denota al conjunto de funciones holomorfas en $B(0,r)\subset\mathbb{C}$.

El caso en que $\beta_n=1$ para todo $n\in\mathbb{N}_0$, y $r=1$, es el llamado, por antonomasia, espacio de Hardy. A continuación, justificamos esta definición.
Para ello, definimos en H^2_β el siguiente producto interno:

$$\langle f,g\rangle=\sum_{n=0}^{\infty}\beta_n^2a_n\bar{b}_n,\tag{2.3.27}$$

con $f,g\in H^2_\beta$ tales que $f(z)=\sum_{n=0}^{\infty}a_nz^n$ y $g(z)=\sum_{n=0}^{\infty}b_nz^n$ son sus expresiones en series de potencias en 0, para un radio de convergencia que ahora acotaremos. Con el producto interno así definido, note el lector que la serie de números reales y positivos $\sum_{n=0}^{\infty}\beta_n^2|a_n|^2$ es precisamente la norma al cuadrado de f, esto es, $||f||^2_\beta$.
Estudiemos ahora el radio de convergencia de la serie de potencias con que hemos expresado f en la definición de H^2_β. Por la condición necesaria de convergencia de la sucesión de términos reales de una serie convergente, tenemos que

$$\lim_{n\to\infty}\beta_n^2|a_n|^2=0\quad\Longrightarrow\quad\lim_{n\to\infty}\beta_n|a_n|=0.\tag{2.3.28}$$

Por tanto, existe un $N\in\mathbb{N}$ tal que, para todo $n\geq N$, tenemos

$$\beta_n|a_n|\leq 1\quad\Longrightarrow\quad|a_n|\leq\beta_n^{-1}.\tag{2.3.29}$$

Entonces, por el criterio de Hadamard sobre series de potencias, el radio de convergencia de f satisface:

$$R_f^{-1}=\limsup_{n\to\infty}|a_n|^{1/n}\leq\limsup_{n\to\infty}\beta_n^{-1/n}=(\liminf_{n\to\infty}\beta_n^{1/n})^{-1}.\tag{2.3.30}$$

Así, estamos en condiciones de enunciar el siguiente lema.

Lema 2.3.9. *Dada $f\in H^2_\beta$ expresable, en términos de serie de potencias centradas en el 0, en la forma*

$$f(z)=\sum_{n=0}^{\infty}a_nz^n,\tag{2.3.31}$$

entonces, su radio de convergencia R_f satisface

$$R_f\geq R:=\liminf_{n\to\infty}\beta_n^{1/n}.\tag{2.3.32}$$

Nótese que esto justifica la elección del $R > 0$ de la definición del espacio de Hardy ponderado H^2_β.
Nosotros, por definición, consideramos el radio de convergencia de los elementos de H^2_β como el propio límite inferior de β. Con ello, el radio de la bola de definición de todos los elementos de este espacio queda determinado por dicha sucesión.
Obsérvese también que la elección $\liminf_{n\to\infty}(\beta_n)^{\frac{1}{n}} > 0$ evita situaciones no interesantes, con radios de convergencia nulos. Como último apunte, el caso $\liminf_{n\to\infty}(\beta_n)^{\frac{1}{n}} = \infty$ determina un espacio de Hardy ponderado de funciones enteras.
Por el criterio de Hadamard, si $\liminf_{n\to\infty}\beta_n^{1/n} = R > 0$, la serie de potencias dada en 2.3.31 converge absolutamente en $B(0, R)$, y normalmente (luego uniformemente) en $\overline{B}(0, \rho)$, para todo $\rho < R$.
Ahora, vamos a probar la completitud de H^2_β, necesaria si queremos demostrar su propiedad reproductora.

Proposición 2.3.10. *H^2_β es un espacio de Hilbert.*

Demostración. Empezamos diciendo que, con la definición (2.3.26), H^2_β es un espacio vectorial, para lo que basta aplicar la desigualdad triangular sobre los módulos involucrados en la segunda expansión de dicha definición. Por otra parte, es un espacio con producto interno dado por (2.3.27), ya que satisface la sesquilinealidad, la positividad, y que si, para algún f que se extiende como $f(z) = \sum_{n=1}^\infty a_n z^n$ en algún entorno del 0, se tiene que $\langle f, f\rangle = \sum_{n=1}^\infty \beta_n^2 |a_n|^2 = 0$, entonces $a_n = 0$ para todo $n \in \mathbb{N}$, pues $\beta_n > 0$ para todo $n \in \mathbb{N}$. Por lo tanto, $f(z) = 0$ para todo $z \in B(0, R_f)$, con $R_f > 0$ su radio de convergencia. Por supuesto, $\langle f, g\rangle = \overline{\langle g, f\rangle}$ para todo $f, g \in H^2_\beta$.
Nos falta ver la completitud. Para ello, es suficiente probar que existe un isomorfismo isométrico entre H^2_β, y el espacio $l^2_\beta(\mathbb{C})$, definido como el siguiente conjunto:

$$l^2_\beta(\mathbb{C}) = \{a = (a_0, a_1, \ldots, a_n, \ldots) : a_j \in \mathbb{C},\ \ j \in \mathbb{N}_0,\ \sum_{i=0}^\infty |a_i|^2\beta_i^2 < \infty\}. \tag{2.3.33}$$

La correspondencia biyectiva está clara, pues dada $f \in H^2_\beta$, es decir, analítica en $B(0, R)$ con algún $R > 0$, y tal que, dada su expansión en serie de potencias centradas en 0, tenemos una sucesión $\{a_i\}_{i=0}^\infty \subset \mathbb{C}$ de los coeficientes de dicha expansión que satisface

$$\sum_{i=0}^\infty |a_i|^2\beta_i^2 < \infty, \tag{2.3.34}$$

entonces bastaría tomar $\{a_i\}_{i=0}^\infty \in l^2_\beta(\mathbb{C})$.
En el otro sentido es exactamente igual: dado $\{a_i\}_{i=0}^\infty \in l^2_\beta(\mathbb{C})$, construimos una función f analítica en $B(0, R)$ a través de una serie de potencias dada por

$$f(z) = \sum_{i=0}^\infty a_i z^i, \tag{2.3.35}$$

la cual satisface la finitud de la norma en H^2_β por satisfacer la de la norma en $l^2_\beta(\mathbb{C})$. Así, la biyección está clara.
Notemos además que, tal y como lo hemos definido, las normas son exactamente las mismas, haciendo la lectura de que en $l^2_\beta(\mathbb{C})$ hablamos de sucesiones de valores complejos, y en H^2_β hablamos de coeficientes de series de potencias. Entonces, tenemos una biyección que además es isometría. Y, como $l^2_\beta(\mathbb{C})$ es completo, entonces H^2_β también. □

Ahora, vamos a describir una característica esencial de las funciones de H^2_β. Particularizamos al caso $R = 1$, pues éste ilustra lo que queremos mostrar, y es un resultado que no utilizaremos más adelante.

Teorema 2.3.11. *Sean $f \in H^2_\beta$, con β tal que $\beta_n = 1$ para todo $n \in \mathbb{N}$, de tal forma que f sea representable en todo $B(0,1)$ por la siguiente serie de potencias:*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \quad a_n \in \mathbb{C} \ , \ z \in B(0,1). \tag{2.3.36}$$

Entonces,

$$\sup_{r\in(0,1)} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta < \infty. \tag{2.3.37}$$

Recíprocamente, si f es analítica en $B(0,1)$ y el último superior es finito, entonces $f \in H^2_\beta$, con $\beta_n = 1$ para todo $n \in \mathbb{N}_0$.

Demostración. Consideramos una función analítica f en $B(0,1)$ que satisfaga las hipótesis. Entonces, podemos expresar f en serie de potencias centrada en 0 de la siguiente forma

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \quad z \in B(0,1). \tag{2.3.38}$$

Sabemos, en virtud del criterio de Hadamard, que esta serie de potencias converge en $B(0,1)$ absolutamente, y uniformente en $\overline{B}(0,r)$ para todo $r \in [0,1)$.
Tomamos $z = re^{i\theta}$ para algún $r \in [0,1)$ y $\theta \in [0,2\pi]$. Entonces,

$$f(re^{i\theta}) = \sum_{n=0}^{\infty} a_n r^n e^{in\theta}, \tag{2.3.39}$$

es una serie de Fourier absolutamente convergente en $[0,2\pi]$, con coeficientes $a_n r^n$ si $n \geq 0$, y coeficientes nulos si $n < 0$.
Denotamos por $\tilde{L}^2_{\mathbb{C}}[0,2\pi]$ al espacio de funciones de cuadrado integrable en $[0,2\pi]$, sobre el que definimos la siguiente norma:

$$||g||_2 = \left(\frac{1}{2\pi} \int_0^{2\pi} |g(t)|^2 dt \right)^{1/2}. \tag{2.3.40}$$

Esta consideración nos permite recurrir a la isometría lineal $U : \tilde{L}^2_{\mathbb{C}}[0,2\pi] \to l^2(\mathbb{Z})$, dada como $U(f) = \hat{f}$, con $\hat{f}$ el elemento de $l^2(\mathbb{Z})$ cuyos términos son los coeficientes de la serie de Fourier de f, i.e., para todo $n \in \mathbb{N}$,

$$\hat{f}_n = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} dt. \tag{2.3.41}$$

Como U es isometría, y dado que $f(re^{i\theta}) \in \tilde{L}^2_{\mathbb{C}}[0,2\pi]$ para cada $r \in [0,1)$ fijo,

$$\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta = ||f||_2^2 = ||U(f)||_{l^2}^2 \tag{2.3.42}$$

$$= ||\hat{f}||_{l^2}^2 = \sum_{n=0}^{\infty} |a_n|^2 r^{2n}. \tag{2.3.43}$$

Por la última igualdad, el superior de las integrales

$$\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta,$$

cuando $r \in [0,1)$, coincide con el superior de las últimas series dadas, que en todo caso será menor o igual que $\sum_{n=0}^{\infty} |a_n|^2$, la cual es finita siempre y cuando f esté en H^2_β.

Recíprocamente, supongamos que $f \notin H^2_\beta$, entonces la serie $\sum_{n=0}^{\infty} |a_n|^2$ diverge. Esto implica que si se hace tender $r \in (0,1)$ a 1, las sumas

$$\sum_{n=0}^{\infty} |a_n|^2 r^{2n},$$

se hacen arbitrariamente grandes, con lo que

$$\sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta = \sup_{0<r<1} \sum_{n=0}^{\infty} |a_n|^2 r^{2n} = \infty,$$

como queríamos probar.

□

Aunque nos hemos quedado en el caso de $B(0,1)$, este resultado se puede extender a bolas $B(0,R)$ con $R>0$ arbitrario, lo que ilustra el tipo de función que hay en un espacio de Hardy ponderado: funciones analíticas en una bola cuyo radio R está determinado por la sucesión β, y con una acotación uniforme de la integral dada en (**??**) sobre circunferencias de radio en $(0,R)$.
Estudiamos ahora la condición de RKHS de un espacio H^2_β, con $\beta = \{\beta_n\}_{n=0}^{\infty}$ una sucesión tal que $\liminf_{n\to\infty} \beta_n^{1/n} = R \in (0,\infty)$. Así, para todo $w \in B(0,R)$ definimos la función que sigue

$$k_w(z) = k(w,z) = \sum_{n=0}^{\infty} \frac{\overline{w}^n z^n}{\beta_n^2}. \tag{2.3.44}$$

Para empezar, probamos que $k_w \in H^2_\beta$. Esto se sigue de su propia definición, ya que dados sus coeficientes $a_n = \overline{w}^n/\beta_n^2$ en su serie de potencias centrada en el origen, se tiene

$$\sum_{n=0}^{\infty} |a_n|^2 \beta_n^2 = \sum_{n=0}^{\infty} \left|\frac{\overline{w}^n}{\beta_n^2}\right|^2 \beta_n^2 = \sum_{n=0}^{\infty} \frac{|\overline{w}|^{2n}}{\beta_n^2}. \tag{2.3.45}$$

Notemos que $|w| < R$, luego podemos tomar $r_0 \in (|w|, R)$. Entonces, existe un $n_0 \in \mathbb{N}_0$ tal que $\beta_n^{1/n} > r_0$ para todo $n \geq n_0$. De esta forma,

$$\left|\frac{\overline{w}^n}{\beta_n}\right| = \left(\frac{|\overline{w}|}{\beta_n^{1/n}}\right)^n < \left(\frac{r_0}{\beta_n^{1/n}}\right)^n < \zeta^n, \;\; n \geq n_0, \tag{2.3.46}$$

con $\zeta \in (0,1)$ constante. Así, hemos acotado los términos de la última serie de (2.3.45) por términos de una serie geométrica convergente.
De este modo, así definido, hemos probado que $k_w \in H^2_\beta$.

Teorema 2.3.12. *El espacio H^2_β es un RKHS con núcleo $k : B(0,R) \times B(0,R) \to \mathbb{C}$ dado por* (2.3.44).

Demostración. Sea $f \in H^2_\beta$, analítica en $B(0,R)$, para el $R>0$ sobre el que se define k. Escribimos su serie de potencias como

$$f(w) = \sum_{n=0}^{\infty} a_n w^n, \;\; w \in B(0,R). \tag{2.3.47}$$

Entonces, para todo $w \in \mathbb{C}$ tal que $|w| < R$, se tiene que

$$\langle f, k_w \rangle = \sum_{n=0}^{\infty} \beta_n^2 \left[a_n \cdot \left(\frac{w^n}{\beta_n^2}\right)\right] = \sum_{n=0}^{\infty} a_n w^n = f(w), \tag{2.3.48}$$

lo que prueba la condición de reproducibilidad en H^2_β a través de lo que, como acabamos de ver, es su núcleo reproductor: $k(w,z)$. Esto demuestra que H^2_β es un RKHS. □

Hay diversos ejemplos particulares de espacios de Hardy ponderados de enorme importancia en la teoría de funciones. Enumeramos algunos de ellos:

1. El *espacio de Segal-Bargmann*, que es el espacio de Hardy ponderado con la sucesión $\beta_n = \sqrt{n!}$ para $n \in \mathbb{N}_0$. Como $\liminf_{n\to\infty}(n!)^{1/2n} = \infty$, entonces este espacio está formado por funciones enteras, y su núcleo reproductor viene dado por

$$k(w,z) = \sum_{n=0}^{\infty} \frac{\overline{w}^n z^n}{n!} = e^{\overline{w}z}. \tag{2.3.49}$$

2. El *espacio de Dirichlet* se obtiene tomando los pesos $\beta_n = \sqrt{n}$ para $n \in \mathbb{N}$, $\beta_0 = 1$. De ello, se tiene que

$$\liminf_{n\to\infty} \left[n^{1/2n}\right] = 1, \tag{2.3.50}$$

con lo que este espacio contiene funciones analíticas en $B(0,1)$. En cuanto al núcleo reproductor, resulta ser

$$k(w,z) = \sum_{n=1}^{\infty} \frac{\overline{w}^n z^n}{n} + 1 = \log(1-\overline{w}z) + 1. \tag{2.3.51}$$

Capítulo 3

Aprendizaje Estadístico y RKHS

En este capítulo, trataremos algunas nociones de la teoría de Aprendizaje Estadístico desde el punto de vista de la aplicación que los RKHSs tienen sobre ésta. La idea general de esta teoría es la de estudiar, sobre un conjunto X de datos de entrada que dependerán del problema a tratar, una aplicación $L(x, y, f(x))$, que llamaremos *pérdida*, y que describe la discrepancia entre la función de predicción dada por $f(x)$, para cada elemento x de dicho espacio de entrada, y la observación dada por y en el punto $x \in X$. A la pérdida L le asociamos un *riesgo* entendido como el promediado de la futura pérdida dada por f. Por supuesto, concretaremos las pertinentes definiciones para todos estos objetos.
Nos centraremos en los métodos de *Máquinas de Soporte Vectorial*, o *Support Vector Machines* (SVM). Estos métodos son técnicas de *aprendizaje supervisado* basadas en la construcción de hiperplanos en un conjunto en correspondencia directa con X, de cara a hacer la pertinente clasificación de datos, en función de la información que estemos manejando. En particular, la idea que a nosotros nos interesa, y en la que basaremos nuestro ejemplos, es la de clasificar los datos en un espacio X arbitrario. Veremos que esto se puede conseguir enviando dichos elementos de X a espacios de Hilbert con núcleo reproductor que, además, aportan una condición de clasificación (como sería la de fijar $b \in H$, tomar $g(\cdot) = \langle \cdot, b \rangle_H$, y construir un hiperplano con esta expresión). A partir de los numerosos resultados de representación demostrados, entre los que destaca el teorema de representación de Mercer, estos espacios presentan la enorme ventaja de que los hiperplanos de H que fijan nuestra clasificación, vienen dados por combinaciones lineales como sigue

$$\sum_i \alpha_i k(x_i, x), \tag{3.0.1}$$

con k el correspondiente núcleo[1] de H, $\alpha_i \geq 0$ unos parámetros positivos, y x y x_i valores del espacio de datos de entrada (con los x_i prefijados, y x el dato arbitrario de entrada sobre el que hacer una predicción). Esta construcción nos permite controlar, con una sola función k, la cercanía que los nuevos datos $x \in X$ presentan respecto a datos previos o presentes del mismo conjunto.
Note el lector que el estudio que sigue es profundamente teórico. Esto quiere decir que lo que a nosotros nos interesa es solamente la teoría del Análisis Funcional, con livianas referencias a la Teoría de la Medida, que está involucrada en muchas de estas construcciones, aportando algunos ejemplos triviales que lo ilustren.

3.1. Formulación general en clave de RKHS

Necesitamos sentar una serie de ideas para poder describir el uso del Análisis Funcional en el estudio de un problema de Aprendizaje Estadístico. Consideraremos que todos los valores x del

[1] Recordemos que el espacio de núcleos en X sobre un espacio de Hilbert H conforman un cono convexo.

input forman parte de un conjunto X. Mentando a los ejemplos de la introducción, X puede estar formado por los posibles valores de determinados ensayos clínicos, o todas las localizaciones de las viviendas en una ciudad. Consideraremos que Y es un conjunto, también conocido, de los datos de respuesta. Por ejemplo, de nuevo en relación a los ejemplos anteriores, puede tomar formas como $\{\text{positivo}, \text{negativo}\}$, o como el rango de todos los posibles precios de una vivienda, para lo que valdría tomar el conjunto $\mathbb{R}^+$.
En el método convencional, se parte de una secuencia finita de pares $D := \{(x_1, y_1), \ldots, (x_n, y_n)\}$, para algún $n \in \mathbb{N}$, de datos de entrada y salida conocidos, y se denota por $f : X \to Y$ a la *función de aprendizaje* que, presumiblemente, será una buena aproximación de los datos de respuesta frente a un $x \in X$ arbitrario.
En este trabajo, supondremos que los pares $(x, y) \in X \times Y$ son independientes y generados por una distribución de probabilidad dada por la probabilidad P, a priori desconocida, sobre el conjunto $X \times Y$, con una σ-álgebra que variará en función del modelo de estudio. Los valores $x \in X$ vendrán dados por una distribución de probabilidad marginal P_X, también a priori desconocida. Además, asumiremos que los valores respuesta $y \in Y$ están estocásticamente generados por una distribución de probabilidad condicionada a los valores de X, que denotaremos por $P(\cdot|x)$. Esto formaliza el hecho de que, en general, la información contenida en el valor de entrada x puede no ser suficiente para determinar el valor de salida, lo que engloba las dos situaciones extremas: que a cada $x \in X$ le corresponda biunívocamente un valor de Y, y que el input sea totalmente irrelevante para los valores respuesta. También, la asunción de que la distribucion $P(\cdot|x)$ sobre el conjunto de valores de salida Y es a priori desconocida incide en el hecho de que, de antemano, no tenemos ninguna descripción razonable de la relación entre los valores de X y de Y. Esta es la diferencia fundacional entre los modelos de Aprendizaje Estadístico y los *modelos paramétricos*, en que la relación entre los valores de entrada y de salida es uno de los puntos de partida, a partir de una función $f \in \mathcal{F}$ tomada de un espacio finito y conocido de funciones con valores en un cuerpo $\mathbb{K}$.
Hasta ahora, solo hemos descrito la naturaleza, con vaga generalidad, de los datos involucrados en nuestro estudio. Para mayor concreción de lo expuesto, supondremos que el conjunto Y será siempre un subconjunto de $\mathbb{R}$, y que f tomará valores reales. Introducimos ahora la función que medirá la calidad de nuestras predicciones dadas por f, $L(x, y, f(x))$, a la que nos referiremos como la *función de pérdida*, la cual será una función $L : X \times Y \times \mathbb{R} \to [0, \infty)$ tanto menor cuanto mejor aproximacion nos dé f. Es intuitivo pensar, y así será para nosotros, que la evaluación de la calidad de f dada por L no vendrá dada por el conocimiento de $L(x, y, f(x))$ para cada par (x, y) en $X \times Y$, sino por el promedio de estos valores sobre un espacio de medida dado. Por supuesto, esto se podría hacer de muchas formas, pero la de uso más extendido en la teoría de Aprendizaje Estadístico es la de calcular la pérdida *esperada* de f, i.e., la cantidad

$$\mathcal{R}_{L,P}(f) := \int_{X\times Y} L(x, y, f(x))dP(x, y) = \int_X \int_Y L(x, y, f(x))dP(x|y)dP_X, \tag{3.1.1}$$

a la cual llamaremos *riesgo de* f, y donde $\mathcal{F}$ es el espacio de funciones en el que ubicamos a los elementos f.
Nosotros nos restringimos a métodos SVM. La esencia de estos es encontrar el llamado *objetivo de aprendizaje* (*learning goal*), por el cual una vez enviados los elementos de X a un espacio de Hilbert H, queremos encontrar f^*, *función objetivo*, que alcanza, o más se aproxima, al valor más pequeño del riesgo, i.e., al valor

$$\mathcal{R}^*_{L,P} = \inf_{f:X\to\mathbb{R}} \mathcal{R}_{L,P}(f). \tag{3.1.2}$$

A lo largo del trabajo, llamaremos indistintamente *funciones objetivo* o *minimizadores* a las funciones f que nos dan $\inf_{f\in\mathcal{F}} \mathcal{R}_{L,P}(f)$, o que mejor se aproximan a éste.
No obstante, como antes hemos dicho, la distribución de probabilidad dada por P es a priori desconocida, con lo que, en la práctica, se juega con muestras finitas de pares conocidos con valores de

entrada y de salida $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$, calculando un *riesgo empírico*, de la forma:

$$\mathcal{R}_{L,D}(f) := \frac{1}{n}\sum_{i=1}^{n} L(x_i, y_i, f(x_i)). \tag{3.1.3}$$

A pesar de que, para un f fijo, por la ley fuerte de los grandes números tengamos una mejor aproximación al riesgo real $\mathcal{R}_{L,P}(f)$ para un cardinal cada vez mayor de D, i.e., para una suma cada vez más extendida, es fácil caer en el error del *overfitting*[2]. Este error consiste en que la función objetivo f^* podría memorizar solamente lo que pase en los pares de la muestra conocida, dando una aproximación muy errada de valores que se alejan un poco de estos.
Un método común con el que afrontar este problema es el de seleccionar determinados espacios de funciones $\mathcal{F} = \{f : X \to \mathbb{K}\}$ con los que, aunque podamos caer en overfitting, al menos podamos controlarlo, y con ello evitarlo. De esta forma, en lugar de plantear la teoría con una función de riesgo que recorra la totalidad de las funciones de X en $\mathbb{K}$, lo restringimos a $\mathcal{F}$. No obstante, este planteamiento sigue dando dos problemas:

1. En aquellas situaciones en que nos interese conocer también la distribución P que desconocemos en un principio, el limitarnos a ciertos espacios $\mathcal{F}$ no nos da ninguna ventaja, pues esto solo aplaca la carencia que teníamos para encontrar una función que minimice $\mathcal{R}_{L,D}$, nada nos dice sobre P. En otras palabras, no hay ninguna garantía de que el *error del modelo* o *error de aproximación*:
$$\mathcal{R}^*_{L,P,\mathcal{F}} - \mathcal{R}^*_{L,P}, \tag{3.1.4}$$
sea suficientemente bajo o, siendo más prácticos, sea controlable por alguna cota o algún comportamiento asintótico.

2. Con lo que sabemos hasta ahora de una SVM, el cálculo de $\inf_{f\in\mathcal{F}}\mathcal{R}_{L,D}(f)$ puede ser impracticable, o al menos no hay nada que nos diga lo contrario. Este punto lleva a afrontar prácticamente cualquier modelo de aprendizaje a través de un modelo alternativo de más fácil cálculo, dado por una pérdida L' distinta, con el que no perdamos el control sobre el error de aproximación, es decir:
$$\mathcal{R}_{L,P}(f) - \mathcal{R}^*_{L,P} \leq \mathcal{R}_{L',P}(f) - \mathcal{R}^*_{L',P}, \tag{3.1.5}$$
de tal forma que una acotación de este error para el modelo alternativo también sirve para el modelo original.

Ambos puntos pueden ser hábilmente confrontados con el uso de RKHSs como los espacios de funciones en que restringir nuestra búsqueda de f^*. Una de las principales razones es lo práctico que se hace el proceso de búsqueda de $\inf_{f\in\mathcal{F}}\mathcal{R}_{L,D}(f)$, a través de la búsqueda de minimizadores $f_{D,\lambda}$ de la expresión

$$\inf_{f\in H}\left(\lambda||f||_H^2 + \mathcal{R}_{L,D}(f)\right), \tag{3.1.6}$$

para un $\lambda > 0$, donde H es el correspondiente RKHS, y donde $\lambda||f||_H^2$ es el llamado *término de regularización*, usado para penalizar a las funciones f con una norma en el RKHS H demasiado grande. Uno de los motivos de esta penalización es el de que las funciones objetivo f que padecen de overfitting tienden a tener normas en H muy grandes, de tal forma que este término puede erradicar este error. A la expresión (3.1.6) sobre la que buscamos un ínfimo, le llamaremos *riesgo empírico regularizado* o, simplemente, RER.
Este manejo práctico vía RKHSs para encontrar la función objetivo se fundamenta en la propiedad reprodutora del núcleo $k : X \times X \to \mathbb{R}$ de H, lo que reduce la búsqueda de funciones $f_{D,\lambda}$ a expresiones del tipo

$$f_{D_\lambda} = \sum_{i=1}^{n}\alpha_i k(x_i, \cdot), \tag{3.1.7}$$

[2]La mejor traducción de *overfitting* al castellano, en este contexto, podría ser la de sobrecarga, en tanto en cuanto un exceso de valores input nos desajusta predicciones futuras.

con $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Nótese que, tanto las funciones de H dadas por $k(\cdot, x_i)$, como los parámetros α_i, para $i \in \{1, \ldots, n\}$, dependen de los datos de entrada, así como del espacio de Hilbert, pero de nada más. La esencia de esta simplificación radica en la estructura del subespacio (1.3.13), denso en el espacio RKHS correspondiente. Por supuesto, este enfoque es válido para la búsqueda de funciones objetivo en espacios dimensionalmente infinitos H, tales como, y esto lo veremos, el dado por la reproducción del *núcleo gaussiano de base radial*, o simplemente núcleo RBF (*Radial Basis Function kernel*), dado como sigue

$$k_\gamma(x, x') := \exp(-\gamma^{-2}||x - x'||_2^2), \quad x, x' \in X, \tag{3.1.8}$$

con $X = \mathbb{R}^d$ para algún $d \in \mathbb{N}$, y con $\gamma > 0$ un parámetro de anchura.

3.2. Funciones de pérdida y de riesgo

En esta sección, sentamos las ideas y resultados más importantes sobre funciones de pérdida y de riesgo.
Empecemos recordando que, dado un espacio topológico (X, τ), la *σ-álgebra de Borel* viene dada por $\sigma(X) := \sigma(\tau)$, i.e., puede ser definida como la σ-álgebra engendrada por los abiertos de esta topología. De esta forma, a toda medida $\mu : \sigma(X) \to [0, \infty]$ se le llama *medida de Borel en X*.
En los conjuntos X e Y nos limitaremos a medidas de Borel, inducidas de la topología usual de $\mathbb{R}^n$, con $n = 1$ para el espacio Y. Además, veremos que suponer que Y sea un cerrado en dicha topología de $\mathbb{R}$ simplifica el análisis.
Recordemos que la meta es encontrar una función objetivo $f : X \to \mathbb{R}$ tal que, para cada par $(x, y) \in X \times Y$, el valor $f(x)$ sea una buena aproximación de y en x. Vamos a ver con qué criterio hablamos de buenas o malas aproximaciones, para lo que necesitamos empezar definiendo correctamente lo que entendemos por función de pérdida. Recordemos que en el apéndice B se recogen los resultados necesarios acerca de medidas producto.

Definición 3.2.1. Sean $d \in \mathbb{N}$, $(X, \mathcal{A})$ un espacio de medida, restringido del espacio de Borel $(\mathbb{R}^d, \sigma)$, e $Y \subset \mathbb{R}$ un subespacio cerrado, también con la medida de Borel restringida. Entonces, una función $L : X \times Y \times \mathbb{R} \to [0, \infty)$ es una **función de pérdida**, o simplemente **pérdida**, si es medible.

Una buena aproximación no viene dada por la coincidencia de los valores x e y a través de f, sino que nos interesa *promediar* dicha pérdida de cara a futuras predicciones. Esto motiva la necesidad de la *función riesgo*.

Definición 3.2.2. Sea $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una función pérdida y P una probabilidad en $X \times Y$. Entonces, para una función medible $f : X \to \mathbb{R}$, el *L-riesgo* o *L-función de riesgo* se define como la integral

$$\mathcal{R}_{L,P}(f) := \int_{X \times Y} L(x, y, f(x)) dP(x, y) = \int_X \int_Y L(x, y, f(x)) dP(x|y) dP_X(x). \tag{3.2.1}$$

Notemos que la función L de la anterior definición 3.2.2 es una función positiva, con lo que la integral (3.2.1) siempre existe, sea finita o infinita. Además, en adelante consideraremos que nuestro espacio Y de datos de salida sea un espacio cerrado en $\mathbb{R}$. Esto implica que Y es un espacio métrico completo y con base numerable, lo que justifica la exitencia de una probabilidad condicionada *regular*, que da sentido a la integral interna de la segunda expresión de (3.2.1).
Es inevitable la breve incursión en la teoría de la probabilidad que sigue. No podremos dar todos los detalles, porque el rigor requerido excede ampliamente el alcance de este trabajo, pero se puede encontrar una exposición completa en [8].

Consideremos $(S, \mathcal{D})$ y $(T, \mathcal{B})$ dos espacios de medida. Sea P una probabilidad definida en el producto $S \times T$, con σ-álgebra producto $\mathcal{A} = \mathcal{D} \times \mathcal{B}$. Sean X e Y las proyecciones usuales de $S \times T$ en S y T, respectivamente. Sea $\mathcal{C} = \{X^{-1}(D) : D \in \mathcal{D}\}$, esto es, sea $\mathcal{C}$ la σ-álgebra más pequeña en que X es medible. Notemos que $X^{-1}(D) = D \times T$ para todo $D \subset S$. Llamamos a $\mu = P \circ X^{-1}$, definida en $\mathcal{D}$, la *distribución marginal* de P en S, denotada por P_x. Su definición se formaliza con las siguientes condiciones:

1. Para todo $x \in S$, P_x es probabilidad en $(T, \mathcal{B})$.

2. Para todo $B \in \mathcal{B}$, $x \mapsto P_x(B)$ es $\mathcal{D}$-medible como aplicación de S en $\mathbb{R}$.

3. Para cada $D \in \mathcal{D}$ y $B \in \mathcal{B}$, $P(D \times B) = \int_D P_x(B) d\mu(x)$.

De esta definición se sigue inmediatamente, a modo de ejemplo, que si P es una medida producto, definida sobre los elementos que engendran nuestro σ-álgebra producto como $P(A \times B) = (\mu \times \nu)(A \times B) = \mu(A) \cdot \nu(B)$, entonces $P_x = \nu$.
Damos a continuación el lema que justificará la existencia de la probabilidad condicionada involucrada en nuestro L-riesgo, así como la forma de escindir dicha integral a través de probabilidades condicionadas y marginales.

Lema 3.2.3. *Sea (X, σ) un espacio de medida, y sea Y un espacio métrico completo, con base numerable y dotado de una σ-álgebra de Borel generada por él mismo, $\sigma(Y)$. Sea P una probabilidad en $\sigma \otimes \sigma(Y)$. Entonces, existe una aplicación $P(\cdot|\cdot) : \sigma(Y) \times X \to [0, 1]$ tal que*

i) $P(\cdot|x)$ es una probabilidad en $\sigma(Y)$ para todo $x \in X$.

ii) $x \mapsto P(B|x)$ es medible para todo $B \in \sigma(Y)$.

iii) Para todos $A \in \sigma$, $B \in \sigma(Y)$, tenemos

$$P(A \times B) = \int_A P(B|x) dP_X(x). \tag{3.2.2}$$

*A dicha aplicación $P(\cdot|\cdot)$ se le llama **probabilidad condicionada regular** de P.*

Fijémonos, por la definición dada de distribución marginal, en que este lema nos asegura que, dado un espacio de medida cualquiera X, y dado un subespacio Y cerrado de $\mathbb{R}$, luego métrico, completo y con base numerable en su topología de subespacio, entonces existe la probabilidad condicionada regular $P(\cdot|x)$ que está en la integral interna de la definición de L-riesgo.
Por lo tanto, tiene sentido la definición de L-riesgo que hemos aportado por medio de la integral (3.2.1). Esto es así por la expresión (3.2.2), y también por la aplicación del teorema de Tonelli-Fubini. Dicho teorema, enunciado en el resultado 1.5.1, es aplicable por la medibilidad supuesta del riesgo L, y por el hecho de que las medidas X e Y son σ-finitas, pues los tomamos como subconjuntos de $\mathbb{R}^n$, para algún $n \in \mathbb{N}$ (no necesariamente el mismo para X e Y) con medidas probabilísticas.
Sea ahora $D = \{(x_1, y_1), \ldots, (x_n, y_n)\} \in (X \times Y)^n$ una muestra finita de pares de datos de entrada y de salida. Definimos la medida que nos permitirá formalizar una función de riesgo a partir de D.

Definición 3.2.4. Sea (X, σ) un espacio de medida, y sea $x \in X$. La *medida de Dirac* δ_x vendrá definida por $\delta_x(A) := 1$ si y solo si $x \in A$, y $\delta_x(A) = 0$ en caso contrario.

De esta forma, estamos en condiciones de definir la *medida empírica* sobre tales muestras finitas, y con ella su función empírica de L-riesgo.

Definición 3.2.5. Sea $n \in \mathbb{N}$. Dada $D = \{(x_1, y_1), \ldots, (x_n, y_n)\} \in (X \times Y)^n$, definimos la *medida empírica* de este conjunto de datos como

$$D := \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i, y_i)}, \tag{3.2.3}$$

donde $\delta_{(x_i,y_i)}$ denota a la medida de Dirac en $(x_i, y_i) \in X \times Y$.
Dada ahora una función $f : X \to \mathbb{R}$ medible con respecto a esta medida D, definimos la *función L-riesgo empírico* como

$$\mathcal{R}_{L,D}(f) = \frac{1}{n} \sum_{i=1}^{n} L(x_i, y_i, f(x_i)). \tag{3.2.4}$$

La motivación teórica del uso de estos L-riesgos empíricos estriba en que $\mathcal{R}_{L,D}(f)$ se aproximará a $\mathcal{R}_{L,P}(f)$ para tamaños progresivamente mayores de la muestra finita D. De esta manera, podemos ver $\mathcal{R}_{L,D}(f)$ como una aproximación práctica del promedio de la función de pérdida sobre las observaciones en D. Justifiquemos esta afirmación: tomamos $(x_1, y_1), \ldots, (x_n, y_n)$ como n variables aleatorias independientes, y distribuidas con la misma probabilidad P en $X \times Y$; dadas la medibilidad en la σ-álgebra producto en $X \times Y \times \mathbb{R}$ y en la σ-álgebra de Borel en $\mathbb{R}$ de la función riesgo $L : X \times Y \times \mathbb{R} \to \mathbb{R}$, y la medibilidad de las funciones $f : X \to \mathbb{R}$ que manejaremos, entonces, la aplicación $L(\cdot, \cdot, (f \circ \pi_X)(\cdot))$, con $\pi_X : X \times Y \to X$ la proyección en X, es medible, en las σ-álgebras producto en $X \times Y$, y de Borel en $\mathbb{R}$; y, por último, todo ello nos da n variables aleatorias $L(x_i, y_i, f(x_i, y_i))$ independientes e igualmente distribuidas tales que $E\left(L(x_i, y_i, f(x_i, y_i))\right) = \mathcal{R}_{L,P}(f) < \infty$. Entonces, se cumplen las hipótesis de la ley fuerte de los grandes números, con lo que $\mathcal{R}_{L,D}(f) \to \mathcal{R}_{L,P}(f)$ para $n \to \infty$.
Además de formalizar la teoría con la que escribir estas pérdidas y estos promedios, no podemos perder de vista el objetivo de esta construcción: buscamos el riesgo más bajo de entre todas las funciones f. A continuación, formalizamos esto por medio de la *función de decisión de Bayes*.

Definición 3.2.6. Sea $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una función de pérdida y P una probabilidad en $X \times Y$. Entonces, el L-riesgo inferior

$$\mathcal{R}^*_{L,P} := \inf\{\mathcal{R}_{L,P}(f) | f : X \to \mathbb{R} \;\; \text{medible}\}, \tag{3.2.5}$$

es llamado **riesgo de Bayes** con respecto a P y L. Además, si existe, a una función medible $f^*_{L,P} : X \to \mathbb{R}$ con $\mathcal{R}_{L,P}(f^*_{L,P}) = \mathcal{R}_{L,P}$ le llamamos **función de decisión de Bayes**.

Es claro que uno de los puntos clave en la resolución de problemas de aprendizaje estadísito es la elección de una función de pérdida adecuada, a partir de la cual el L-riesgo y todas las consideraciones teóricas podrán ser más o menos complejas.
A modo de ejemplos, presentamos en primer lugar el caso estándar de clasificación binaria, utilizada en todas aquellas situaciones en que interese el etiquetar los elementos de un determinado conjunto, X, de dos formas posibles, por medio del conjunto de datos de salida $Y = \{1, -1\}$. También contemplamos el método ampliado de clasificación con un peso α, así como un enfoque generalista del tratamiento de la regresión por mínimos cuadrados.

Ejemplo 3.2.7 (Clasificación binaria estándar)**.** Sea pues el conjunto output $Y := \{1, -1\}$, X un conjunto arbitrario no vacío de elementos que queremos clasificar en dos tipos, y P una distribución de probabilidad en $X \times Y$. Reescribiendo, de una forma más informal, el propósito de este problema de aprendizaje, el objetivo es predecir la etiqueta $y \in Y$ del par (x, y), si $x \in X$ es observado.
La función de pérdida más común en esta situación es la llamada *pérdida de clasificación* L_{cl} : $Y \times \mathbb{R} \to [0, \infty)$ donde, notemos, no hay dependencia explícita con los $x \in X$, ya que

$$L_{\text{cl}}(y, t) := 1_{(-\infty,0]}(y \cdot \text{signo}(t)), \quad y \in Y, t \in \mathbb{R}, \tag{3.2.6}$$

donde es habitual usar la convención signo(0) := 1. Básicamente, lo que hace esta función de pérdida es devolver un valor 0 (i.e., contribución nula al L_{cl}-riesgo) si $y \in Y$ y $t \in \mathbb{R}$ coinciden en signo, y 1 si no lo hacen. Sin restringir aún el espacio de función sobre el que buscamos nuestra solución, sea $f : X \to \mathbb{R}$ una función medible (recodemos que, salvo si se dice lo contrario, en $\mathbb{R}$ siempre consideramos su topología estándar y su σ-álgebra de Borel). Veamos que el significado informal de la función pérdida es coherente con la forma que adquiere la función L_{cl}-riesgo:

$$\begin{aligned}
&\mathcal{R}_{L_{\text{cl}},P}(f) = \int_X \int_Y L_{\text{cl}}(y, f(x)) dP(y|x) dP_X(x) = \int_X \left[\int_Y 1_{(-\infty,0]}(y \cdot \text{signo}(f(x))) dP(y|x) \right] dP_X(x) \\
&= \int_X \left[\theta(x) 1_{(-\infty,0]}(f(x)) + (1-\theta(x)) 1_{[0,\infty)}(f(x)) \right] dP_X(x) \\
&= P(\{(x,y) \in X \times Y : \text{signo}(f(x)) \neq y\}), \qquad (3.2.7)
\end{aligned}$$

donde se ha denotado $P(y = 1|x) = \theta(x)$. Esta pérdida L_{cl} da el mismo *peso* a los dos tipos de errores que en una clasificación binaria pueden ocurrir: $(y = 1, f(x) < 0)$ y $(y = -1, f(x) > 0)$. No obstante, imagínese la situación de que esta clasificación binaria se diera para la detección de dos tipos de errores informáticos, uno más dañino que el otro. Es claro que, en ese caso, lo mejor sería dar un peso mayor al fallo más pernicioso. Esta situación se estudia en el ejemplo siguiente.

Para el L-riesgo que ahora nos ocupa, el menor valor del L-riesgo viene dado por

$$\mathcal{R}^*_{L_{\text{cl}},P} = \int_X \min\{\theta, 1-\theta\} dP_X, \qquad (3.2.8)$$

luego la función $f : X \to Y$ es una función de decisión de Bayes si y solo si $(2\theta(x)-1) \cdot \text{signo}(f(x)) \geq 0$ para P_X-casi todo $x \in X$. Esto se ve fácilmente si escribimos, para un $x \in X$ fijo y $\theta(x) = \theta \in [0,1]$:

$$\min\{\theta, 1-\theta\} = \begin{cases} 1-\theta & si \quad \theta \geq 1/2, \\ \\ \theta & si \quad \theta < 1/2. \end{cases} \qquad (3.2.9)$$

Como el integrando del L-riesgo será de la forma θ si $f(x) < 0$, y de la forma $1-\theta$ si $f(x) \geq 0$, llegamos a la función de decisión de Bayes dada.

Hemos visto así un ejemplo clásico e ilustrativo de la formalización probabilística y funcional de una pérdida y un riesgo en un método de Aprendizaje Estadístico.

Ejemplo 3.2.8 (Clasificación binaria con peso)**.** Sea $Y := \{-1, 1\}$ el conjunto de salida, y sea un parámetro $\alpha \in (0,1)$, al que llamaremos peso. La *pérdida de clasificación con peso* α, $L_{\alpha,\text{cl}} : Y \times \mathbb{R} \to [0,\infty)$ (notemos que seguimos sin tener dependencia explícita de L con los $x \in X$) está definida como sigue

$$\mathrm{L}_{\alpha,\text{cl}} = \begin{cases} 1-\alpha & si \qquad y = 1 \ \text{ y } \ t < 0, \\ \\ \alpha & si \qquad y = -1 \ \text{ y } \ t \geq 0, \\ \\ 0 & si \quad (y,t) \in (\{1\} \times [0,\infty)) \cup (\{-1\} \times (-\infty, 0)), \end{cases} \qquad (3.2.10)$$

para todo $y \in Y$, $t \in \mathbb{R}$. Notemos que $2 \cdot L_{1/2,\text{cl}} = L_{\text{cl}}$, es decir, la pérdida de clasifición es un caso particular de la pérdida con peso $\alpha = 1/2$. Con la justificación dada para cambiar los órdenes de integración en las funciones de L-riesgo, estamos en condiciones de escribir la función L-riesgo de este modelo de la siguiente forma, teniendo en cuenta, además, que en la forma explícita de la función L tampoco aparece la $y \in Y$. Basta con separar la integral en los casos $y = 1$ y $y = -1$ como sigue

$$\begin{aligned}
&\mathcal{R}_{L_{\alpha,\text{cl}}}(f) = \int_{\{y=1\}} \int_{X \times Y} L(y, f(x)) dP(x,y) + \int_{\{y=-1\}} \int_{X \times Y} L(y, f(x)) dP(x,y) \\
&= (1-\alpha) \int_{f<0} \eta dP_X + \alpha \int_{f \geq 0} (1-\eta) dP_X, \qquad (3.2.11)
\end{aligned}$$

donde se ha denotado $\eta(x) = P(y = 1|x)$, para todo $x \in X$. Repitiendo un razonamiento idéntico al de la función de pérdida de clasificación binaria ordinaria, llegamos a que f es función de decisión de Bayes si y solo si $(\eta(x) - \alpha) \cdot \text{signo}(f(x)) \geq 0$. Entonces, el L-riesgo de Bayes de este planteamiento es:

$$\mathcal{R}^*_{L_{\text{cl}},P} = \int_X \min\{(1-\alpha)\eta, \alpha(1-\eta)\} dP_X. \tag{3.2.12}$$

De la misma forma que en el primer ejemplo, este tipo de pérdidas y L-riesgos permiten predecir valores y del conjunto $Y = \{-1, 1\}$. En el siguiente ejemplo, generalizamos a conjuntos de datos de salida reales generales.

Ejemplo 3.2.9 (Regresión por mínimos cuadrados)**.** Nos proponemos diseñar una pérdida L y un L-riesgo con conjunto de salida $Y := \mathbb{R}$, para duplas $(x, y) \in X \times Y$ distribuidas con una probabilidad P en $X \times Y$, siempre y cuando x sea un dato observado. La forma más sencilla de plantearlo es mediante la *pérdida de mínimos cuadrados* $L_{MC} : Y \times \mathbb{R} \to [0, \infty)$, definida como:

$$L_{MC}(y, t) := (y - t)^2, \quad y \in Y, t \in \mathbb{R}. \tag{3.2.13}$$

Es decir, *penalizamos* las diferencias entre los valores y y t cuadráticamente. Entonces, dada $f : X \to \mathbb{R}$ una función medible, el L_{MC}-riesgo es

$$\mathcal{R}_{L_{MC},P} \int_X \int_Y (y - f(x))^2 \, dP(y|x) dP_X(x). \tag{3.2.14}$$

Minimizando (3.2.14) con respecto a $f(x)$, simplemente mencionamos que f es una función de decisión de Bayes si y solo si $f(x)$ es la esperanza de la variable aleatoria Z dada por los valores $y \in Y$. No entraremos en el detalle de esto, pues requerimos de nociones de teoría de la probabilidad que se salen del tema de este trabajo.

A continuación, nos centramos en un somero estudio de distintas propiedades importantes, tales como la convexidad o la medibilidad, de las funciones pérdida.

Definición 3.2.10. Diremos que una pérdida $L : X \times Y \times \mathbb{R} \to [0, \infty)$ es una **pérdida de Nemitski** si y solo si existen una función medible $b : X \times Y \to [0, \infty)$ y una función creciente $h : [0, \infty) \to [0, \infty)$ tales que

$$L(x, y, t) \leq b(x, y) + h(|t|), \quad (x, y, t) \in X \times Y \times \mathbb{R}. \tag{3.2.15}$$

Además, diremos que L es un **pérdida Nemitski de orden** $p \in (0, \infty)$ si existe una constante $c > 0$ tal que

$$L(x, y, t) \leq b(x, y) + c|t|^p, \quad (x, y, t) \in X \times Y \times \mathbb{R}. \tag{3.2.16}$$

Por último, si P es una probabilidad en $X \times Y$ con $b \in L^1(P)$ (i.e., b integrable con respecto a la probabilidad P), decimos que L es una **pérdida Nemitski P-integrable**.

Como primera consecuencia de la definición 3.2.10, dada P la probabilidad con que están distribuidos los datos en $X \times Y$, y suponiendo que L es P-integrable Nemitski, tenemos que $\mathcal{R}_{L,P}(f) < \infty$ para todo $f \in L^\infty(P_X)$. En particular, $\mathcal{R}_{L,P}(0) < \infty$, es decir, el L-riesgo de la función constante igual a 0 es finito. Más adelante veremos que esto asegura que el inferior de los riesgos sobre L y P, $\mathcal{R}^*_{L,P}$, es finito en cualquier función f del espacio de funciones en que buscamos dicho ínfimo.
Pasamos a definir lo que entenderemos por convexidad de las funciones pérdida.

Definición 3.2.11. Una pérdida $L : X \times Y \times \mathbb{R} \to [0, \infty)$ es *(estrictamente) convexa* si y solo si $L(x, y, \cdot) : \mathbb{R} \to [0, \infty)$ es (estrictamente) convexa para todos $x \in X$ e $y \in Y$.

Relacionamos ahora la convexidad de las funciones pérdida con la de las funciones riesgo. Denotamos por $\mathcal{L}^0(X)$ al conjunto de funciones medibles que van de un conjunto X no vacío, a $\mathbb{R}$. Equivalentemente, cuando queramos remarcar la distribución de probabilidad P en X que estemos considerando, denotaremos a este conjunto como $\mathcal{L}^0(P)$.

Lema 3.2.12. *Sea $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una pérdida (estrictamente) convexa y sea P una distribución de probabilidad en $X \times Y$. Entonces, $\mathcal{R}_{L,P} : \mathcal{L}^0(X) \to [0, \infty]$ es (estrictamente) convexa.*

Demostración. Es consecuencia directa de la linealidad y la monotonía de las integrales respecto a la distribución P en $X \times Y$, ya que, si $f, g \in \mathcal{L}^0(X)$, y si $t \in (0, 1)$, se tiene, por la propia definición de L-riesgo:

$$\begin{aligned}
&\mathcal{R}_{L,P}(tf + (1-t)g) = \int_{X\times Y} L(x, y, tf(x) + (1-t)g(x))dP(x, y) \\
&\leq \int_{X\times Y} [tL(x, y, f(x)) + (1-t)L(x, y, g(x))]\, dP(x, y) = t \int_{X\times Y} L(x, y, f(x))dP(x, y) \\
&+ (1-t) \int_{X\times Y} L(x, y, g(x))dP(x, y) = t\mathcal{R}_{L,P}(f) + (1-t)\mathcal{R}_{L,P}(g). \qquad (3.2.17)
\end{aligned}$$

□

A continuación, tratamos la continuidad de los L-riesgos a partir de la continuidad de las funciones pérdida, así como algún otro resultado que nos permite hablar de propiedades interesantes de dichos riesgos. Para ello, empezamos definiendo lo que entendemos por pérdida continua.

Definición 3.2.13. Una pérdida $L : X \times Y \times \mathbb{R} \to [0, \infty)$ es *continua* si $L(x, y, .) : \mathbb{R} \to [0, \infty)$ es continua para todos $x \in X$, $y \in Y$.

Consideremos una función de pérdida continua $L : X \times Y \times \mathbb{R} \to [0, \infty)$, así como una sucesión $\{f_n\}$ de funciones medibles de X en $\mathbb{R}$ (lo que, con el tiempo, iremos viendo como elementos del RKHS correspondiente) convergente puntualmente a una función $f : X \to \mathbb{R}$ (necesariamente medible). Entonces, por la continuidad secuencial de L, tenemos la convergencia en $\mathbb{R}$ dada por $L(x, t, f_n(x)) \to L(x, y, f(x))$ para $n \to \infty$, y para todo $(x, y) \in X \times Y$. Sin embargo, no tenemos ninguna garantía de que esto implique la convergencia de sus correspondientes integrales, es decir, de los L-riesgos $\mathcal{R}_{L,P}(f_n) \to \mathcal{R}_{L,P}(f)$. Enunciamos un resultado más débil.

Lema 3.2.14. *Sea $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una función de pérdida continua, P una probabilidad en $X \times Y$, y $\{f_n\}$ una sucesión en $\mathcal{L}^0(P_X)$, con P_X la probabilidad marginal en X, tal que converge en dicha probabilidad a un elemento $f \in \mathcal{L}^0(P_X)$. Entonces,*

$$\mathcal{R}_{L,P}(f) \leq \liminf_{n\to\infty} \mathcal{R}_{L,P}(f_n). \qquad (3.2.18)$$

Demostración. Notemos que estamos tratando elementos $f_n : X \to \mathbb{R}$ medibles. Por definición de límite inferior, debe existir una subsucesión $\{f_{n_k}\}_{k=1}^{\infty}$ tal que

$$\liminf_{n\to\infty} \mathcal{R}_{L,P}(f_n) = \lim_{k\to\infty} \mathcal{R}_{L,P}(f_{n_k}). \qquad (3.2.19)$$

La convergencia en probabilidad a f se mantiene para la subsucesión $\{f_{n_k}\}_{k=1}^{\infty}$, lo que implica que existe una subsucesión de ésta, $\{f_{n_{k_m}}\}_{m=1}^{\infty}$, que converge P_X-casi seguro en X a f. Simplificaremos la subsucesión resultante de este doble paso como $\{f_{n_k}\}_{k=1}^{\infty}$, sin olvidar que nos quedamos con los índices n_k tales que la convergencia casi seguro anterior se cumple.
Ahora bien, como L es una función de pérdida continua, entonces es continua secuencialmente, con lo que $L(x, y, f_{n_k}(x)) \to L(x, y, f(x))$, para casi todo $(x, y) \in X \times Y$ (respecto a P). Utilizamos el

lema de Fatou, por el que, dada la convergencia secuencial anterior en términos de la pérdida L, se tiene que

$$\begin{aligned}
\mathcal{R}_{L,P}(f) &:= \int_{X\times Y} \lim_{k\to\infty} L(x,y,f_{n_k}(x))dP(x,y) \\
&= \int_{X\times Y} \liminf_{k\to\infty} L(x,y,f_{n_k}(x))dP(x,y) \leq \liminf_{k\to\infty} \int_{X\times Y} L(x,y,f_{n_k}(x))dP(x,y) \\
&= \liminf_{k\to\infty} \mathcal{R}_{L,P}(f_{n_k}) = \liminf_{n\to\infty} \mathcal{R}_{L,P}(f_n),
\end{aligned}$$

con lo que tenemos la desigualdad buscada, que es

$$\mathcal{R}_{L,P}(f) \leq \liminf_{n\to\infty} \mathcal{R}_{L,P}(f_n). \tag{3.2.20}$$

□

El enunciado del lema anterior 3.2.14 nos sugiere que si somos capaces de mayorar la sucesión $L(\cdot,\cdot,f_n(\cdot))$ por alguna función integrable en la medida dada, entonces, por el teorema de convergencia dominada, obtendríamos la convergencia de $\mathcal{R}_{L,P}(f_n)$ a $\mathcal{R}_{L,P}(f)$. Veamos que las pérdidas Nemitski desempeñan un papel fundamental para estas convergencias.

Lema 3.2.15 (Continuidad de L-riesgos)**.** *Sea P una distribución de probabilidad en $X\times Y$, y sea $L : X\times Y\times\mathbb{R}\to[0,\infty)$ una pérdida continua y P-integrable Nemitski. Entonces:*

1. *Sean $f_n : X\to\mathbb{R}$ medibles y acotadas para todo $n\in\mathbb{N}$, para las que existe una constante $B>0$ con $||f_n||_\infty \leq B$ para todo $n\in\mathbb{N}$. Si dicha sucesión $\{f_n\}$ converge P_X-casi seguro a una función $f : X\to\mathbb{R}$, entonces*
$$\lim_{n\to\infty} \mathcal{R}_{L,P}(f_n) = \mathcal{R}_{L,P}(f). \tag{3.2.21}$$

2. *La aplicación $\mathcal{R}_{L,P} : L^\infty(P_X)\to[0,\infty)$ está bien definida y es continua.*

3. *Si L es una pérdida de Nemitski de orden $p\in[1,\infty)$, entonces la aplicación $\mathcal{R}_{L,P} : L^p(P_X)\to[0,\infty)$ está bien definida y es continua.*

Demostración. 1. Como $\{f_n\}$ converge P_X-casi seguro a f, y f_n es medible para todo $n\in\mathbb{N}$, es claro que f también será medible.

Por otra parte, f también estará acotada por B. Esto se deduce directamente de la convergencia P_X-casi seguro de $\{f_n\}$ a f. Para probarlo, sea $X'\subset X$ tal que $P_X(X\setminus X')=0$ y tal que $\lim_{n\to\infty} f_n(x) = f(x)$ para todo $x\in X'$, y consideremos un $\varepsilon>0$ arbitrario. Entonces, existe un $N_x\in\mathbb{N}$ tal que

$$|f(x)| \leq |f_{N_x}(x) - f(x)| + |f_{N_x}(x)| \leq \varepsilon + B, \tag{3.2.22}$$

donde se ha usado la acotación de las f_n y la desigualdad triangular. Como esta desigualdad se cumple para todo $\varepsilon>0$, y como x está en un conjunto X' cuyo complementario tiene probabilidad nula, entonces $||f||_\infty \leq B$.

Ahora bien, como L es una aplicación continua, podemos extender la convergencia P_X-casi seguro de $\{f_n\}$ a f a la siguiente convergencia, ahora P-casi seguro en $X\times Y$:

$$\lim_{n\to\infty} |L(x,y,f_n(x)) - L(x,y,f(x))| = 0. \tag{3.2.23}$$

Dado que L es una pérdida de Nemitski, por la desigualdad triangular sobre la norma ordinaria en $\mathbb{R}$ y por la condición (3.2.15), tenemos que

$$|L(x,y,f_n(x)) - L(x,y,f(x))| \leq 2b(x,y) + h(|f_n(x)|) + h(|f(x)|) \leq$$

$$\leq 2b(x,y) + 2h(B), \quad (x,y) \in X \times Y, n \in \mathbb{N}, \tag{3.2.24}$$

donde se ha usado la monotonía de la función h, y la acotación de las normas del infinito f_n y f por la constante $B > 0$. Como $2b(x,y) + 2h(B)$, también por la definición de pérdida de Nemitski P-integrable, es P-integrable (nótese que $2h(B)$ es una función constante, y que estamos integrando sobre una medida que es una probabilidad), entonces, estamos en condiciones de aplicar el teorema de la convergencia dominada para confirmar la convergencia de $\mathcal{R}_{L,P}(f_n)$ a $\mathcal{R}_{L,P}(f)$.

2. La buena definición se sigue inmediatamente de la definición de pérdida P-integrable de Nemitski y de un argumento muy similar al del punto anterior. Como $b \in L^1(P)$, se tiene que, para un $f \in L^\infty(P_X)$ cualquiera y usando la monotonía de la integral,

$$\mathcal{R}_{L,P}(f) = \int_{X\times Y} L(x,y,f(x))dP(x,y) \leq \int_{X\times Y} b(x,y)dP(x,y)+$$

$$+\int_{X\times Y} h(|f(x)|) \leq \int_{X\times Y} b(x,y)dP(x,y) + P(X\times Y)h(M), \tag{3.2.25}$$

donde se ha usado la monotonía de h, y la acotación de f por una constante $M > 0$ fruto de que $f \in L^\infty(P_X)$. Esto demuestra la buena definición.
La continuidad es una consecuencia directa del primer punto, pues en este probábamos la continuidad secuencial de $\mathcal{R}_{L,P}(\cdot)$ bajo convergencia P_X-casi seguro de elementos $f_n : X \to \mathbb{R}$, y la convergencia uniforme, i.e., en $L^\infty(P_X)$, implica dicha convergencia casi seguro.

3. Ahora, el enunciado restringe la pérdida L de Nemitski a una pérdida de Nemitski de orden $p \in [1,\infty)$, lo cual recordemos que implica, por (3.2.15), que

$$L(x,y,f(x)) \leq b(x,y) + c|t|^p, \quad (x,y) \in X\times Y, \tag{3.2.26}$$

de lo que se sigue

$$\mathcal{R}_{L,P}(f) = \int_{X\times Y} L(x,y,f(x))dP(x,y) \leq \int_{X\times Y} b(x,y)dP(x,y) + \int_{X\times Y} c|f(x)|^p dP(x,y), \tag{3.2.27}$$

donde $c > 0$, y donde el primer término es finito porque $b \in L^1(P)$. Con respecto al segundo término, escindiendo la integral como hicimos en la definición de L-riesgo, y usando también la linealidad de estas integrales, se tiene

$$\int_{X\times Y} c|f(x)|^p dP(x,y) = c\int_Y \left(\int_X |f(x)|^p dP_X(x)\right) dP(y|x). \tag{3.2.28}$$

Esta última expresión es finita porque $f \in L^p(P_X)$, y de ello se deriva una constante integrada sobre una medida probabilísitca, lo que siempre es integrable. Esto prueba la buena definición de la aplicación propuesta.
En cuanto a la continuidad, usaremos el resultado bien conocido que afirma que la convergencia en $L^p(P_X)$ (o lo que en teoría de la probabilidad se suele llamar *convergencia en media de orden* p) implica la convergencia en probabilidad P_X. Entonces, podemos aplicar el lema 3.2.14 sobre una sucesión $\{f_n\} \subset \mathcal{L}^0(P_X)$, convergente en probabilidad P_X a $f : X \to \mathbb{R}$, de tal forma que

$$\mathcal{R}_{L,P}(f) \leq \liminf_{n\to\infty} \mathcal{R}_{L,P}(f_n), \tag{3.2.29}$$

sobre nuestra pérdida L continua, y sobre nuestra probabilidad P en $X \times Y$. Pues bien, si aplicamos este mismo razonamiento sobre la función pérdida definida en $X \times Y \times \mathbb{R}$ como:

$$\tilde{L}(x,y,t) := b(x,y) + c|t|^p - L(x,y,t), \tag{3.2.30}$$

que también es continua (por ser suma de funciones continuas de $\mathbb{R}$ en $[0,\infty)$, para cada $(x,y)\in X\times Y$ fijo), entonces, aplicando (3.2.30), y denotando por $\mathcal{R}_{\tilde{L},P}$ al $\tilde{L}$-riesgo:

$$\mathcal{R}_{\tilde{L},P}(f) \leq \liminf_{n\to\infty} \mathcal{R}_{\tilde{L},P}(f_n), \tag{3.2.31}$$

donde

$$\begin{aligned}\mathcal{R}_{\tilde{L},P}(f) &= \int_{X\times Y} (b(x,y) + c|f(x)|^p - L(x,y,f(x)))\, dP(x,y) = \\ &= ||b||_{L^1(P)} + c||f||_p^p - \mathcal{R}_{L,P}(f),\end{aligned} \tag{3.2.32}$$

y también

$$\liminf_{n\to\infty} \mathcal{R}_{\tilde{L},P}(f_n) = \liminf_{n\to\infty} \left(||b||_{L^1(P)} + c||f_n||_p^p - \mathcal{R}_{L,P}(f_n)\right). \tag{3.2.33}$$

Por lo tanto, por continuidad de $||\cdot||_p^p$, por linealidad del límite, y por (3.2.31), llegamos a

$$-\mathcal{R}_{L,P}(f) \leq -\liminf_{n\to\infty} \mathcal{R}_{L,P}(f_n), \tag{3.2.34}$$

o equivalentemente

$$\mathcal{R}_{L,P}(f) \geq \limsup_{n\to\infty} \mathcal{R}_{L,P}(f_n). \tag{3.2.35}$$

Así, por (3.2.29) y por (3.2.35), se tiene:

$$\limsup_{n\to\infty} \mathcal{R}_{L,P}(f_n) \leq \mathcal{R}_{L,P}(f) \leq \liminf_{n\to\infty} \mathcal{R}_{L,P}(f_n), \tag{3.2.36}$$

lo que demuestra la existencia del límite, y que este límite es, precisamente, $\mathcal{R}_{L,P}(f)$. Esto prueba la continuidad de $\mathcal{R}_{L,P} : L^p(P_X) \to [0,\infty)$.

□

A continuación, entramos en una faceta un tanto más cuantitativa de la noción de continuidad de estas funciones pérdida, dando una forma analítica explícita a una nueva idea que nos ayudará en nuestro desarrollo: las *pérdidas localmente Lipschitz continuas*.

Definición 3.2.16. Sea $L : X\times Y\times\mathbb{R} \to [0,\infty)$ una función pérdida. Diremos que L es **localmente Lipschitz continua** si y solo si para todo $a\geq 0$ existe una constante $c_a \geq 0$ tal que

$$\sup_{x\in X, y\in Y} |L(x,y,t) - L(x,y,t')| \leq c_a |t-t'|, \quad t,t'\in[-a,a]. \tag{3.2.37}$$

Denotaremos por $|L|_{a,1}$ a la constante c_a más pequeña que satisfaga la desigualdad anterior para todos $t,t'\in[-a,a]$ (notemos que tal $|L|_{a,1}$ debe existir, dado que el conjunto de tales constante c_a está acotado inferiormente por 0 y es no vacío, para lo que restaría aplicar el axioma de completitud). Finalmente, si tenemos $|L|_1 := \sup_{a\geq 0}|L|_{a,1} < \infty$, decimos directamente que L es **Lipschitz continua**.

De cara a acotar funciones minimizadores que sean soluciones generales SVM, notemos que si tomamos un $t\in\mathbb{R}$ cualquiera, cogiendo $a = |t| \geq 0$, $c_a = L_{|t|,1}$, y $t' = 0$ en la definición 3.2.16 anterior, tenemos que

$$\begin{aligned}&\sup_{x\in X, y\in Y} |L(x,y,t) - L(x,y,t')| \leq c_a|t-t'| \implies |L(x,y,t) - L(x,y,0)| \leq L_{|t|,1}|t| \implies \\ &\implies L(x,y,t) \leq L(x,y,0) + L_{|t|,1}|t|, \quad (x,y)\in X\times Y, t\in\mathbb{R}.\end{aligned} \tag{3.2.38}$$

Esta desigualdad nos permite enunciar la siguiente proposición.

Proposición 3.2.17. *Sea $L : X\times Y\times\mathbb{R}\to[0,\infty)$ una pérdida localmente Lipschitz continua, y P la distribución dada sobre $X\times Y$. Supongamos también que $\mathcal{R}_{L,P}(0) < \infty$. Entonces, L es una pérdida P-integrable Nemitski. En particular, lo será con orden $p = 1$.*

Demostración. Si L es localmente Lipschitz continua, por el razonamiento anterior tenemos que

$$L(x, y, t) \leq L(x, y, 0) + L_{|t|,1}|t|, \quad (x, y) \in X \times Y, t \in \mathbb{R}. \tag{3.2.39}$$

Así, tomando en la definición de pérdida de Nemitski P-integrable la función $b(x, y) = L(x, y, 0)$ y la constante $c = L_{|t|,1}$ para el término $c|t|^p$, tenemos la acotación deseada. Como, por hipótesis, $\mathcal{R}_{L,P}(0) < \infty$, tenemos que esta función b, con la integración dada sobre la medida P en $X \times Y$, está en $L^1(P)$. Entonces, L es P-integrable Nemitski y, en particular, de orden $p = 1$. □

Una vez obtenidas las propiedades de continuidad más importantes de funciones pérdida y riesgo, pasamos a un nuevo apartado con mayor interés práctico, recurriendo de nuevo a nuestro RKHSs.

3.3. Existencia y unicidad de soluciones generales SVM

Fijamos nuestra mirada en el estudio de las propiedades estructurales de los SVMs. Fundamentalmente, estudiaremos la existencia y la unicidad de las soluciones de algunos problemas de SVM, así como una posible cota sobre dichas soluciones, y un importante teorema de representación en nuestros RKHSs subyacentes.
En la práctica, los métodos de SVM toman como función objetivo a una función $f_{D,\lambda}$, a partir de una muestra D finita en $(X \times Y)^n$ para algún $n \in \mathbb{N}$, y de un parámetro $\lambda > 0$ de control del término de penalización, dado por la norma definida en el espacio de funciones H. Es así que las funciones a minimizar toman, normalmente, la forma del llamado **riesgo empírico regularizado**, o **RER**, que es

$$\mathcal{R}^{\text{reg}}_{L,D,\lambda}(f) := \lambda||f||^2_H + \mathcal{R}_{L,D}(f), \qquad f \in H. \tag{3.3.1}$$

No obtante, los resultados que nos ocupan se extenderán, salvo que se haga notar lo contrario, a las funciones riesgo generales $\mathcal{R}_{L,P}$. Como ya dijimos, nos interesa la teoría que hay detrás de estos métodos, no su implementación. La ley fuerte de los grandes números ya se encarga de justificar la práctica con RERs. Sin embargo, a la hora de hablar de representación de soluciones vía la propiedad reproductora del núcleo de un RKHS, sí que interesa volver la mirada a estos riesgos empíricos.
Denotamos por H al espacio de funciones de X (el conjunto de datos de entrada) en el cuerpo $\mathbb{R}$. Nos limitamos a espacios H como RKHSs. Llegados a este punto, las preguntas naturales que hacer serían:

1. ¿Existen minimizadores? Y de ser así, ¿son únicos?
2. ¿Qué formas posibles hay de representar el minimizador $f_{P,\lambda}$ a partir de muestras finitas?

Denotaremos por $L : X \times Y \times \mathbb{R} \to [0, \infty)$ a las funciones pérdida, H al RKHS con núcleo medible en X en el que buscamos la función minimizadora, y P la distribución de probabilidad en $X \times Y$ con la que se distribuyen los datos de entrada y de salida. Dado $\lambda > 0$, a la función $f_{P,\lambda,H} \in H$ de X en $\mathbb{R}$ que satisface

$$\lambda||f_{P,\lambda,H}||^2_H + \mathcal{R}_{L,P}(f_{P,\lambda,H}) = \inf_{f \in H} \left(\lambda||f||^2_H + \mathcal{R}_{L,P}(f)\right), \tag{3.3.2}$$

le llamamos **solución general de la SVM** o **función de decisión general de la SVM**, a la cual denotaremos en adelante como $f_{P,\lambda}$ para abreviar, definida sobre H salvo que se haga notar lo contrario.
Observemos que, si $f_{P,\lambda}$ es una solución general de nuestra SVM, entonces

$$\lambda||f_{P,\lambda}||^2_H \leq \lambda||f_{P,\lambda}||^2_H + \mathcal{R}_{L,P}(f_{P,\lambda}) = \inf_{f \in H} \left(\lambda||f||^2_H + \mathcal{R}_{L,P}(f)\right) \leq \mathcal{R}_{L,P}(0), \tag{3.3.3}$$

de tal forma que tenemos una primera acotación

$$||f_{P,\lambda}||_H \leq \sqrt{\frac{\mathcal{R}_{L,P}(0)}{\lambda}}, \tag{3.3.4}$$

lo cual tiene una interesante consecuencia: fijado $\mathcal{R}_{L,P}(0)$, la acotación sobre el minimizador en norma de H queda controlada por el parámetro λ.
Entramos ahora a estudiar bajo qué condiciones existe una única $f_{P,\lambda}$. En el caso de funciones de pérdida convexas, la unicidad, asumiendo de partida su existencia, es inmediata.

Proposición 3.3.1 (Unicidad de soluciones SVM)**.** *Sean $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una función de pérdida convexa, H el RKHS definido por un núcleo medible en X, y P una probabilidad en $X \times Y$ con $\mathcal{R}_{L,P}(f) < \infty$ para algún $f \in H$. Entonces, para todo $\lambda > 0$, existe a lo más una solución $f_{P,\lambda}$ de la SVM propuesta.*

Demostración. Por reducción al absurdo, supongamos que $f \mapsto \lambda||f||_H^2 + \mathcal{R}_{L,P}(f)$ admite dos minimizadores f_1 y f_2 en H, , tales que $f_1 \neq f_2$. Utilizando la identidad del paralelogramo, es claro que

$$||\frac{1}{2}(f_1 + f_2)||_H^2 = \frac{1}{2}||f_1||_H^2 + \frac{1}{2}||f_2||_H^2 - ||\frac{1}{2}(f_1 - f_2)||_H^2 < \frac{1}{2}||f_1||_H^2 + \frac{1}{2}||f_2||_H^2. \tag{3.3.5}$$

Así, la convexidad de $f \mapsto \mathcal{R}_{L,P}(f)$ (derivada de la convexidad, por hipótesis, de la función pérdida) y la igualdad $\lambda||f_1||_H^2 + \mathcal{R}_{L,P}(f_1) = \lambda||f_2||_H^2 + \mathcal{R}_{L,P}(f_2)$ muestra que, si tomamos $\tilde{f} := \frac{1}{2}(f_1 + f_2)$, entonces

$$||\tilde{f}||_H^2 + \mathcal{R}_{L,P}(\tilde{f}) < ||f_1||_H^2 + \mathcal{R}_{L,P}(f_1), \tag{3.3.6}$$

es decir, que f_1 no puede ser un minimizador de $f \mapsto ||f||_H^2 + \mathcal{R}_{L,P}(f)$, lo que va en contra de nuestra hipótesis. Por lo tanto, de existir, existe un único minimizador como solución de nuestra SVM. □

Nuestro siguiente objetivo es el de ver que para funciones de pérdida convexas e integrables de tipo Nemitski, siempre existe una solución general del problema SVM. Antes de abordar el teorema, definimos lo que entenderemos por semicontinuidad inferior o superior, así como el epigrafo de una función de un conjunto X no vacío en $\mathbb{R}$.

Definición 3.3.2. Sea X un espacio topológico y sea $f : X \to \mathbb{R} \cup \{\infty\}$. Entonces f es inferiormente semicontinua (resp. superiormente semicontinua) si y solo si los conjuntos $\{x \in X : f(x) \leq a\}$ (resp. $\{x \in X : f(x) \geq a\}$) son cerrados para todo $a \in \mathbb{R}$.
Por su parte, definimos el *epigrafo* de una función $f : X \to \mathbb{R}$ como el conjunto:

$$\mathrm{epi}(f) := \{(x, a) \in X \times \mathbb{R} : a \geq f(x)\} \tag{3.3.7}$$

El teorema que ahora enunciamos es de importancia capital para entender el fundamento de este trabajo, y para entender la aplicación de nuestros RKHSs en los métodos de SVM. Éste prueba la existencia de, al menos, una solución general de un problema de SVM.

Teorema 3.3.3 (Existencia de soluciones SVM para pérdidas de Nemitski)**.** *Sea $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una función de pérdida convexa y sea P una distribución en $X \times Y$ tal que L es una pérdida P-integrable Nemitski. Sea también H un RKHS dado por un núcleo en X que es acotado y medible. Entonces, para cualquier $\lambda > 0$, existe al menos una solución general de la SVM $f_{P,\lambda}$.*

Demostración. Antes de pasar a su demostración, necesitamos los siguientes resultados auxiliares.

Lema 3.3.4 (Caracterización de los funcionales lineales de $E \times \mathbb{R}$ en $\mathbb{R}$)**.** *Sea E un espacio de Banach. Los funcionales lineales de $E \times \mathbb{R}$ en $\mathbb{R}$ son de la forma:*

$$(x, a) \to y(x) + \alpha a, \quad a \in \mathbb{R}, x \in E, \quad \text{para algún } y \in E', \tag{3.3.8}$$

es decir, la topología débil en $E \times \mathbb{R}$ sería la misma que la topología producto en $E \times \mathbb{R}$, engendrada por la topología débil en E, y la topología usual en $\mathbb{R}$.

Demostración. Es bien sabido que $E \times \mathbb{R}$ es una espacio de Banach con la norma definida por:

$$||(x,a)|| = ||x||_E + |a|, \tag{3.3.9}$$

para todo $(x,a) \in E \times \mathbb{R}$. Sea $z \in (E \times \mathbb{R})'$. Definimos $l_E : E \to \mathbb{R}$ por:

$$l_E(x) = z((x,0)), \tag{3.3.10}$$

para todo $x \in E$. Como z es lineal, también lo es l_E. Además, para todo $x \in E$,

$$|l_E(x)| = |z((x,0))| \leq ||z||_{E'} \cdot ||(x,0)||, \tag{3.3.11}$$

donde se ha usado la continuidad de z como funcional en $E \times \mathbb{R}$. De ello se sigue la continuidad de l_E como aplicación lineal de E en $\mathbb{R}$. En adelante, denotamos como $\hat{x}$ al único elemento de E' tal que $\hat{x} = l_E$.
De forma análoga, definiendo:

$$l_{\mathbb{R}}(y) = z((0,y)), \tag{3.3.12}$$

para todo $y \in \mathbb{R}$, se deduce que existe un único $\hat{y} \in \mathbb{R}$ tal que $\hat{y} = l_{\mathbb{R}}$.
Entonces,

$$z((x,y)) = z((x,0)) + z((0,y)) = l_E(x) + l_{\mathbb{R}}(y) = \hat{x}(x) + \hat{y}(y), \tag{3.3.13}$$

para todo $(x,y) \in E \times \mathbb{R}$. Denotamos entonces esta aplicación como $z = (\hat{x}, \hat{y})$.
Recíprocamente, si tomamos $\tilde{x} \in E'$ y $\tilde{y} \in \mathbb{R}'$, entonces, para todo $(x,y) \in E \times \mathbb{R}$, se tiene

$$|(\hat{x},\hat{y})((x,y))| = |\hat{x}(x) + \hat{y}(y)| \leq |\hat{x}(x)| + |\hat{y}(y)| \leq$$

$$\leq ||\hat{x}||_{E'} \cdot ||x||_E + ||\hat{y}||_{E'} \cdot ||y||_E \leq (||\hat{x}||_{E'} + ||\hat{y}||_{E'}) \cdot (||x||_E + ||y||_E) = ||(x,y)|| \cdot ||(\hat{x},\hat{y})||, \tag{3.3.14}$$

donde se han usado la continuidad de los funcionales involucrados, la desigualdad triangular, y la definición de norma de suma directa finita de espacios de Banach. De aquí se sigue que $(\hat{x}, \hat{y}) \in (E \times \mathbb{R})'$.
Con este último razonamiento, hemos caracterizado la topología producto del dual topológico de E por $\mathbb{R}$ como el dual topológico del producto $(E \times \mathbb{R})$. De ello llegamos a la caracterización (3.3.8), con tan solo especificar la forma que adquieren los funcionales lineales en $\mathbb{R}$. □

Lema 3.3.5. *Sea E un espacio de Banach reflexivo, y sea $f : E \to \mathbb{R} \cup \{\infty\}$ una aplicación convexa e inferiormente semicontinua. Si existe un $M > 0$ para el que el conjunto $A_E = \{x \in E : f(x) \leq M\}$ satisface $A_E \neq \emptyset$, y tal que A_E es acotado, entonces f tiene un mínimo absoluto en A_E.*

Demostración. Notemos que, por definición de función inferiormente semicontinua, A_E es cerrado. También, como f es convexa, es inmediato que A_E es convexo. Como A_E es acotado, necesariamente debe existir un $x_0 \in E$ tal que x_0 no esté en A_E. Probamos que x_0 no pertenence a la adherencia débil de A_E. Por el teorema de separación de Hahn-Banach, dado E un espacio normado de Banach, y dados $A_E \neq \emptyset$ cerrado y convexo, y $\{x_0\}$ compacto y también convexo, entonces existe un funcional lineal $\phi \in E'$ que separa x_0 de los puntos de A_E: existen $c_1 > c_2$, $c_1, c_2 \in \mathbb{R}$, tales que $\phi(y) > c_1 > c_2 > \phi(x_0)$, para todo $y \in A_E$. Ahora bien, como la topología débil en E es la menos fina que hace continuas a los elementos de E', entonces, el conjunto

$$C := \{x \in E : \phi(x) < c_1\}, \tag{3.3.15}$$

es obviamente débilmente abierto, y por construcción de ϕ tenemos $C \subset E \setminus A_E$ y $x_0 \in C$. Se deduce de esto que x_0 no es débilmente adherente a A_E, con lo que $E \setminus A_E$ es débilmente abierto. Entonces, A_E es débilmente cerrado.
Además, como A_E es acotado, y dada B la bola cerrada unidad en E, sabemos que existe un $a > 0$ tal que $A_E \subset aB$. Por el teorema A.0.22, como E es reflexivo, sabemos que B es débilmente compacto,

con lo que aB también lo es. Entonces, A_E es un subconjunto débilmente cerrado de un conjunto débilmente compacto, es decir, A_E es débilmente compacto.
Veamos que f es una aplicación débilmente inferiormente semicontinua. Notemos, en primer lugar, que epi(f) es convexo, dada la convexidad de f. Además, como f es inferiormente semicontinua (en la topología de partida en E), entonces epi(f) también es cerrado, pues epi(f) $=$ $\pi_E^{-1}(\{x \in E : f(x) \leq a\})$, con π_E la proyección $E \times \mathbb{R} \to E$. Por un razonamiento análogo al de la prueba de que A_E es débilmente compacto, podemos afirmar que epi(f) es débilmente cerrado en $E \times \mathbb{R}$. Por el lema 3.3.4, la condición de débilmente cerrado de epi(f) en $E \times \mathbb{R}$ implica que f es débilmente inferiormente semicontinua.
Resumiendo, sabemos lo siguiente:

a) A_E es débilmente compacto en E.

b) $f : E \to \mathbb{R} \cup \{\infty\}$ es débilmente inferiormente semicontinua.

Para acabar, y en virtud de que f es débilmente inferiormente semicontinua, y que A_E es débilmente compacto, razonamos por reducción al absurdo. Llamamos al ínfimo de f en A_E por $a = \inf_{x \in A_E} f(x)$, el cual existe por ser f inferiormente semicontinua en un compacto. Si el valor a no se alcanza, entonces, para cada $x \in A_E$, existe un $\varepsilon > 0$ tal que $f(x) - \varepsilon > a$. Como f es inferiormente semicontinua con la topología débil de E, existe un entorno V_x de x, en dicha topología, tal que $a < f(y)$ para todo $y \in V_x$. Entonces, $\cup_{x \in A_E} V_x$ es un recubrimiento abierto de A_E, luego por la compacidad de A_E existe un recubrimiento finito de la forma $A_E \subset \cup_{i=1}^n V_{x_i}$. Entonces, se sigue que $a' = \min_{i=1,\dots,n} f(x_i) - \varepsilon_i > a$, con los $\varepsilon_i > 0$ anteriores de la caracterización de ínfimo. Además, como $A_E \subset \cup_{i=1}^n V_{x_i}$, se tiene $a' < f(x)$ para todo $x \in A_E$. Esto contradice que a sea el ínfimo, pues hay otra cota inferior de A_E, a', mayor que a. Entonces, f alcanza el valor a. □

Volvemos a la demostración de 3.3.3:
Por hipótesis, k es medible , luego por el lema 1.4.3 tenemos que H consiste en funciones de X en $\mathbb{R}$ que son medibles. También, como k es acotado, haciendo uso del lema 1.4.2, tenemos que la aplicación id : $H \to L^\infty(P_X)$ es continua. Se sabe que una función convexa y a valores reales en un intervalo abierto de $\mathbb{R}$ es continua en el intervalo. Como $L(x, y, .) : \mathbb{R} \to \mathbb{R}$ es una función convexa para todo $(x, y) \in X \times Y$ (definición 3.2.11), y como $L(x, y, t) < \infty$ para todo $(x, y, t) \in X \times Y \times \mathbb{R}$, entonces L es continua en todo $\mathbb{R}$ respecto a su argumento real t, fijando los otros dos en $X \times Y$. Por lo tanto, por el lema 3.2.15, sabemos que la aplicación $\mathcal{R}_{L,P} : L^\infty(P_X) \to \mathbb{R}$ es continua. Así, con tan solo componer con la aplicación continua id : $H \to L^\infty(P_X)$, podemos decir que la aplicación $\mathcal{R}_{L,P} : H \to \mathbb{R}$ es continua.
Además, por el lema 3.2.12, tenemos también la convexidad de esta última aplicación.
Todo lo anterior, junto a la continuidad y convexidad de la aplicación $f \mapsto \lambda||f||_H^2$, implica la continuidad y convexidad de $f \mapsto \lambda||f||_H^2 + \mathcal{R}_{L,P}(f)$. Definimos ahora el siguiente conjunto:

$$A := \{f \in H : \lambda||f||_H^2 + \mathcal{R}_{L,P}(f) \leq M\}, \tag{3.3.16}$$

con $M := \mathcal{R}_{L,P}(0)$. Es claro que la función constante igual a 0, como elemento de H, está en A, y por tanto $A \neq \emptyset$. Además, si $f \in A$, es inmediato que, por la propia definición de A, tenemos que $\lambda||f||_H^2 \leq M$ o, en otras palabras, $A \subset (M/\lambda)^{1/2} B_H$, con B_H la bola cerrada unidad en H. Resumiendo, tenemos que A es un conjunto no vacío y acotado en el espacio de Hilbert H. La existencia del mínimo absoluto en todo H nos la da el lema 3.3.5. Notemos que H, por ser de Hilbert, es un espacio de Banach reflexivo (vía el teorema de representación de Riesz). □

Lo que acabamos de probar es clave en nuestro análisis: con pedir convexidad, continuidad y P-integrabilidad de Nemitski a L, si el espacio de funciones en el que buscamos nuestro minimizador $f_{P,\lambda}$ es un RKHS, con un núcleo reproductor acotado y medible, entonces podemos afirmar que existe una solución general de nuestra SVM y, además, por la proposición 3.3.1, sabemos que es única, dado

que L es convexa.
Este trabajo quedaría incompleto si no se mencionaran las pérdidas supervisadas y no supervisadas, así como, dentro del primer grupo, las pérdidas *margin-based*. Simplemente, las vamos a definir para dar un interesante resultado que permite, a partir del teorema 3.3.3, controlar el tamaño de los minimizadores a partir de la función pérdida y del parámetro de penalización.

Definición 3.3.6 (Pérdidas supervisadas y no supervisadas)**.** Sea $L : Y \times \mathbb{R} \to [0, \infty)$ una pérdida sin dependencia explícita en X. Diremos que L es una *pérdida supervisada* si y solo si es medible.
Sea ahora $L : X \times \mathbb{R} \to [0, \infty)$ otra función pérdida, sin dependencia explícita en los datos de salida. Diremos que L es una *pérdida no supervisada* si y solo si L es medible.

La motivación de esta terminología radica en la dependencia o independencia que el L-riesgo adquiere con respecto a las $y \in Y$. En el caso no supervisado, dicho L-riesgo toma la forma:

$$\mathcal{R}_{L,P}(f) = \int_X L(x, f(x))dP_X(x), \tag{3.3.17}$$

donde desaparece dicha dependencia respecto a los elementos de Y.

Definición 3.3.7 (pérdida *margin-based* (m.b.))**.** Sea $L : Y \times \mathbb{R} \to [0, \infty)$ una pérdida supervisada. Diremos que L es una pérdida margin-based (m.b.) si existe una *función representante* $\varphi : \mathbb{R} \to [0, \infty)$ tal que:

$$L(y, t) = \varphi(yt), \quad y \in Y, t \in \mathbb{R}. \tag{3.3.18}$$

Corolario 3.3.8. *Sea $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una pérdida convexa, y localmente Lipschitz continua. Sea también P una distribución de probabilidad en $X \times Y$, tal que el L-riesgo satisface $\mathcal{R}_{L,P}(0) < \infty$. Sea H un RKHS con núcleo medible y acotado en X. Entonces, para todo $\lambda > 0$, existe una única solución general SVM $f_{P,\lambda} \in H$.*
Además, si L es una función pérdida convexa m.b., representada por la función $\varphi : \mathbb{R} \to [0, \infty)$, entonces:

$$||f_{P,\lambda}||_H \leq \left(\frac{\varphi(0)}{\lambda}\right)^{1/2}. \tag{3.3.19}$$

Demostración. La unicidad de la solución general SVM es una consecuencia directa de la proposición 3.3.1.
En cuanto a la existencia, por la proposición 3.2.17 tenemos que L, por ser una pérdida localmente Lipschitz continua, es P-integrable de Nemistki, luego tenemos todas las hipótesis del teorema 3.3.3. Esto asegura la existencia de la solución general SVM $f_{P,\lambda}$.
Por la desigualdad (3.3.4), como $\mathcal{R}_{L,P}(0) = \varphi(0)$, ya tendríamos la acotación deseada. □

En el marco de espacios de funciones con núcleo reproductor, tratamos ahora la existencia y representación de soluciones SVM empíricas, que son aquellas soluciones SVM que cumplen la condición de ínfimo restringidos a muestras finitas $D := \{(x_1, y_1) \ldots, (x_n, y_n))\}$ sobre la función de L-riesgo (3.2.4).

Teorema 3.3.9 (Teorema de representación)**.** *Sea $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una pérdida convexa, $D := \{(x_1, y_1) \ldots, (x_n, y_n))\} \in (X \times Y)^n$, y H un RKHS sobre X. Entonces, para todo $\lambda > 0$, existe una única solución empírica SVM.*
Además, existen $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ tales que:

$$f_{D,\lambda}(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i), \quad x \in X. \tag{3.3.20}$$

Demostración. 1. *Existencia*: Dado que la convergencia en norma de H implica la convergencia puntual de los elementos de H, se puede probar la continuidad del L-riesgo empírico $\mathcal{R}_{L,D} : H \to [0,\infty)$ debido a la continuidad de la pérdida L (por ser convexa y finita en todos sus puntos), siguiendo un razonamiento idéntico al expuesto para el teorema 3.3.3. La existencia se sigue, también, con un razonamiento análogo al de aquel teorema, y particularizado sobre este L-riesgo.

2. *Representación*: Tomamos el conjunto finito $X' = \{x_1, \dots, x_n\}$, así como

$$H' = \vee\{k(\cdot, x_i) : i = 1, \dots, n\}. \tag{3.3.21}$$

Al igual que ocurría con (1.3.13), H' es el RKHS de núcleo reproductor $k|_{X'\times X'}$. Entonces, aplicando el punto anterior, podemos afirmar que existe una solución empírica $f_{P,\lambda,H'} \in H'$.
Tomamos el complemento ortogonal $H'^{\perp}$ de H', dentro del espacio general H. Dado $f \in H'^{\perp}$, entonces, como $k(\cdot, x_i) \in H'$ para todo $i = 1, \dots, n$, utilizando la propiedad reproductora de k, llegamos a

$$f(x_i) = \langle f, k(\cdot, x_i)\rangle = 0, \quad i = 1, \dots, n. \tag{3.3.22}$$

Denotamos por $P_{X'} : H \to H'$ a la proyección ortogonal de H en H'. Por la igualdad anterior, se tiene que

$$\mathcal{R}_{L,D}(P_{X'}f) = \frac{1}{n}\sum_{i=1}^{n} L(x_i, y_i, P_{X'}f(x_i)) = \mathcal{R}_{L,D}(f), \tag{3.3.23}$$

donde se ha usado la definición de L-riesgo empírico, y que $P_{X'}f(x_i) = f(x_i)$ para todos $f \in H$, $i \in \{1, \dots, n\}$, por definición de H'. Como $P_{X'}$ es proyección ortogonal, tenemos que $||P_{X'}f||_H \leq ||f||_H$, para todo $f \in H$.
Todo ello nos lleva a que

$$\inf_{f\in H}\left(\lambda||f||_H^2 + \mathcal{R}_{L,D}(f)\right) \leq \inf_{f\in H'}\left(\lambda||f||_H^2 + \mathcal{R}_{L,D}(f)\right) =$$

$$= \inf_{f\in H}\left(\lambda||P_{X'}f||_H^2 + \mathcal{R}_{L,D}(P_{X'}f)\right) \leq \left(\inf_{f\in H}\lambda||f||_H^2 + \mathcal{R}_{L,D}(P_{X'}f)\right). \tag{3.3.24}$$

De este modo, hemos obtenido la igualdad en la anterior cadena de desigualdades. Por lo tanto, $f_{D,\lambda,H'}$ minimiza la aplicación

$$f \mapsto \lambda||f||_H^2 + \mathcal{R}_{L,D}(f) \tag{3.3.25}$$

en H, y la unicidad de $f_{D,\lambda,H}$ implica que $f_{D,\lambda,H'} = f_{D,\lambda,H}$.
De este modo, $f_{D,\lambda} := f_{D,\lambda,H} \in H'$, con lo que la representación de esta solución empírica en términos de los $k(\cdot, x_i)$ es consecuencia de la definición de H'.

3. *Unicidad*: Se sigue inmediatamente de la convexidad de L y de la proposición 3.3.1.

□

Teorema 3.3.10 (Soluciones SVM no triviales)**.** *Sea $L : X \times Y \times \mathbb{R} \to [0,\infty)$ una pérdida convexa y P una distribución en $X \times Y$ tal que L es P-integrable Nemitski. Sea también H un RKHS de núcleo acotado y medible en X, y tal que $\mathcal{R}^*_{L,P,H} < \mathcal{R}_{L,P}(0)$. Entonces, para todo $\lambda > 0$, se tiene que $f_{P,\lambda} \neq 0$.*

Demostración. Por hipótesis, tiene que existir un elemento $g \in H$, $g \neq 0$, tal que $\mathcal{R}_{L,P}(g) < \mathcal{R}_{L,P}(0)$. Sea $\alpha \in [0,1]$. Entonces, por convexidad de L (y por tanto de $\mathcal{R}$), tenemos:

$$\lambda||\alpha g||_H^2 + \mathcal{R}_{L,P}(\alpha g) \leq \lambda\alpha^2||g||_H^2 + \alpha\mathcal{R}_{L,P}(g) + (1-\alpha)\mathcal{R}_{L,P}(0) =: h(\alpha), \tag{3.3.26}$$

donde $g = \alpha \cdot g + (1-\alpha)\cdot 0$.
Denotamos $\mathcal{R}_{L,P}(0) = b$, $\mathcal{R}_{L,P}(g) = c$ $(b > c)$, y $d = \lambda||g||_H^2 > 0$. Usamos ahora la relación $\mathcal{R}^*_{L,P,H} < \mathcal{R}_{L,P}(0)$. El polinomio cuadrático $\alpha \to h(\alpha)$, con $h : [0,1] \to [0,\infty)$, alcanza su mínimo en

$\alpha^* \in (0,1]$. Esto es así porque que la derivada de $h(\alpha)$ es $h'(\alpha) = 2d\alpha + (c-b)$, con lo que $h'(0) < 0$. Entonces:

$$\lambda||\alpha^* g||_H^2 + \mathcal{R}_{L,P}(\alpha^* g) \leq h(\alpha^*) < h(0) = \lambda||0||_H^2 + \mathcal{R}_{L,P}(0) = \mathcal{R}_{L,P}(0), \tag{3.3.27}$$

donde g tiene que ser necesariamente distinto de 0, pues $\mathcal{R}_{L,P}(0) > \mathcal{R}_{L,P}(g)$. Esto implica que el minimizador $f_{P,\lambda}$ es distinto de cero. □

3.4. *Large RKHS*. Núcleos universales

En este apartado, resumimos y demostramos los resultados que, a mi juicio, mejor ilustran la capacidad que tienen los RKHSs de limitar nuestra búsqueda de minimizadores a estos espacios. Esto es posible a través de la densidad que algunos de estos RKHSs presentan en el espacio de funciones continuas, por medio de los llamados *núcleos universales*. Esto tiene interesantes consecuencias para la formulación geométrica de, por ejemplo, los problemas de clasificación vistos anteriormente, a través de hiperplanos definidos con una aplicación característica de dicho núcleo universal.
También, generalizamos esta noción de universalidad del núcleo por medio de series de Taylor, para afirmar la existencia de algunos de estos núcleos, con expresiones de sobra conocidas. Por otra parte, mostraremos la capacidad de los RKHS H con estos núcleos universales para expresar funciones riesgo de Bayes restringiendo la búsqueda, estrictamente, a H. Ya para acabar, damos dos interesantes resultados de densidad de los RKHS en espacios de funciones tales como los espacios L^p y los de funciones continuas que se anulan en el infinito. Como último apunte, esta presencia de los RKHS, vía densidad, en espacios más grandes, es lo que califica a esta clase de espacios de Hilbert como *large RKHS*.
Empezamos definiendo lo que se entiende como RKHS grande (large RKHS). En todo el análisis que sigue nos limitamos, como es costumbre ya, a espacios de funciones $f : X \to \mathbb{R}$, para X arbitrario y no vacío. Denotamos por $C(X)$ al espacio de funciones continuas de la forma $f : X \to \mathbb{R}$. También, por simplicidad, nos limitaremos a X métrico y compacto.

Definición 3.4.1. Sea un núcleo continuo k y con valores reales en un espacio métrico y compacto X. Diremos que k es **núcleo universal** en X, o simplemente **universal**, si el RKHS H de k es denso, con respecto a la norma $||\cdot||_\infty$, en el espacio de funciones continuas de la forma $f : X \to \mathbb{R}$. En otras palabras, k es universal si para toda función $g \in C(X)$ y para todo $\varepsilon > 0$, existe una $f \in H$ tal que

$$||f - g||_\infty \leq \varepsilon. \tag{3.4.1}$$

Una forma equivalente de aportar esta última definición es a través de un espacio y una aplicación característicos cualesquiera, H_0 y $\phi_0 : X \to H_0$ respectivamente, para el núcleo k. Para ello, recordemos la forma con que, vía (1.3.11), caracterizábamos un RKHS cualquiera a través de su núcleo reproductor con el conjunto que sigue

$$H := \{f : X \to \mathbb{K} : \exists w \in H_0 \text{ con } f(x) = \langle w, \phi_0(x)\rangle_{H_0} \text{ para todo } x \in X\}, \tag{3.4.2}$$

con lo que la definición 3.4.1 puede darse en los siguientes términos: k es universal si y solo si, para todo $g \in C(X)$ y para todo $\varepsilon > 0$, existe un $w \in H_0$ tal que

$$||\langle w, \phi_0(\cdot)\rangle_{H_0} - g||_\infty \leq \varepsilon. \tag{3.4.3}$$

Veremos más adelante que esta observación adquiere su relevancia a la hora de construir núcleos universales.
Entramos ahora en las propiedades geométricas de los núcleos reproductores universales. Necesitamos una definición de separabilidad general, que nos dará posteriormente un criterio de separabilidad sobre los compactos de X. Recordemos que hablar de núcleos equivale a hablar de núcleos reproductores, a través de su asociación con un único RKHS.

Definición 3.4.2. Sea k un núcleo en un espacio métrico X con RKHS H. Diremos que k **separa a los conjuntos disjuntos** $A, B \subset X$ si y solo si existe una función $f \in H$ tal que $f(x) < 0$ si $x \in B$ y $f(x) > 0$ si $x \in A$.

Proposición 3.4.3. *Sea (X, d) un espacio métrico y compacto, y k un núcleo universal en X. Entonces, k separa a todos los subconjuntos compactos disjuntos de X.*

Demostración. Sean $C, D \subset X$ compactos disjuntos, y definimos la siguiente función:

$$h(x) := \frac{\text{dist}(x, B)}{\text{dist}(x, A) + \text{dist}(x, B)} - \frac{\text{dist}(x, A)}{\text{dist}(x, A) + \text{dist}(x, B)}, \tag{3.4.4}$$

donde definimos la distancia de un conjunto D de X a un punto cualquiera $x \in X$ como es habitual,

$$\text{dist}(x, D) := \inf_{y \in D} \text{dist}(x, y). \tag{3.4.5}$$

Esta función h es continua, dada la continuidad de la aplicación $x \in X \mapsto \text{dist}(x, A)$, $A \subset X$, y teniendo en cuenta también que, como A y B son compactos y disjuntos, entonces las funciones $\text{dist}(x, A)$ y $\text{dist}(x, B)$ no se pueden anular al mismo tiempo. Además, es inmediato que $h(x) = -1$ si $x \in B$ y $h(x) = 1$ si $x \in A$.
Ahora, usamos la propiedad de universalidad del núcleo k sobre el RKHS H. Ello implica que, dada la densidad de H en $C(X)$, y como $h \in C(X)$, tenemos que existe $f \in H$ tal que $||f - h||_\infty \leq 1/2$. Entonces, como h solo toma los valores 1 en A y -1 en B, es inmediato que, entonces, $f(x) \geq 1/2$ para todo $x \in A$, y que $f(x) \leq -1/2$. Por tanto, tomando $f \in H$ como función separante, hemos demostrado el resultado. □

Una consecuencia simple de esta proposición 3.4.3 es que *siempre* podremos separar dos puntos distintos $x \neq x'$ en X por medio de una función del RKHS H. Este hecho es esencial a la hora de hablar de la inyetividad de aplicaciones características asociadas a núcleos univerales.
A pesar de la simpleza de su demostración, la proposición 3.4.3 tiene implicaciones geométricas que dan gran parte del sentido al uso de RKHS en los problemas de clasificación. Para ver esto, tomemos un núcleo universal k con espacio característico H_0 y una aplicación característica $\phi_0 : X \to H_0$. Tomemos ahora un subconjunto finito $X' := \{x_1, \ldots, x_n\} \subset X$, para algún $n \in \mathbb{N}$, y consideremos un subconjunto de elecciones $y_1, \ldots, y_n$ del espacio de salida de un problema SVM de clasificación estándar, i.e., $Y := \{-1, 1\}$. Como X' es finito, entonces es compacto, luego la proposición 3.4.3 garantiza la existencia de una función f en el RKHS H asociado a k tal que $y_i f(x_i) > 0$ para todo $i = 1, \ldots, n$, y esto, mentando al ejemplo 3.2.7, nos permite *clasificar* los valores de X', distinguiéndolos del resto de elementos de X. En consecuencia, por el resultado 1.3.11, al poder representar dicha función de decisión f a través del producto interno de H_0 como $f = \langle w, \phi_0(\cdot) \rangle_{H_0}$, para un $w \in H_0$ adecuado, podemos centrar nuestro estudio en el conjunto $\{(\phi_0(x_1), y_1), \ldots, (\phi_0(x_n), y_n)\}$, el cual puede ser separado en H_0 por medio de un hiperplano determinado por w de la forma

$$y_i \langle w, \phi(x_i) \rangle_{H_0} > 0, \;\; i \in \{1, \ldots, n\}, \tag{3.4.6}$$

los cuales serían los valores de X que se corresponderían, precisamente, con los datos x_i de X' que queremos *clasificar positivamente* (i.e., que no penalizan en el L-riesgo).
En resumen, hemos convertido un problema de clasificación de un conjunto X, al que pedimos que sea métrico y compacto, en una clasificación geométrica dada por un hiperplano en un espacio de Hilbert con núcleo reproductor H_0. No obstante, es necesario hacer notar que este razonamiento se ha hecho para un conjunto de muestras muy concreto (un conjunto finito X' en un conjunto métrico y compacto). Esto, con lo que sabemos, no se puede generalizar directamente a conjuntos de muestras arbitrarios en $\mathbb{R}^n$, con $n \in \mathbb{N}$.
Esta lectura de la geometría mediante hiperplanos sugiere la siguiente pregunta: ¿tales núcleos universales van a existir siempre? La respuesta, en espacios métricos y compactos, es que sí, e incluso

podemos garantizar que algunos de los núcleos estándar más conocidos, como el núcleo RBF (radial basis function), es un caso. Yendo más lejos con la pregunta, dada la definición 3.4.1: ¿podría extenderse esta noción a espacios *topológicos* compactos? La respuesta, que no trabajaremos aquí por sobrepasar el alcance de este trabajo, está en que se puede extender a espacios topológicos de estas características, con un matiz: no existe núcleo universal si la topología no está generada por una métrica. Necesitamos, para ver esto, un lema que enunciará propiedades nucleares en el análisis que viene.

Lema 3.4.4. *Sea X un espacio métrico y compacto y sea k un núcleo en X. Entonces, las siguientes afirmaciones se cumplen:*

1. *Cualquier aplicación característica asociada a k es inyectiva.*
2. *Se tiene que $k(x,x) > 0$ para todo $x \in X$.*
3. *Cualquier restricción de k a un compacto $X' \subset X$ es universal.*
4. *El núcleo normalizado $k^* : X \times X \to \mathbb{R}$, definido como:*

$$k^*(x,x') := \frac{k(x,x')}{\sqrt{k(x,x)k(x'x')}}, \quad x, x' \in X, \tag{3.4.7}$$

es universal.

Demostración. Denotamos por H_0 y por ϕ a un espacio y a una aplicación caractarísticos arbitrarios de k. Las tres primeras afirmaciones son directas de la definición de núcleo universal y de la proposición 3.4.3. La inyectividad se sigue de la condición separante de al menos un elemento del RKHS de k para dos conjuntos unipuntuales $\{x\}$ y $\{x'\}$ de X cualesquiera, y de la representación de los elementos de H a través de 1.3.11. Para ver esto, tomamos $x \neq x'$ cualesquiera en X y $\phi_0 : X \to H_0$ una aplicación característica de k. Por 3.4.3, existe un $f \in H$ que cumple lo siguiente:

$$f(x) \neq f(x') \implies \langle w, \phi(x) \rangle_{H_0} \neq \langle w, \phi(x') \rangle_{H_0}, \tag{3.4.8}$$

para algún $w \in H_0$, donde se ha usado (1.3.11). Así, $\phi_0(x) \neq \phi(x')$, y tenemos la inyectividad deseada. Para los otros tres puntos, particularizamos H_0 en el RKHS asociado H. Para ver que $k(x,x) > 0$ para todo $x \in X$, recurriendo de nuevo a 3.4.3, como X es compacto, debe existir un $f \in H$ tal que $f(x) > 0$ para todo $x \in X$. Basta tomar, en el enunciado de dicha proposición, el compacto X y el vacío. Entonces, utilizando la propiedad reproductora de k sobre f en cada $x \in X$, y usando la desigualdad de Cauchy-Schwarz, se tiene que

$$0 < |f(x)| = |\langle f, k(\cdot,x) \rangle_H| \leq ||f||_H \cdot ||k(\cdot,x)||_H, \tag{3.4.9}$$

y como $k(x,x) = \langle \phi(x), \phi(x) \rangle_H \geq 0$, entonces $k(x,x) > 0$ para todo $x \in X$.
En cuanto al tercer punto, nótese que si k está asociado a un RKHS H que es denso en $C(X)$ con la norma del supremo, es inmediato que si restringimos las funciones que son elementos de estos espacios al compacto $X' \subset X$, seguimos teniendo dicha densidad. Nótese que $||f|_{X'} - g|_{X'}||_\infty \leq ||f - g||_\infty$.
Veamos ya el último punto. Consideramos que la aplicación característica $\phi : X \to H$ es la canónica. Definimos $a(x) := k(x,x)^{-1/2}$, para todo $x \in X$, lo cual tiene sentido por el segundo punto. Entonces, por el lema 1.1.5, la aplicación $\phi' := a \cdot \phi : X \to H$ es aplicación caraterística de k^*, ya que

$$\langle \phi'(x'), \phi'(x) \rangle_H = \frac{1}{\sqrt{k(x,x)k(x',x')}} \langle \phi(x'), \phi(x) \rangle_H = \frac{1}{\sqrt{k(x,x)k(x',x')}} k(x,x'), \tag{3.4.10}$$

con lo que k^* es un núcleo, y H es uno de sus espacios característicos.
En cuanto a su universalidad, tomemos $h \in C(X)$ y $\varepsilon > 0$. Como k es acotado, al ser continuo de un

espacio compacto a $\mathbb{R}$, entonces $d := ||a||_\infty < \infty$. De esta forma, utilizando además que $a(x) > 0$ para todo $x \in X$ por el segundo punto de este lema, y dado que k es universal con aplicación característica ϕ, entonces $||f - \frac{h}{a}||_\infty = ||\langle f, \phi\rangle_H - \frac{h}{a}||_\infty \leq \frac{\varepsilon}{d}$ para cierto $f \in H$. Además, ya hemos visto que $\phi' = a \cdot \phi$ es aplicación característica del núcleo k^*, y que el espacio H es característico de k^*, de modo que

$$||f - h||_\infty = ||\langle f, a(\cdot)\phi(\cdot)\rangle_H - h||_\infty \leq ||a||_\infty \cdot ||\langle f, \phi(\cdot)\rangle_H - \frac{h}{a}||_\infty \leq d \cdot \frac{\varepsilon}{d} = \varepsilon. \tag{3.4.11}$$

Entonces, por definición, k^* es universal. □

Para investigar la existencia de esta clase de núcleos, demostramos una condición suficiente para la universalidad de un núcleo reproductor, para la que el teorema de Stone-Weierstrass es clave.

Teorema 3.4.5 (Un test de universalidad)**.** *Sea X un espacio métrico y compacto, y sea k un núcleo continuo en X con $k(x, x) > 0$ para todo $x \in X$. Supongamos que existe una aplicación característica inyectiva $\phi : X \to l^2(\mathbb{N})$ para el núcleo k, y escribimos como $\phi_n : X \to \mathbb{R}$ a su componente n-ésima, de tal modo que $\phi(x) = (\phi_n(x))_{n=1}^\infty$, para todo $x \in X$. Si el espacio dado por*

$$A = \vee \{\phi_n : n \in \mathbb{N}\}, \tag{3.4.12}$$

forma un álgebra, entonces k es un núcleo universal.

Demostración. Se tiene que

$$||(\phi_n(x))||_{l^2}^2 = k(x, x) > 0, \ \ x \in X. \tag{3.4.13}$$

Además, la continuidad de k en X implica que, para todo $n \in \mathbb{N}$, $\phi_n : X \to \mathbb{R}$ es continua, en virtud del resultado de caracterización de núcleos continuos 1.4.17. Esto prueba que $A \subset C(X)$.
También, la inyectividad de la aplicación característica ϕ implica que A, tal y como está definida, separa puntos de X. Basta tomar $x, x' \in X$ cualesquiera, y una de las ϕ_n, con $n \in \mathbb{N}$, que nos dé $\phi_n(x) \neq \phi_n(x')$. Tal ϕ_n tiene que existir o, de lo contrario, ϕ no sería inyectiva. Entonces, teniendo en cuenta que estamos tratando con funciones que llegan a $\mathbb{R}$, en virtud del teorema de Stone-Weierstrass, podemos afirmar que A es denso en $C(X)$ en norma $|| \cdot ||_\infty$.
Sea ahora $h \in C(X)$, y sea $\varepsilon > 0$. Existe una función $f \in A$, es decir, de la forma

$$f = \sum_{i=1}^m a_i \phi_{n_i}, \tag{3.4.14}$$

para un cierto $m \in \mathbb{N}$, tal que $||f - h||_\infty \leq \varepsilon$. Dado ahora $n \in \mathbb{N}$, definamos la sucesión de números reales dada por $w_n = a_n$ si existe un índice i en la definición anterior de f tal que $n_i = n$, y por $w_n = 0$ en caso contrario. Entonces, tenemos que $w = (w_n)_n \in l^2(\mathbb{N})$, pues w tiene un número finito de componentes no nulas. Además, se tiene que

$$f(\cdot) = \langle w, \phi(\cdot)\rangle_{l^2(\mathbb{N})}. \tag{3.4.15}$$

Entonces, repitiendo un razonamiento visto cuando definíamos la univeralidad de los núcleos, tenemos que

$$||\langle w, \phi(\cdot)\rangle_{l^2(\mathbb{N})} - h||_\infty \leq \varepsilon, \tag{3.4.16}$$

lo que equivale a que k sea un núcleo universal. □

A partir del teorema anterior 3.4.5, estamos en condiciones de construir, vía series de Taylor, ejemplos de núcleos universales.

Corolario 3.4.6. *Sea $r \in [0, \infty)$ fijo, y sea $f : (-r, r) \to \mathbb{R}$ una función $\mathcal{C}^\infty((-r, r))$, tal que admite la serie de Taylor que sigue en un entorno del origen,*

$$f(t) = \sum_{n=0}^{\infty} a_n t^n, \quad t \in (-r, r). \tag{3.4.17}$$

Fijemos el compacto $X := \{x \in \mathbb{R}^d : ||x||_2 \leq \sqrt{r}\}$, y consideramos $\langle \cdot, \cdot \rangle$ el producto escalar usual en $\mathbb{R}^d$. Si tenemos que $a_n > 0$ para todo $n \in \mathbb{N}_0$, entonces la función dada por

$$k(x, x') = f(\langle x, x' \rangle), \quad x, x' \in X, \tag{3.4.18}$$

forma un núcleo universal en los compactos de X.

Demostración. Ya vimos en el lema 1.1.8 que una función k, en estas condiciones, es un núcleo en X, y que admite como espacio característico a $l^2(\mathbb{N}_0^d)$, con una aplicación característica $\phi : X \to l^2(\mathbb{N}_0^d)$ dada por:

$$\phi(x) := \left(\sqrt{a_{j_1+\cdots+j_d} c_{j_1}, \ldots, c_{j_d}} \prod_{i=1}^{d} x_i^{j_i} \right)_{j_1,\ldots,j_d \in \mathbb{N}_0}, \quad x \in X, \tag{3.4.19}$$

donde se ha mantenido la misma notación que en 1.1.8. Entonces, el núcleo definido por esta aplicación ϕ, dado como $k(x, x') = \langle \phi(x'), \phi(x) \rangle_{l^2(\mathbb{N}_0^d)}$, es continuo, y $a_0 > 0$ implica que $k(x, x) > 0$, para todo $x \in X$.
Veamos la inyectividad de ϕ. Por reducción al absurdo, supongamos que $\phi(x) = \phi(x')$ para ciertos $x \neq x'$ en X. Entonces, por (3.4.19), tendríamos $x = x'$, ya que $a_n > 0$ para todo $n \in \mathbb{N}$, y $c_i > 0$ para todo $i \in \mathbb{N}_0$, lo que contradice la hipótesis. Por tanto, ϕ es inyectiva.
Así, como el conjunto de polinomios forma un álgebra, y como la suma en $l^2(\mathbb{N}_0^d)$ se define componente a componente, entonces el espacio que sigue

$$\vee \left\{ \phi_{j_1,\ldots,j_d} : j_1, \ldots, \mathrm{j}_d \in \mathbb{N}_0 \right\}, \tag{3.4.20}$$

forma un álgebra. Por lo tanto, por el teorema anterior 3.4.5, k es núcleo universal. □

Presentamos ahora, en forma de corolario, una lista de núcleos universales, cuyas condiciones de núcleo y universalidad se siguen de los resultados vistos. Nos seguimos restringiendo a subespacios compactos, ahora en $\mathbb{R}^n$, para algún $n \in \mathbb{N}$.

Corolario 3.4.7 (Ejemplos de núcleos universales). *Sea X un subepacio compacto de $\mathbb{R}^n$, $b > 0$, y $a > 0$. Denotamos por $\langle \cdot, \cdot \rangle$ al producto escalar usual en $\mathbb{R}^n$. Entonces, las siguientes funciones de $X \times X$ a $\mathbb{R}$ son núcleos universales en X:*

1. *Nucleo exponencial:*
$$k(x, x') = \exp\left(\langle x, x' \rangle\right). \tag{3.4.21}$$

2. *Núcleo gaussiano (RBF):*
$$k_b(x, x') = \exp\left(-b^{-2}||x - x'||_2^2\right). \tag{3.4.22}$$

3. *Núcleo binomial:*
$$k(x, x') = (1 - \langle x, x' \rangle)^{-a}. \tag{3.4.23}$$

Demostración. A través de la expresión (1.1.13), y de las proposiciones 1.1.9 y 1.1.10, ya hemos demostrado o se sigue directamente del lema 1.1.8 que estas aplicaciones son núcleos.
Con los razonamientos anteriores, para los que se utilizó el hecho de que estamos tratando con núcleos tipo Taylor, por el corolario 3.4.6 se sigue directamente su universalidad.

□

A continuación, daremos tres resultados que desembocan en una de las razones de ser de este trabajo: *el ínfimo del L-riesgo asociado a una pérdida L continua y P-integrable Nemitski, estará en el mismo RKHS H si el núcleo asociado es universal.* Para ello, recordemos que el problema de SVM que nosotros estamos tratando consiste en encontrar el ínfimo del L-riesgo de nuestro problema de SVM, esto es, omitiendo el término de penalización

$$\mathcal{R}^*_{L,P} = \inf\{\mathcal{R}_{L,P}(f) : f \in F\}, \tag{3.4.24}$$

sobre F un espacio de funciones medibles arbitrario.
El primer resultado trata los riesgos de Bayes a través de funciones decisión medibles y *acotadas*. En esta proposición, así como en el teorema que sigue, no hace falta imponer que X sea métrico y compacto. Más adelante sí que tendremos que hacerlo.

Proposición 3.4.8. *Sea $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una pérdida y P una distribución de probabilidad en $X \times Y$ tal que L es una pérdida P-integrable Nemitski. Entonces, se tiene la igualdad*

$$\mathcal{R}^*_{L,P,L^\infty(P_X)} = \mathcal{R}^*_{L,P}. \tag{3.4.25}$$

Demostración. Veamos que si cogemos una función medible de X en $\mathbb{R}$ tal que haga el L-riesgo finito, existe una sucesión en $L^\infty(P_x)$ cuyas imágenes por $\mathcal{R}_{L,P}$ convergen a $\mathcal{R}_{L,P}(f)$. Esto probaría que el ínfimo en el espacio de todas las funciones medibles es el ínfimo en $L^\infty(P_X)$.
Tomemos pues una función medible cualquiera $f : X \to \mathbb{R}$, con respecto a P_X. Entonces, podemos construir una sucesión de funciones medibles y acotadas $\{f_n\} \subset L^\infty(P_X)$ de la forma $f_n = f1_{\{x:|f(x)|>n\}}$, la cual cumple que

$$\begin{aligned}
&|\mathcal{R}_{L,P}(f_n) - \mathcal{R}_{L,P}(f)| \\
&= |\int_{\{|f|>n\}\times Y} (L(x, y, f_n(x)) - L(x, y, f(x)))\, dP(x, y) \\
&+ \int_{\{|f|\leq n\}\times Y} (L(x, y, f_n(x)) - L(x, y, f(x))\, dP(x, y)| \\
&\leq \int_{\{|f|\leq n\}\times Y} |L(x, y, 0) - L(x, y, f(x))| dP(x, y) \\
&\leq \int_{\{|f|\leq n\}\times Y} (b(x, y) + h(0) + L(x, y, f(x)))\, dP(x, y),
\end{aligned}$$

donde se ha usado la desigualdad triangular dos veces, y donde $b(x, y)$ y $h(t)$ son las funciones que nos dan la acotación de la definición de P-integrable Nemitski.
De esta forma, como $\mathcal{R}_{L,P}(f) < \infty$ por hipótesis, y como $b \in L^1(P)$, aplicando el teorema de la convergencia dominada es directa la convergencia $\mathcal{R}_{L,P}(f_n) \to \mathcal{R}_{L,P}(f)$ para $n \to \infty$. Y esto prueba la proposición. □

Usando la anterior proposición 3.4.8, probamos una importante característica sobre el espacio F: con pérdidas continuas y P-integrables Nemitski, $\mathcal{R}^*_{L,P,F} = \mathcal{R}^*_{L,P}$ para F un denso en $L^\infty(P_X)$.

Teorema 3.4.9 (Aproximación puntual por conjuntos densos)**.** *Sea P una distribución sobre $X \times Y$, $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una pérdida continua y P-integrable Nemitski, y $F \subset L^\infty(P_X)$. Supongamos también que, para todo $h \in L^\infty(P_X)$ existe una sucesión $\{f_n\} \subset F$ tal que* $\sup_{n\in\mathbb{N}} ||f_n||_\infty < \infty$ *y*

$$\lim_{n\to\infty} f_n(x) = h(x), \tag{3.4.26}$$

para P_X-casi todo $x \in X$. Entonces, $\mathcal{R}^*_{L,P,F} = \mathcal{R}^*_{L,P}$.

Demostración. Por la prosicion anterior 3.4.8, ya sabemos que $\mathcal{R}^*_{L,P,L^\infty(P_x)} = \mathcal{R}^*_{L,P}$. Además, por la contención $F \subset L^\infty(P_X)$, es clara la desigualdad $\mathcal{R}^*_{L,P,L^\infty(P_x)} \leq \mathcal{R}^*_{L,P,F}$. Veamos la desigualdad contraria.
Fijemos una función $h \in L^\infty(P_X)$. Sea $\{f_n\}$ una sucesión que cumpla las condiciones del enunciado. Por el lema 3.2.15, dado que se cumplen todas las condiciones del primer punto, tenemos la convergencia que sigue

$$\lim_{n\to\infty} \mathcal{R}_{L,P}(f_n) = \mathcal{R}_{L,P}(h). \tag{3.4.27}$$

Por un razonamiento análogo al de la proposición anterior, llegamos al aserto enunciado, pues tal convergencia implica la desigualdad $\mathcal{R}^*_{L,P,F} \leq \mathcal{R}_{L,P,L^\infty(P_X)}$. □

Y ahora, a partir del teorema 3.4.9, estamos en condiciones de probar que los $RKHS$ con núcleos universales en X nos dan el L-riesgo de Bayes para pérdidas continuas y P-integrables Nemitski. Volvemos, entonces, a suponer que X es métrico y compacto.

Corolario 3.4.10. *Sea X un espacio métrico y compacto, H el RKHS de un núcleo univesal en X, y P una distribución en $X \times Y$. Sea también $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una pérdida P-integrable Nemitski. Entonces:*

$$\mathcal{R}^*_{L,P,H} = \mathcal{R}^*_{L,P}. \tag{3.4.28}$$

Demostración. Usaremos la igualdad del espacio de funciones continuas $C(X)$ y del espacio de funciones continuas con soporte compacto, que denotamos por $C_c(X)$, dado que estamos manejando un espacio X compacto.
Así, fijamos $f \in L^\infty(P_X) \subset L^1(P_X)$. Conocida la densidad de $C_c(X)$ en $L_1(P_X)$, por la igualdad anterior tenemos densidad de $C(X)$ en $L_1(P_X)$. Entonces, existe una sucesión $\{f_n\} \subset C(X)$ con $||f_n - f||_1 \to 0$ para $n \to \infty$.
Truncamos ahora los elementos f_n a alturas $\pm||f||_\infty \in \mathbb{R}$, para todo $n \in \mathbb{N}$. Esto significa que, fijado un $x \in X$, consideramos $f_n(x)$ como es, siempre que $|f_n(x)| \leq ||f||_\infty$; $f_n(x) = ||f||_\infty$ si $f_n(x) > ||f||_\infty$, y $f_n(x) = -||f||_\infty$ si $f_n(x) < -||f||_\infty$. Sin pérdida de generalidad, nos limitamos a una subsucesión $(f_n) \subset C(X)$ adecuada de la sucesión anterior en $C(X)$ tal que $||f_n||_\infty \leq ||f||_\infty$ para todo $n \geq 1$, y $f_n(x) \to f(x)$ para P_X-casi todo $x \in X$. Esto último es posible porque la convergencia en $L^1(P_X)$ implica la convergencia de cierta subsucesión casi seguro respecto a la medida dada en X.
Ahora, entra en juego la universalidad de H: tal propiedad nos confirma la existencia, sobre cada $f_n \in C(X)$, de una función $g_n \in H$, para todo $n \in \mathbb{N}$, tal que $||f_n - g_n||_\infty \leq 1/n$. Entonces, se tiene que

$$||g_n||_\infty = ||\,(f_n - g_n)\,||_\infty + ||f_n||_\infty \leq \frac{1}{n} + ||f||_\infty, \tag{3.4.29}$$

para todo $n \in \mathbb{N}$. De esto se sigue que $||g_n||_\infty \leq ||f||_\infty + 1$ para todo $n \in \mathbb{N}$. Además, de la convergencia uniforme $||g_n - f||_\infty \to 0$ se sigue la convergencia P_X-casi seguro en X de la forma $g_n(x) \to f(x)$.
Entonces, $\{g_n\}$ es una sucesión que cumple las condiciones del teorema anterior 3.4.9 sobre $F = H$, luego tenemos la igualdad buscada: $\mathcal{R}^*_{L,P,H} = \mathcal{R}^*_{L,P}$. □

Observemos que, por los resultados anteriores, la universalidad de un núcleo con RKHS H garantiza que

$$\inf_{f\in H}\mathcal{R}_{L,P}(f) = \mathcal{R}^*_{L,P}, \tag{3.4.30}$$

sobre cualquier pérdida P-integrable Nemitski.
No obstante, este resultado se ha limitado a espacios métricos y compactos X. Ello implica que, con lo que sabemos, espacios como $\mathbb{R}^n$, para algún $n \in \mathbb{N}$, se quedan fuera del análisis, así como cualquier conjunto *infinito* con topología discreta, las cuales son situaciones interesantes en estos estudios de problemas de SVM. Para suplir esta carencia, aportamos el siguiente teorema, el cual nos permitirá, a través de pérdidas P-integrables Nemitski de orden $p \in [1, \infty)$ y continuas, identificar los L-riesgos de Bayes de entre todas las funciones medibles con los de un denso F de $L^p(P_X)$.

Teorema 3.4.11 (Aproximación por funciones p-integrables)**.** *Sea P una distribución de probabilidad en $X \times Y$ y $L : X \times Y \times \mathbb{R} \to [0, \infty)$ una pérdida continua y P-integrable Nemitski de orden $p \in [1, \infty)$. Entonces, para todo denso $F \subset L^p(P_X)$, se tiene que*

$$\mathcal{R}^*_{L,P,F} = \mathcal{R}^*_{L,P}. \tag{3.4.31}$$

Demostración. Sabemos que $L^\infty(P_X) \subset L^p(P_X)$. También, se ha probado que $\mathcal{R}_{L,P,L^\infty(P_X)} = \mathcal{R}_{L,P}$, en virtud de la proposición 3.4.8. Estos dos hechos implican que $\mathcal{R}_{L,P,L^p(P_X)} = \mathcal{R}_{L,P}$. Entonces, por la densidad de F en $L^p(P_X)$, y la continuidad de $\mathcal{R}_{L,P} : L^p(P_X) \to [0, \infty)$ (por el lema 3.2.15), tenemos la convergencia deseada en $\mathbb{R}$ de las imágenes de elementos de F hacia cualquier $\mathcal{R}_{L,P,L^p(P_X)}(f)$, para todo $f \in L^p(P_X)$. Entonces, por la primera observación, tenemos la igualdad de riesgos de Bayes enunciada. □

Aunque no se ha especificado ninguna forma concreta que estos espacios F pudieran tomar, bajo ciertas condiciones sobre la integrabilidad del núcleo se puede afirmar que dicha F puede ser un RKHS. No entramos en el detalle de esto porque estos resultados nos llevarían muy lejos.

Apéndice A

Resultados adicionales de Análisis Funcional

En este apéndice vamos a comenzar presentando el resultado elemental de compleción de un espacio con producto escalar a un espacio de Hilbert.

Teorema A.0.1 (Compleción a un espacio de Hilbert)**.** *Si F es un espacio con producto escalar $\langle\cdot,\cdot\rangle_F$, entonces existe un espacio de Hilbert H con producto escalar $(\cdot,\cdot)_H$ y una aplicación $A : F \to H$ tales que:*

1 A es inyectiva.

2 A es lineal.

3 $(Ax, Ay)_H = \langle x, y\rangle_F$ para todos $x, y \in F$.

4 $A(F) = \{Ax : x \in F\}$ es denso en H. Si F es ya completo, entonces $A(F) = H$.

Además, H es único salvo isomorfismo isométrico, y se le denomina ***la compleción de*** *F.*

Notemos que A nos da una biyección entre elementos de F y de $A(F)$, luego podemos pensar en A como una equivalencia entre los elementos x del espacio F y los elementos de $A(F)$ denso en H, con la diferencia de que en el segundo tenemos la ventaja de manejar un espacio completo. Intuitivamente, hemos llenado los huecos que los límites de las sucesiones de Cauchy podían dejar en F.
Dado este primer resultado de compleción a espacios de Hilbert, damos ahora algunos resultados de continuidad y ortogonalidad con importancia en nuestro análisis. Empezamos dando un par de teoremas que son piedra angular a la hora de constatar la continuidad de aplicaciones entre espacios de Banach (y por ende de Hilbert).

Teorema A.0.2 (Teorema de la Aplicación Abierta)**.** *Sean E, F dos espacios de Banach y sea $A : E \to F$ una aplicación lineal continua y sobreyectiva. Entonces, $A(G)$ es abierto en F para todo G abierto en E.*

Teorema A.0.3 (Teorema del Grafo Cerrado)**.** *Sean E, F espacios de Banach y $A : E \to F$ una aplicación lineal tal que el grafo de A, dado por*

$$gra(A) := \{(x, A(x)) : x \in E\} \subset E \times F, \tag{A.0.1}$$

es cerrado en $E \times F$. Entonces, A es continua en E.

Damos ahora un resultado de mucha utilidad práctica que es consecuencia del teorema del grafo cerrado.

Proposición A.0.4. *Si E y F son espacios de Banach y $A : E \to F$ es una aplicación lineal, entonces las afirmaciones siguientes son equivalentes:*

1. *$gra(A)$ es cerrado.*
2. *Si $(x_n)_n \subset E$ es una sucesión tal que $\lim_{n\to\infty} x_n = 0$ en E y $\lim_{n\to\infty} A(x_n) = y$ en F, entonces necesariamente $y = 0$.*
3. *A es continua.*

Tratemos ahora algunas de las nociones básicas de ortogonalidad en espacios de Hilbert. Empezamos enunciando lo que, en un espacio vectorial arbitrario, se entiende por ortogonalidad, para después particularizar resultados nucleares en nuestro desarrollo. Supondremos que los espacios definidos se construyen sobre el cuerpo arbitrario $\mathbb{K}$, y que, a menos que se diga lo contrario, la norma dada será la del espacio más grande. Necesitamos algunas definiciones previas.

Definición A.0.5. Sean H es un espacio vectorial y $A \subset H$. Se dice que A es *convexo* si, para todos $x, y \in A$ y $0 \leq t \leq 1$, $tx + (1-t)y \in A$.

Esta definición no es más que decir que el conjunto A mantiene dentro de sí mismo a todos los elementos del segmento $\{tx + (1-t)y : 0 \leq t \leq 1\}$ con extremos x e y en A.

Teorema A.0.6. *Sean H un espacio con producto interno, y M un subconjunto no vacío convexo y completo de H. Sea $h \in H$ un elemento arbitrario del espacio general. Entonces, existe un único $h_0 \in M$ tal que:*

$$||h - h_0|| = \operatorname{dist}(h, M) = \inf\{||h - k|| : k \in M\}. \tag{A.0.2}$$

En nuestro caso, la aplicación del teorema A.0.6 se dará sobre H espacio de Hilbert, y con M convexo y, habitualmente, cerrado en H. Estas hipótesis implican las de dicho teorema.
De hecho, si el subconjunto M no solo es convexo, sino que es un subespacio vectorial de H, aún más puede ser dicho. Nótese que, evidentemente, un subespacio vectorial nunca será vacío (por contener al elemento neutro para la suma), por lo que no hace falta incluir esta información en las premisas de lo siguiente.

Teorema A.0.7. *Sea M un subespacio lineal cerrado de H espacio de Hilbert, sea $h \in H$, y sea $h_0 \in M$ el único elemento de M tal que* $\operatorname{dist}(h, M) = ||h_0 - h||$. *Entonces, $h - h_0 \perp M$. Recíprocamente, si $h_0 \in M$ es tal que $h - h_0 \perp M$, entonces $||h - h_0|| = \operatorname{dist}(h, M)$.*

Definimos ahora lo que es el subespacio ortogonal de un determinado subconjunto del espacio de Hilbert arbitrario H.

Definición A.0.8. Sea $A \subset H$. El conjunto $A^\perp = \{f \in H : f \perp g,\ \forall g \in A\}$ es el espacio ortogonal de A en H.

Es inmediato ver que, para todo subconjunto $A \subset H$, $A^\perp$ es un subespacio vectorial cerrado. Para ver esto último, considérese el funcional continuo $\phi_f(\cdot) := \langle \cdot, f \rangle_H : H \to \mathbb{K}$, el cual es acotado, como se puede ver al aplicar la desigualdad de Cauchy-Schwarz, luego es continuo. Por tanto, podemos escribir $A^\perp = \cap_{f\in A} \phi_f^{-1}(\{0\})$, donde $\phi_f^{-1}\{0\}$ es cerrado por ser contraimagen de un cerrado por una aplicación continua, y donde, por tanto, tenemos una intersección de cerrados, luego es cerrado.
Con el teorema anterior, y dado $M \subset H$ un subespacio vectorial cerrado de H, estamos en condiciones de definir una aplicación $P_M : H \to M$ que envíe cada elemento $h \in H$ al elemento $h_0 \in M$ enunciado en el teorema anterior. Este elemento se llama *proyección ortogonal* de h en M, y a la aplicación P_M

le daremos el nombre de *proyección ortogonal* de H en M. Queda claro, de los resultados anteriores, el porqué de esta terminología.
Para cualquier subespacio vectorial cerrado M de H, la aplicación proyección ortogonal P_M es una tranformación lineal y contractiva ($||P_M|| \leq 1$), además de idempotente[1].
Aclarado lo anterior, damos ahora resultados, menos triviales, que nos permiten caracterizar con facilidad el subespacio ortogonal de un subespacio ortogonal, o la densidad de un subespacio de H.

Corolario A.0.9. *Si M es un subespacio vectorial cerrado de un espacio de Hilbert H, entonces $(M^{\perp})^{\perp} = M$.*

Corolario A.0.10. *Si A es un subespacio vectorial de un espacio de Hilbert H, entonces A es denso en H si y solo si $A^{\perp} = \{0\}$.*

Puestos de manifiesto los resultados que, en este trabajo, más relevancia tienen en términos de continuidad y ortogonalidad en espacios de Hilbert, entramos ahora en un breve estudio sobre operadores compactos.
Empezamos definiendo lo que entendemos por operador compacto en un espacio de Banach. Claramente, se aplica de forma análoga en espacios de Hilbert.

Definición A.0.11. Sean E y F espacios de Banach. Sea $S : E \to F$ un operador acotado y lineal. Diremos que S es compacto si y solo si $\overline{SB_E}$ es compacto en la topología de F, donde B_W es la bola unidad cerrada en el espacio W. Denotaremos por $C(E, F)$ al conjunto de operadores lineales que sean compactos.

Lo que estamos diciendo con esta definición es que un operador $S : E \to F$ entre espacios de Banach es compacto si manda subconjuntos acotados de E a conjuntos relativamente compactos de F, entendiendo por relativamente compactos todos aquellos conjuntos cuya adherencia es compacta en la topología de F.
Un operador lineal y acotado $S : E \to F$ es compacto si $\dim(\mathrm{Im}S) < \infty$, esto es, si $\mathrm{Im}S$ tiene dimensión finita. No obstante, el recíproco no es cierto en general. Llamaremos *operadores de rango finito* a aquellos $S : E \to F$ lineales y acotados cuya imagen sea un espacio de dimensión finita.

Una forma equivalente de definir un operador compacto S entre espacios de Banach E y F (en particular, espacios metrizables), se da a través de la compacidad secuencial. Dado $S : E \to F$ una aplicación para la que E y F son espacios vectoriales de Banach, S será operador compacto si y solo si para toda sucesión $(f_n)_{n=1}^{\infty} \subset E$ acotada en la norma inducida por E, existe una subsucesión $(f_{n_k})_{k=1}^{\infty}$ tal que $(Af_{n_k})_{k=1}^{\infty}$ converge a un elemento $g \in F$.
Damos ahora un resultado nuclear en todo el desarrollo presentado de núcleos sobre espacios de Hilbert: *el teorema de representación de Fréchet-Riesz*. Obsérvese que, dados H un espacio de Hilbert, la aplicación $\langle \cdot, x \rangle_H \to \mathbb{K}$ para todo $x \in H$, definida por $y \in H \mapsto \langle y, x \rangle$, es un funcional lineal y acotado en todo H, esto es, un elemento del dual de H, que denotamos H'. El teorema mencionado, que ahora enunciamos, nos dice que *todos* los funcionales lineales y acotados en H son precisamente de esta forma.

Teorema A.0.12 (Teorema de representación de Fréchet-Riesz)**.** *Sea H un K-espacio de Hilbert y H' su dual topológico. Entonces, la aplicación $\mathcal{P} : H \to H'$, dada por*

$$\mathcal{P}(x) := \langle \cdot, x \rangle, \tag{A.0.3}$$

para todo $x \in H$, es isométrica y sobreyectiva. Además, si $\mathbb{K} = \mathbb{R}$, $\mathcal{P}$ es un isomorfismo isométrico.

[1]Una aplicación *idempotente* sobre un espacio de Hilbert H se define como un operador A lineal y acotado de H en H tal que $A^2 = A$. De hecho, se dice que A es una *proyección* si es idempotente y si $\ker A = (\mathrm{ran}A)^{\perp}$. Nótese que esta es una caracterización general de proyecciones entre espacios de Hilbert, en ningún momento hemos hablado de proyección ortogonal.

Como mero apunte de notación, es frecuente encontrar el abuso de notación $\langle u, x\rangle_{H',H}$ cuando sencillamente estamos evaluando en $x \in H$ el funcional $u : H \to \mathbb{K}$. De igual manera, hay otro abuso de notación incluido en este trabajo por comodidad de escritura: pese a no haber, a priori, un producto escalar definido en el dual topológico de un espacio de Hilbert, cuando nosotros aplicamos dicho teorema de representación A.0.12, denotaremos directamente por $\langle u, v\rangle_{H'}$ a los productos escalares $\langle x, y\rangle_H$ con $\mathcal{P}(x) = u$ y $\mathcal{P}(y) = v$. Note el lector que este último producto interno sí que tiene sentido. Es tan solo una cuestión de notación.
A partir del enunciado de A.0.12, pasamos a tratar lo que es el adjunto de un operador entre espacios de Hilbert.

Definición A.0.13. Sea $S : H_1 \to H_2$ una aplicación lineal y acotada entre los $\mathbb{K}$-espacios de Hilbert H_1 y H_2. Llamaremos **adjunto** de S al operador $S' : f \in H_2' \mapsto (fS) \in H_1'$, donde H_1' y H_2' son los duales topológicos de H_1 y H_2, respectivamente.
Cogemos $\mathbb{K} = \mathbb{R}$. Si $I_i : H_i \to H_i'$, para $i = 1, 2$, son los isomorfismos isométricos del teorema de representación de Fréchet-Riesz, al operador $S^* := I_1^{-1} S' I_2 : H_2 \to H_1$ le llamamos **adjunto de S en el sentido de espacios de Hilbert**, o simplemente adjunto de S, si no hay peligro de confusión.

Dado el $S : H_1 \to H_2$ del último enunciado, su adjunto en el sentido de espacios de Hilbert viene *caracterizado* por

$$\langle Sx, y\rangle_{H_2} = \langle x, S^* y\rangle_{H_1}, \qquad x \in H_1, y \in H_2. \tag{A.0.4}$$

Además, denotando $* : \mathcal{L}(H_1, H_2) \to \mathcal{L}(H_2, H_1)$ a la aplicación que manda operadores lineales y acotados de H_1 a H_2 a sus adjuntos en el sentido de Hilbert, tenemos lo siguiente.

Corolario A.0.14. *La aplicación $*$ arriba definida, dados los $\mathbb{R}$-espacios de Hilbert H_1 y H_2, es un isomorfismo isométrico que satisface, para todo $S \in \mathcal{L}(H_1, H_2)$, que $S^{**} = S$.*
Además, dados $S \in \mathcal{L}(H_1, H_2)$, y $R \in \mathcal{L}(H_2, H_3)$ se tiene que $(RS)^* = S^* R^*$.

Dada la propiedad anterior, añadimos algo de terminología:

Definición A.0.15. Sea H un espacio de Hilbert. Dado $T \in \mathcal{L}(H)$, diremos que T es *autoadjunto* si $T = T^*$, diremos que es *positivo* si $\langle Tx, x\rangle \geq 0$ para todo $x \in H$, y diremos que T es *estrictamente positivo* si es positivo y si, además, la desigualdad es estricta para todo $x \neq 0$.

Lema A.0.16. *Dados H_1, H_2 espacios de Hilbert. Sea $S \in \mathcal{L}(H_1, H_2)$. Entonces, SS^* y S^*S son autoadjuntos y positivos. Además, si S es inyectivo, entonces S^*S es estrictamente positivo y, análogamente, si S^* es inyectivo, entonces SS^* es estrictamente positivo.*

Demostración. La primera parte del lema es trivial. Pasamos directamente a suponer que S es inyectivo. Recordando que S va de H_1 a H_2 (y S^* va al revés), empezamos cogiendo un $f \in H_1$ arbitrario tal que $S^*S(f) = 0$. Entonces, por (A.0.4), tenemos que

$$0 = \langle S^*S(f), f\rangle_{H_1} = \langle S(f), S(f)\rangle_{H_2}. \tag{A.0.5}$$

Como estamos tratando con productos escalares, la igualdad $\langle S(f), S(f)\rangle_{H_2} = 0$ implica que $S(f) = 0$. Como S es inyectivo por hipótesis, entonces $f = 0$, lo cual demuestra que S^*S es estrictamente positivo. Para la última afirmación se razona análogamente. □

Damos a continuación un resultado imprescindible para la caracterización de los RKHSs con núcleo de cuadrado integrable.

Proposición A.0.17. *Dados H_1, H_2 dos espacios de Hilbert, y $f : H_1 \to H_2$ un aplicación lineal, entonces f es inyectiva si y solo su adjunto $f^* : H_2' \to H_1'$ tiene una imagen densa en la topología débil de H_1'.*

Demostración. Supongamos primero que f^* tiene imagen densa en H_1'. Tomamos $x \in H_1$ tal que $f(x) = 0$, y queremos ver que $x = 0$, lo que probaría la inyectividad de f dada su linealidad. Como $f(x) = 0$, entonces

$$\langle g, f(x)\rangle_{H_2', H_2} = 0, \ \ \forall g \in H_2', \tag{A.0.6}$$

con lo que, tomando el adjunto,

$$\langle f^*(g), x\rangle_{H_1', H_1} = 0, \tag{A.0.7}$$

luego por la densidad supuesta de $\mathrm{Im}(f^*)$ en H_1' en la topología débil en H_1', tenemos que $\langle h, x\rangle_{H_1', H_1} = 0$, para todo $h \in H_1'$, lo que prueba que $x = 0$, y tenemos la inyectividad de f.
Veamos el recíproco. Razonamos por reducción al absurdo. Supongamos que f es inyectiva y que $f^*(H_2')$ no es densa en H_1'. Por un resultado bien conocido que es consecuencia del teorema de Hahn-Banach, esta no densidad implica que existe un elemento $x \in H_1$ tal que $\langle f^*(g), x\rangle_{H_1', H_1} = 0$, para todo $g \in H_2'$. Entonces, tomando el adjunto, $\langle g, f(x)\rangle_{H_2', H_2} = 0$, de nuevo para todo $g \in H_2'$. Luego $f(x) = 0$, con lo que f no puede ser inyectiva. Y hemos llegado al absurdo. □

Hecha la exploración por la compacidad y la representación en clave de aplicaciones entre espacios de Hilbert, pasamos al último bloque de este apéndice. Tras lecturas del Conway [5], y del Schwartz-Dunford [7], pasamos a enunciar y probar algunos de los resultados más importantes de las topologías débiles en los espacios de Banach, así como la teoría de reflexividad de espacios de Banach. Denotaremos por E' al dual topológico, objeto de nuestro actual estudio, de un espacio vectorial topológico (e.v.t.) arbitrario E, sobre un cuerpo $\mathbb{K}$, generalmente $\mathbb{R}$ o $\mathbb{C}$. Por supuesto, toda esta teoría se particulariza para espacios de Banach en el cuerpo central de este trabajo. De hecho, gran parte de este apartado se desarrollará, directamente, con espacios de Banach.

Definición A.0.18. Si E es un e.v.t. localmente convexo, la **topología débil** en E, que denotaremos por $\sigma(E, E')$, es la topología generada por la familia de seminormas $\mathcal{S} = \{|f| : f \in E'\}$.

En adelante, nos centramos en el caso de espacios de Banach. A raíz de esta última definición, nos referiremos a los distintos objetos topológicos en el marco de esta topología con el calificativo débil (por ejemplo, un cerrado en la topología $\sigma(E, E')$ será un cerrado débil en E). Entramos a definir la propiedad de reflexividad para espacios de Banach.
Si aplicamos la construcción de la definicón A.0.18 sobre el dual topológico E', a través de la familia de seminormas en E' que sigue

$$\mathcal{U} = \{F_x \in E'' : F_x(f) = f(x) \ \ \forall f \in E'\} \subset E'', \tag{A.0.8}$$

entonces tenemos una nueva topología, ahora sobre E', a la que nos referimos como *topología $*$-débil.* Nótese que, dada la condición de convexidad local y de Hausdorff del espacio de Banach E, podemos identificar, vía el teorema de Hahn Banach, esta familia de seminormas $\mathcal{U} \subset E''$ con el propio espacio E.
Recordemos que la inmersión natural de un espacio de Banach E en su doble dual E'' (el cual es también un espacio de Banach, con tal de pedir que E sea normado) viene definida por la siguiente aplicación:

$$\begin{aligned} \gamma : \ \ & E \to E'' \\ & x \mapsto \hat{x}, \end{aligned}$$

donde $\hat{x}$ denota a la siguiente aplicación, que envía elementos de E' en $\mathbb{K}$:

$$\hat{x}(y) = y(x), \ \ y \in E'. \tag{A.0.9}$$

Esta aplicación γ es una isometría.

Definición A.0.19. Un espacio de Banach E es reflexivo si la inmersión natural de E en E'' (que más arriba hemos denotado por γ) es sobreyectiva.

Conocida la condición isométrica de γ, sabemos que γ es inyectiva. Entonces, un espacio de Banach es reflexivo si y solo si la inmersión natural γ es un isomorfismo isométrico.
A continuación, damos un teorema que nos permite caracterizar a los espacios de Banach reflexivos a través de la compacidad de su bola cerrada unidad. Antes, enunciamos el teorema de Alaoglu, en clave de teoría de espacios normados.
Necesitamos definir lo que es una red en un espacio topológico, y cómo se define la convergencia de una sucesión con respecto a ellas. Esto nos permitirá hablar con propiedad de convergencia en la demostración del teorema de Alaoglu.

Definición A.0.20. Un **sistema directo** es un conjunto I con una relación de orden $\succ$ que satisface:

1. Si $a, b \in I$, entonces existe un $c \in I$ tal que $c \succ a$ y $c \succ b$.

2. $\succ$ satisface las propiedades reflexiva, transitiva, y antisimétrica.

Una **red** en S es una aplicación que va del sistema directo I a S, y la denotamos por $(x_i)_{d\in I}$.
Diremos que una red $(x_i)_{i\in I}$ converge a $x \in S$, y lo denotamos por $x_i \to x$, si para todo entorno V de x en S existe un $b \in I$ tal que $x_i \in V$ para todo $i \succ b$.

Teorema A.0.21 (Teorema de Alaoglu)**.** *Si E es un $\mathbb{C}$-espacio vectorial normado, entonces la bola unidad cerrada $B_{E'}$ en su dual E' es $*$-débilmente compacto.*

Demostración. Para cada $x \in E$ definimos

$$D_x = \{z \in \mathbb{C} : |z| \leq ||x||\}, \tag{A.0.10}$$

que es la bola cerrada de radio $||x||$ en el cuerpo $\mathbb{K}$. Cada D_x es compacto en $\mathbb{C}$ respecto a su topología usual, luego por el teorema de Tychonoff sobre compacidad tenemos que $D = \prod_{x\in E} D_x$ es compacto en la topología producto.
Los elementos de D son sucesiones $(\mu_x)_{x\in E}$ tales que $\mu_x \in D_x$, para todo $x \in E$. De hecho, μ es una aplicación de E en D que satisface $|\mu_x| \leq ||x||$, para todo $x \in E$. Así, aunque no tengamos ninguna garantía de que μ sea lineal ni de que $D = \mathbb{C}$, utilizaremos la notación estándar $\langle x, \mu\rangle = \mu_x$. Escrito de esta forma, se tiene que $|\langle x, \mu\rangle| \leq ||x||$, para todo $x \in E$.
Si μ fuera lineal, por la acotación anterior tendríamos $||\mu|| \leq 1$, con lo que $\mu \in B_{E'}$. Es por ello por lo que $B_{E'}$ es el conjunto formado por tales funcionales lineales.
Veamos que $B_{E'}$ es cerrado con respecto a la topología producto en D. Sea $(\mu_i)_{i\in I}$ una red de elementos de $B_{E'}$ convergente en D, de la forma $\mu_i \to \mu \in D$. Como las proyecciones canónicas son continuas en la topología producto, se tiene que

$$\langle x, \mu_i\rangle = \pi_x(\mu_i) \to \pi_x(\mu) = \langle x, \mu\rangle, \quad x \in X. \tag{A.0.11}$$

En particular, dados $x, y \in E$, $a, b \in \mathbb{C}$, tenemos que

$$\langle ax + by, \mu_i\rangle \to \langle ax + by, \mu\rangle. \tag{A.0.12}$$

Sin embargo, cada μ_i es lineal porque $\mu_i \in B_{E'}$ para todo $i \in \mathbb{N}$, luego

$$\langle ax + by, \mu_i\rangle = a\langle x, \mu_i\rangle + b\langle y, \mu_i\rangle \to a\langle x, \mu\rangle + b\langle y, \mu\rangle. \tag{A.0.13}$$

Esto prueba la linealidad de μ. Entonces, $\mu \in B_{E'}$, luego $B_{E'}$ es un subconjunto cerrado de D, y como D es compacto en la topología producto, entonces $B_{E'}$ también lo es respecto a la misma topología.

Veamos ahora que la topología producto en D restringida a $B_{E'}$ está contenida en la topología $*$-débil en E restringida a $B_{E'}$. Fijamos $x \in E$, y tomamos $(\mu_i)_{i\in I}$ una red en $B_{E'}$ tal que $\mu_i \to \mu$ $*$-débilmente. Entonces, en particular,

$$\pi_x(\mu_i) = \langle x, \mu_i\rangle \to \langle x, \mu\rangle = \pi_x(\mu), \tag{A.0.14}$$

luego la proyección canónica π_x es continua con respecto a la topología $*$-débil restringida a $B_{E'}$. Entonces, como γ es la topología menos fina que hace continua a todas las proyecciones canónicas, se tiene $\gamma \subset \sigma$.
Ya hemos probado que $B_{E'}$ es compacto en γ, luego también lo será en σ. Así, $B_{E'}$ es compacto respecto a la topología $*$-débil en E'. □

Teorema A.0.22. *Sea E un espacio de Banach. Entonces, E es reflexivo si y solo si su bola cerrada unidad es compacta en la topología débil de E.*

Demostración. Aunque hemos enunciado la equivalencia por mostrar el resultado completo, solo probamos la implicación que usamos en este trabajo: si E es una espacio de Banach reflexivo, entonces su bola unidad cerrada B_H es débilmente compacta. Por el teorema de Alaoglu A.0.21, la bola cerrada $B_{E''}$ es $*$-débilmente compacto en E'. Por la reflexividad de E, $E \equiv E''$, luego B_E es débilmente compacto en E. □

Apéndice B

Resultados adicionales de Análisis Real

Los elementos de los RKHSs H involucrados en nuestro desarollo son funciones definidas sobre conjuntos arbitrarios X sobre los que, para hablar de mediblidad o integrabilidad, necesitamos una medida. Por ello, damos unas nociones básicas de medida en un conjunto X y de la integrabilidad para una medida arbitraria. Como siempre, empezamos dando las pertinentes definiciones.

Definición B.0.1. Sea X un conjunto no vacío, y sea $\mathcal{F}$ una σ-álgebra de X. Diremos que la aplicación $\mu : \mathcal{F} \to [0, \infty]$ es **una medida en $\mathcal{F}$** si satisface:

1. $\mu(A) \in [0, \infty]$, para todo $A \in \mathcal{F}$.
2. $\mu(\emptyset) = 0$.
3. Si $(A_n)_{n=1}^{\infty}$ es una familia de elementos disjuntos de $\mathcal{F}$ y si $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$, entonces

$$\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n). \tag{B.0.1}$$

Diremos que la medida μ es *finita* si $\mu(X) < \infty$, *infinita* si $\mu(X) = \infty$, y *σ-finita* si $X = A_1 \cup A_2 \cup \ldots$ es unión finita o infinita numerable de elementos de $\mathcal{F}$ tales que $\mu(A_k) < \infty \ \forall k \in \mathbb{N}$.

Se suele denotar por $(X, \mathcal{F}, \mu)$ al espacio de medida, aunque, si no hay peligro de confusión, diremos simplemente que X es un espacio de medida con la medida μ. Además, a medibilidad de una aplicación $f : X \to Y$, con X e Y espacios de medida cualesquiera, se define a través de la pertenenecia en el σ-álgebra de partida de las imágenes inversas de los elementos de la σ-álgebra de llegada.
Dado X un conjunto arbitrario no vacío, y $\mathcal{F}$, denotamos por I_A para todo $A \in \mathcal{F}$ a la **función indicatriz o característica** del conjunto A de la σ-álgebra $\mathcal{F}$, definido como:

$$I_A(x) = \begin{cases} 1 \quad si \qquad x \in A \\ \\ 0 \qquad \text{en otro caso.} \end{cases} \tag{B.0.2}$$

Extendiendo esta idea, llamaremos *función simple* a toda aplicación $f : X \to \mathbb{K}$ para la que existen $a_1, \ldots, a_n \in \mathbb{K}$ y $A_1, \ldots, A_n \in \mathcal{F}$ tales que

$$f(x) = \sum_{i=1}^{n} a_i I_{A_i}(x), \qquad x \in X, \tag{B.0.3}$$

esto es, tal que $\mathrm{Im}(f)$ es un conjunto finito. Es claro que una función simple es medible siempre.
Con lo dado, si X es un espacio de medida con medida μ sobre la σ-álgebra $\mathcal{F}$, y dada la función

medible y positiva $f : X \to [0, \infty)$, definimos la integral de f en X, con respecto a la medida μ, como sigue:

$$\int f d\mu = \sup \sum_i [\inf_{x \in A_i} f(x)] \mu(A_i), \tag{B.0.4}$$

donde el superior se da sobre el conjunto de descomposiciones finitas $(A_i) \subset \mathcal{F}$ del conjunto X. Esta idea es fácilmente generalizable a funciones medibles y no positivas a través de la escritura de cualquier función $f : X \to \mathbb{R}$ como $f = f^+ - f^-$, siendo f^+ y f^- las partes positiva y negativa de f. También, se puede generalizar esta noción de integrabilidad a funciones medibles con valores complejos $f : X \to \mathbb{C}$, sencillamente a través de la restricción de nuestro estudio a su parte real y a su parte imaginaria.
A fuerza de uso, dada la teoría de funciones riesgo hecha en este trabajo, con una estructura de espacio de medida producto subyacente, nos vemos obligados a presentar algunas de las ideas básicas de este tipo de espacios.
Recordemos cómo se construía un producto finito de espacio medibles, restringiéndonos al caso de medidas σ-finitas de partida. No es difícil generalizar estos resultados a productos infinitos, pero, siendo prácticos, a nosotros solo nos interesan los productos finitos.

Definición B.0.2. Sean[1] $(\Omega_1, \mathcal{A}_1), \ldots, (\Omega_n, \mathcal{A}_n)$, para algún $n \in \mathbb{N}$. Diremos que una σ-álgebra $\mathcal{A}$ sobre el producto $\Omega = \Omega_1 \times \cdots \times \Omega_n$ es la **σ-álgebra producto** si se verifica:

1. Las proyecciones $\pi_i : \Omega \to \Omega_i$ son medibles.
2. Dado otro espacio medible $(\Omega', \mathcal{A}')$, toda aplicación dada por

$$F : \Omega' \to \Omega, \tag{B.0.5}$$

   es medible si y solo si sus componentes $F_i = \pi_i \circ F$ son medibles.

De nada sirve esta definición si no probamos que realmente existe una σ-álgebra de este tipo. De hecho, veamos también que es única.

Proposición B.0.3. *La σ-álgebra producto existe y es única. Además, es la generada por los produtos $A_1 \times \cdots \times A_n$, con $A_i \in \mathcal{A}_i$, para todo $i \in \{1, \ldots, n\}$.*

Demostración. Denotamos por $\mathcal{G}$ al espacio de estos productos de elementos de las σ-álgebras A_i, $i = 1, \ldots, n$; y por $\sigma(\mathcal{G})$ a la σ-álgebra generada por $\mathcal{G}$.
Probaremos que $\sigma(\mathcal{G})$ cumple las dos condiciones, lo que demostraría la primera y la última afirmación, a falta de ver la unicidad.

1. Si $A_i \in \mathcal{A}_i$, entonces $\pi_i^{-1}(A_i) = \Omega_1 \times \cdots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \cdots \times \Omega_n$, lo cual es un elemento de $\mathcal{G}$ por definición, luego también lo es de $\sigma(\mathcal{G})$.
2. Si F es medible, $F_i = \pi_i \circ F$ es composición de medibles, con lo que F_i lo es. Recíprocamente, si suponemos que las F_i son medibles y que $A_i \in \mathcal{A}_i$ para cada $i \in \{1, \ldots, n\}$, entonces $F_i^{-1}(A_i) \in \mathcal{A}'$, y:

$$F^{-1}(A_1 \times \cdots \times A_n) = \cap_{i=1}^n F_i^{-1}(A_i), \tag{B.0.6}$$

   lo cual está en $\mathcal{A}'$ por ser intersección numerable de elementos de $\mathcal{A}'$. Por tanto, F es medible, y con esto hemos probado que $\sigma(\mathcal{G})$ satisface las dos propiedades del enunciado.

[1] La notación usada para los espacios de medida es la estándar, i.e., escribimos $(\Omega, \mathcal{A}, \mu)$ para referirnos a un espacio de medida sobre un conjunto Ω, con una σ-álgebra $\mathcal{A}$, y con una medida μ. Cuando no hay duda sobre la medida, simplemente escribiremos $(\Omega, \mathcal{A})$.

Veamos ahora la unicidad. Supongamos que existen dos σ-álgebras sobre Ω, $\mathcal{A}$ y $\mathcal{A}'$, con las propiedades enunciadas. Sobre la segunda condición del enunciado para las σ-álgebra producto (finito), tomemos las aplicaciones $\text{id} : (\Omega, \mathcal{A}') \to (\Omega, \mathcal{A})$, y $\text{id}^{-1} = \overline{\text{id}} : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A}')$, las cuales son medibles ya que, usando la segunda afirmación, para todo $i \in \{1, \dots, n\}$, $\text{id}_i : \Omega \to \Omega_i$ y $\overline{\text{id}}_i : \Omega \to \Omega_i$ son medibles. De esto se sigue, por la propia definición de medibilidad en las aplicaciones id y $\overline{\text{id}}$, que $\mathcal{A} = \mathcal{A}'$, como queríamos. $\square$

A estos productos de elementos de cada una de las σ-álgebras involucradas se les llama **rectángulos medibles** del producto. Otra forma de expresar lo anterior es que la σ-álgebra producto sobre el espacio $\Omega_1 \times \cdots \times \Omega_n$ es la menor σ-álgebra en $\Omega_1 \times \cdots \times \Omega_n$ que contiene a todos los rectángulos medibles.
En nuestro desarrollo, particularizamos para el caso $n = 2$. En este, conocidas las σ-álgebras subyacentes, presentamos el resultado que formaliza el tipo de medida que manejamos en esta situación. Si (X, σ_1, μ_1) y (Y, σ_2, μ_2) son los dos espacios de medida correspondientes, denotaremos por $\mu_1 \otimes \mu_2$ a la medida definida como

$$(\mu_1 \otimes \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2), \tag{B.0.7}$$

para todo $A_1 \in \sigma_1$ y $A_2 \in \sigma_2$. La compleción de esta medida es precisamente la medida y el espacio que manejaremos. No entramos en más detalles por la extensión que ello supondría, pero ya tenemos claro en qué tipo de contexto de medida trabajaremos a la hora de hablar de funciones de pérdida y de riesgo. Aparcamos la construcción del espacio de medida producto, y pasamos a presentar algunos resultados sobre la integración de funciones con valores en un espacio de Banach.
Extendemos la expresión (B.0.3) en un espacio de medida $(X, \mathcal{F})$ con funciones $f : X \to E$, donde E es ahora un espacio de Banach arbitrario. Para ello, basta tomar un número finito de elementos de E de la forma $x_1, \dots, x_n \in E$ y $A_1, \dots, A_n \in \mathcal{F}$ de tal forma que llamaremos *función simple con valores en* E (o función simple si no hay peligro de confusión) a una función de la forma:

$$f = \sum_{i=1}^{n} x_i I_{A_i}. \tag{B.0.8}$$

No es complicado probar (es más un problema de notación que de ingenio) que estas funciones simples no dependen de la descomposición $(A_n) \subset \mathcal{F}$ elegida.
Dicho esto, definimos la medibilidad de una función en X con valores en un espacio de Banach E como sigue:

Definición B.0.4. La función $f : X \to E$ es una **función medible con valores en E** si y solo si existe una sucesión $(f_n)_n$ de funciones simples con valores en E $f_n : X \to E$ tal que $\lim_{n\to\infty} ||f_n(x) - f(x)||_E = 0$ para todo $x \in X$.

El siguiente lema relaciona la caracterización anterior de mediblidad con llegada en E con la medibildad estándar:

Lema B.0.5. *Sea E un espacio de Banach, y $(X, \mathcal{F})$ un espacio de medida. Sea $f : X \to E$. Entonces, las siguientes afirmaciones son equivalentes:*

1. *f es una función medible con valores en E.*

2. *$f(X)$ es separable y $f^{-1}(B)$ es medible para todo conjunto de Borel $B \subset E$ (esto es, para todo abierto en la topología estándar definida en E a partir de la métrica inducida por su norma).*

En casi todas las situaciones que nosotros tratamos, el espacio E (que particularizamos habitualmente en espacio de Hilbert) es separable. En tales casos, las dos afirmaciones recogidas en el lema B.0.5 son equivalentes sin necesidad de imponer que $f(\Omega)$ sea separable.

No es difícil ver que la medibilidad con llegada en un espacio de Banach E se mantiene a través de operaciones estándar, como sumas, productos, o límites. También, se mantiene a través de *determinadas* composiciones: si $f : \Omega \to E$ es una función medible con llegada en E, y $S : E \to F$ es una aplicación lineal y acotada entre espacios de Banach, entonces $S \circ f : \Omega \to F$ es medible con llegada en el espacio de Banach F. En particular a lo anterior, para todo $x' \in E'$ las funciones $\langle x', f(\cdot)\rangle : \Omega \to \mathbb{R}$[2] son medibles con llegada en el espacio de Banach $\mathbb{R}$, lo que equivale a la medibilidad estándar, dado el carácter separable de $\mathbb{R}$. El siguiente resultado nos da también el resultado recíproco.

Teorema B.0.6 (Teorema de medilidad de Petty)**.** *Sea E un espacio de Banach y sea $(\Omega, \mathcal{F})$ un espacio de medida, con medida μ. Entonces, $f : \Omega \to E$ es una función medible con valores en E si y solo si las siguientes condiciones se satisfacen:*

1. *f es débilmente medible, i.e., $\langle x', f(\cdot)\rangle : \Omega \to \mathbb{R}$ es medible para todo $x' \in E'$.*

2. *$f(\Omega)$ es un subconjunto separable de E.*

Ilustremos el teorema de medibilidad de Petty a través del siguiente ejemplo. Sea (X, d) un espacio métrico separable, equipado con la σ-álgebra de Borel dada por la topología inducida por su métrica. Sea $f : X \to E$ una función continua. Es claro que f es débilmente medible (tal y como lo hemos enunciado en el teorema B.0.6). Además, como X es separable y f es continua, tenemos $f(X)$ es separable en E. Para probar esto, basta usar ideas elementales de la continuidad de aplicaciones generales entre espacios topológicos. Una de las posibles caracterizaciones de esta continuidad era la de que, si $g : Y \to Z$ es la aplicación continua entre espacios topológicos, se tiene, a modo de equivalencia, que $f(\overline{A}) \subset \overline{f(A)}$, para todo $A \subset Y$, donde con $\overline{A}$ denotamos a la adherencia de A.
Así pues, en el caso concreto que estamos tratando, el razonamiento es como sigue: sea $A \subset X$ el subconjutno denso y numerable que sabemos que existe por la separabilidad de X. Entonces, claramente $f(A)$ sigue siendo numerable, y por lo caracterización que acabamos de dar de continuidad, $E = f(X) = f(\overline{A}) \subset \overline{f(A)}$, luego $f(A)$ es también denso en E. Por tanto, $f(X)$ es separable, y como ya vimos también que f es débilmente medible, entonces, por el teorema B.0.6, f es una función medible con valores en E.
En cuanto a la integrabilidad con respecto a una medida arbitraria, empezamos dando la noción de esta para funciones simples como en B.0.8. Dada $f : X \to E$ una función simple como en B.0.8, y dada μ una medida σ-finita en X, definimos la *integral de f* como:

$$\int_X f := \sum_{i=1}^{n} \mu(A_i) x_i, \tag{B.0.9}$$

donde esta expresión es independiente de la representación $(A_n)_n$ elegida en $\mathcal{F}$.

Con las nociones básicas dadas para la medibilidad de funciones con llegada en espacio de Banach arbitrarios, pasamos a tratar un concepto esencial en nuestro estudio de núcleos y RKHS: *la integrabilidad de Bochner*. Este concepto es una buena adaptación de las teorías de integrabilidad más conocidas a la teoría de espacios de Banach, con su propia versión del teorema de la convergencia dominada, por ejemplo.

Definición B.0.7. Sea E un espacio de Banach y $(X, \mathcal{F}, \mu)$ un espacio de medida, con μ una medida σ-finita. A una función $f : X \to E$ se le llama **μ-integrable Bochner** (simplemente intebrable Bochner cuando no haya peligro de confusión) si existe una sucesión (f_n) de funciones simples en X con valores en E tal que

[2] Notemos el abuso de notación sobre esta notación de producto escalar, tal y como ya hemos venido aclarando anteriormente. Sencillamente, estamos evaluando los funcionales $x' \in E'$ en las imágenes por f de los elementos de Ω.

$$\lim_{n\to\infty}\int_X ||f_n - f||_E d\mu = 0. \tag{B.0.10}$$

En tal caso, el límite dado por

$$\int_X f d\mu := \lim_{n\to} f_n d\mu \tag{B.0.11}$$

existe y es llamado **la integral de Bochner de f**. Como mero comentario de notación, en el caso en el que el espacio de medida sea un espacio probabilístico, se suele denotar a este límite con $\mathcal{E}_\mu f$.

De la linealidad de los límites involucrados en el espacio de funciones $f : X \to E$ se sigue la linealidad de la integral de Bochner. Además, la integrabilidad Bochner equivale a la integrabilidad de la función $\phi : X \to [0,\infty]$, con $\phi(x) := ||f(x)||_E$, para un $f : X \to E$ arbitrariamente fijado, como vemos a continuación.

Proposición B.0.8. *Sea $f : X \to E$ una función con valores en el espacio de Banach E. Entonces, f será μ-integrable si y solo si $x \to ||f(x)||_E$ es μ-integrable Bochner. Además, en el caso de tener tal integrabilidad, dado F otro espacio de Banach, y $S : E \to F$ una aplicación lineal y acotada, entonces la integral de Bochner conmuta con S, esto es:*

$$S(\int_X f d\mu) = \int_X S f d\mu. \tag{B.0.12}$$

A continuación, damos algunas nociones básicas de la teoría de la transformada de Fourier, con el objetivo bien marcado en el teorema de Plancherel, de uso común en el capítulo segundo (véase, por ejemplo, la demostración del teorema de Paley-Wiener). En el desarrollo que presentamos aquí, empezamos aclarando la notación que usaremos, de cara a simplificar al máximo las expresiones involucradas:

1. La medida que usaremos en todo momento será la *medida de Lebesgue* en $\mathbb{R}^n$. Así, los espacios habituales construidos con esta medida son los espacios de Lebesgue L^p, o $L^p(\mathbb{R}^n)$, que son espacios normados dados por la siguiente norma:

$$||f||_p = \left\{\int_{\mathbb{R}^n} |f|^p dx\right\}^{\frac{1}{p}}, \qquad \text{con} \quad 1 \le p < \infty. \tag{B.0.13}$$

2. Dado un $t \in \mathbb{R}^n$ arbitrario, llamaremos *carácter* e_t a la función definida en $\mathbb{R}^n$ por:

$$e_t(x) = e^{it\cdot x} = exp(i(t_1x_1 + \cdots + t_nx_n)), \qquad x \in \mathbb{R}^n. \tag{B.0.14}$$

   Nótese que e_t es un homomorfismo del grupo aditivo $\mathbb{R}^n$ sobre el grupo multiplicativo de los números complejos que tienen módulo la unidad.

3. Definimos la convolución de dos funciones f y g en $\mathbb{R}^n$ medibles con respecto a la medida de Lebesgue como:

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dm_n(y), \tag{B.0.15}$$

   cuando esta integral exista.

4. Sea α un multiíndice, esto es, un elemento de $\mathbb{N}_0^n$, para algún $n \in \mathbb{N}$. Denotamos:

$$D_\alpha = (i)^{-|\alpha|}D^\alpha = (\frac{1}{i}\frac{\partial}{\partial x_1})^{\alpha_1}\cdots(\frac{1}{i}\frac{\partial}{\partial x_n})^{\alpha_n}, \tag{B.0.16}$$

   donde $|\alpha|$ es la longitud del multiíndice, y donde D^α denota el operador derivada parcial ordinario. Es claro, de lo ya dicho, que $D_\alpha e_t = t^\alpha e_t$, donde $t^\alpha = t_1^{\alpha_1}\cdots t_n^{\alpha_n}$.

Definición B.0.9 (Versión $L^1(\mathbb{R}^n)$.)**.** La **transformada de Fourier** de una función $f \in L^1(\mathbb{R}^n)$ es la función $\hat{f}$ dada por:

$$\hat{f}(t) = \int_{\mathbb{R}^n} f(x)e_{-t}(x)dx. \tag{B.0.17}$$

Es a la aplicación que lleva de f a $\hat{f}$ a la que llamaremos *tranformada de Fourier*. Denotaremos a esta indistintamente por $\hat{f}$ o por $\mathcal{F}(f)$.

Damos ahora la definición de un espacio indispensable en cualquier desarrollo de la teoría de transformadas de Fourier: las funciones de decrecimiento rápido.

Definición B.0.10. El **espacio de funciones de decrecimiento rápido o espacio de Schwartz en** $\mathbb{R}^n$, que denotaremos por $\mathcal{S}_n$ es el espacio de funciones $f \in \mathcal{C}^\infty(\mathbb{R}^n)$ que verifican:

$$\sup_{|\alpha|\leq N} \sup_{x\in\mathbb{R}^n} (1 + |x|^2)^N |D_\alpha f(x)| < \infty, \tag{B.0.18}$$

para todo $N = 0, 1, 2, \ldots$, y donde estamos usando la norma inducida a partir del producto escalar usual en $\mathbb{R}^n$.

Es claro que $\mathcal{S}_n$ tiene estructura vectorial. Obsérvese también que las expresiones dadas por (B.0.18) son seminormas en $\mathcal{S}_n$, con lo que la familia numerable de seminormas dadas por (B.0.18) definen en $\mathcal{S}_n$ una topología localmente convexa y metrizable.
Recordemos que el espacio de funciones infinitamente diferenciables con soporte compacto, que denotaremos por $\mathcal{D}(\mathbb{R}^n)$, es un subespacio vectorial de $\mathcal{S}_n$.
Recordemos también que un *espacio de Fréchet* es un espacio vectorial topológico[3] que es localmente convexo, metrizable, y secuencialmente completo.

Teorema B.0.11. *(a) $\mathcal{S}_n$ es un espacio de Fréchet.*

(b) Sean $g \in \mathcal{S}_n$, P un polinomio y $\alpha \in \mathbb{N}_0^n$ un multiíndice. Entonces, las aplicaciones

$$f \mapsto Pf, \qquad f \mapsto gf, \qquad f \mapsto D_\alpha f \tag{B.0.19}$$

son transformaciones lineales y continuas de $\mathcal{S}_n$ en $\mathcal{S}_n$.

(c) Sean P un polinomio y $f \in \mathcal{S}_n$. Entonces:

$$(P(D)f)^\wedge = P\hat{f}, \qquad \text{y} \qquad (Pf)^\wedge = P(-D)\hat{f}. \tag{B.0.20}$$

(d) La transformación de Fourier es una aplicación continua y lineal de $\mathcal{S}_n$ en $\mathcal{S}_n$[4]*.*

Teorema B.0.12 (El teorema de inversión)**.** *1. Sea $g \in S_n$. Entonces:*

$$g(x) = \int_{\mathbb{R}^n} \hat{g}(t)e_x(t)dt, \qquad x \in \mathbb{R}^n. \tag{B.0.21}$$

2. La transformada de Fourier es una aplicación continua, lineal, inyectiva de S_n en S_n, de período 4, cuya inversa es también continua.

[3]En el desarrollo aquí llevado, hablamos de espacios vectoriales topológicos (i.e., espacios topológicos X en que la suma y el producto por escalares en el cuerpo $\mathbb{K}$ son aplicaciones continuas de $X \times X$ y de $\mathbb{K} \times X$, respectivamente, en X), aunque en nuestra teoría no llegamos a generalizar más que al nivel de espacios de Hilbert, e incluso de Banach en algún momento.

[4]Mejoraremos la parte (d) en el teorema B.0.11.

3. Si $f \in L^1(\mathbb{R}^n)$, $\hat{f} \in L^1(\mathbb{R}^n)$, *y*

$$f_0(x) = \int_{\mathbb{R}^n} \hat{f}(t) e_x(t) dt, \qquad x \in \mathbb{R}^n, \tag{B.0.22}$$

entonces $f(x) = f_0(x)$ *para casi todo* $x \in \mathbb{R}^n$.

Concluimos este apartado con el teorema que sigue.

Teorema B.0.13 (Teorema de Plancherel)**.** *Existe un isomorfismo isométrico y lineal* ψ *de* $L^2(\mathbb{R}^n)$ *en* $L^2(\mathbb{R}^n)$ *unívocamente determinada por la condición:*

$$\psi f = \hat{f}, \qquad \forall f \in L^2(\mathbb{R}^n). \tag{B.0.23}$$

Nótese que la igualdad $\psi f = \hat{f}$ se extiende de S_n a $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, ya que S_n es denso en $L^1(\mathbb{R}^n)$ y en $L^2(\mathbb{R}^n)$. Esto nos da la consistencia deseada: el dominio de ψ es $L^2(\mathbb{R}^n)$, luego junto a la definición dada para $\hat{f}$ sobre $L^1(\mathbb{R}^n)$, tenemos que $\psi f = \hat{f}$ para cualquiera de las dos definiciones. Habitualmente, se extiende la aplicación ψ de $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ a $L^2(\mathbb{R}^n)$. Esta extensión es frecuentemente conocida también como transformada de Fourier (o de *Fourier-Plancherel* para evitar confusiones), y la notación $\hat{f}$ seguirá siendo usada en lugar de ψf, para cualquier $f \in L^2(\mathbb{R}^n)$.

Bibliografía

[1] Alain Berlinet, Christine Thomas-Agnan; *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, segunda edición, Springer, 2004.

[2] Antonio G. García, *Sampling Theory and RKHSs*, Departamento de Matemáticas, Universidad Carlos III de Madrid, 2015.

[3] Frigyes Riesz, Béla Sz.-Nagy, *Functional Analysis*, segunda edición, "Dover Books on Advanced Mathematics", Dover Publications, 1990.

[4] Ingo Steinwart, Andreas Christmann; *Support Vector Machines*, primera edición, "Information Science and Statistics", Springer, 2008.

[5] John B. Conway, *A Course in Functional Analysis*, segunda edición, "Graduate Texts in Mathematics", Springer, 1990.

[6] N. Aronszajn, *Theory of Reproducing Kernels*, "Division of Engineering Science", Harvard University, 1948.

[7] N. Dunford, J. T. Schwartz; *Linear Operators. Part I: General Theory*, tercera edición, "Pure and Applied Mathematics - A Series of Texts and Monographs", Wiley-Interscience, 1988.

[8] R.M. Dudley, *Real Analysis and Probability*, segunda edición, "Cambridge Studies in Advanced Mathematics", Cambridge Univesity Press, 2004.

[9] Robert R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, segunda edición, "Lecture Notes in Mathematics", Springer-Verlag, 1993.

[10] Rubén A. Martínez-Avendaño, Peter Rosenthal; *An Introduction to Operators on the Hardy-Hilbert Space*, primera edición, "Graduate Texts in Mathematics", Springer, 2007.

[11] Saburou Saitoh, Yoshihiro Sawano; *Theory of Reproducing Kernels and Applications*, primera edición, "Developments in Mathematics", Springer, 2016.

[12] Vern I. Paulsen, *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*, primera edición, "Cambridge Studies in Advanced Mathematics", Cambridge University Press, 2016.

[13] W. Rudin, *Functional Analysis*, segunda edición, McGraw-Hill, 2009.

[14] W. Rudin, *Real and Complex Analysis*, tercera edición, McGraw-Hill, 1988.