

Eduardo Javier Águila Hernández

*Algunos Tópicos en Teoría de
Aproximación*

Some Topics on Approximation Theory

Trabajo Fin de Grado
Grado en Matemáticas
La Laguna, Julio de 2023

DIRIGIDO POR
Ruymán Cruz Barroso

Ruymán Cruz Barroso
Departamento de Análisis
matemático
Universidad de La Laguna
38200 La Laguna, Tenerife

Agradecimientos

Quiero empezar por agradecer a mi familia, a mis padres, amigos y tutor Ruymán. Ha sido una época difícil, a algunos se nos ha hecho larga, y siempre han estado con buena cara, ilusión y ánimo para ayudarme a sacar esta memoria. En especial mis padres; y Ruymán, que ha revisado una y otra vez este trabajo sin descanso, y me ha orientado en cada avance de la memoria sobre la importancia de cada resultado. A ellos, con especial cariño.

Eduardo Javier Águila Hernández
La Laguna, 30 de mayo de 2023.

Resumen · Abstract

Resumen

Esta memoria se ha estructurado en tres capítulos y un apéndice. En el primer capítulo se introducen los polinomios de Bernstein, se demuestra el célebre Teorema Aproximación de Weierstrass haciendo uso de éstos, y se establecen algunas propiedades adicionales. En el segundo capítulo, se aborda el problema de mejor aproximación en espacios normados, haciendo especial hincapié en el caso de los espacios Pre-Hilbert. Asimismo, sin mucha extensión, se analizan los polinomios ortogonales, y se resuelve el problema de Aproximación en Mínimos Cuadrados. Finalmente, en el tercer capítulo, se investiga la clausura y sus consencuencias, se analiza brevemente la completitud de las potencias y sistemas trigonométricos en $L^2[a, b]$ y se culmina con los teoremas de Müntz. Finalmente, en el Apéndice, se presenta un código Matlab que computa polinomios de Bernstein y se demuestran las Desigualdades Young, Hölder y Minkowski.

Palabras clave: *Polinomios de Bernstein – Teorema de Aproximación de Weierstrass – Aproximación en espacios normados – Teoría de la clausura y la completitud.*

Abstract

This memory has been structured into three chapters and an appendix. In the first chapter the Bernstein polynomials are introduced, the famous Weierstrass Approximation Theorem is proved using them, and some additional properties are established. In the second chapter, the problem of best approximation in normed spaces is addressed, with special emphasis on the case of Pre-Hilbert spaces. Likewise, without much extension, orthogonal polynomials are analyzed, and the Least Squares Approximation problem is solved. Finally, in the third chapter, the closure and its consequences are investigated, the completeness of trigonometric powers and systems in $L^2[a, b]$ is briefly analyzed, and it ends with Müntz's theorems. Finally, in the Appendix, a Matlab code that computes Bernstein polynomials is presented and the Young, Hölder and Minkowski Inequalities are proved.

Keywords: *Bernstein polynomials – Weierstrass Approximation Theorem – Approximation in normed spaces – Theory of closure and completeness.*

Contenido

Agradecimientos	1
Resumen/Abstract	3
Introducción	7
1. Polinomios de Bernstein y Teorema de Aproximación de Weierstrass	9
1.1. Interpolación Polinómica	10
1.2. Polinomios de Bernstein	14
1.3. Demostración del Teorema de Aproximación de Weierstrass	17
1.4. Propiedades adicionales sobre los Polinomios de Bernstein	18
2. El problema de mejor aproximación	23
2.1. ¿Qué entendemos por mejor aproximación?	23
2.2. Aproximación en espacios normados	24
2.3. Aproximación por mínimos cuadrados	31
3. Teoría de la clausura y la completitud	39
3.1. Teoría de la clausura y la completitud	39
3.2. Completitud de las potencias y sistemas trigonométricos en $L^2[a, b]$	45
3.3. Teorema de la Clausura de Müntz	48
A. Apéndice	53
A.1. Programación en Matlab: computación de polinomios de Bernstein	53
A.2. Desigualdades de Young, Hölder y Minkowski	54
Bibliografía	57
Poster	59

Introducción

Esta memoria trata sobre algunos tópicos de la Teoría de Aproximación. Como fuente principal se usó el libro [4, *Interpolation and Approximation*] de Philip J. Davis. Se hace en general un barrido de este, pero nos hemos centrado especialmente de los Temas 6 a 11.

El objetivo del primer capítulo es demostrar el Teorema de Aproximación de Weierstrass. Se conocen varias demostraciones del mismo, pero nosotros nos centraremos en la demostración de Sergei N. Bernstein, haciendo uso de los polinomios que llevan su nombre. El primer capítulo comienza con una sección donde se aborda la interpolación polinómica. En la segunda sección se introducen los polinomios de Bernstein y se establecen maneras alternativas para su computación. En la tercera sección se demuestra el mencionado Teorema de Aproximación de Weierstrass, y en la última sección se prueban algunas propiedades adicionales que verifican los polinomios de Bernstein. En particular, la convergencia uniforme a las sucesivas derivadas, bajo ciertas condiciones, que los aproximantes respeten la convexidad de la función, y se prueba el Teorema de Voronovski.

En la primera sección del segundo capítulo se motiva el problema de mejor aproximación. En la segunda sección se analiza el problema de mejor aproximación en espacios normados, es decir, partiendo de una estructura matemática abstracta que describa propiedades sobre la mejor aproximación, independientemente del criterio específico de cercanía de aproximación. Se han incluido numerosos ejemplos de los resultados obtenidos. Al final de la misma, se pueden destacar el Teorema de Tonelli y el Teorema de Equioscilación de Chebyshev. En la última sección se analiza el problema de aproximación, esta vez en espacios Pre-Hilbert o Hilbert. Se estudian diversos resultados como las Desigualdades de Cauchy-Schwarz, Hölder, Minkowski o la Ley del Paralelogramo, entre otros. Se abordan algunas propiedades básicas sobre polinomios ortogonales. Se continúa demostrando el problema de aproximación en Mínimos Cuadrados, deduciendo las Ecuaciones Normales, donde la Matriz de Gram juega un papel importante, y por último, una vista breve a propiedades sobre los desarrollos de Fourier.

En el último capítulo se introduce la teoría de clausura y completitud, con respectivos ejemplos y resultados, entre los que se destacan el Teorema 3.1.1, que establece condiciones de equivalencia para que un sistema ortonormal en un espacio pre-Hilbert sea cerrado, o el Teorema de Riesz. A continuación, se introducen algunos resultados de Análisis Funcional, como el Teorema de Banach, que se hacen necesarias para la siguiente sección, donde se hace un pequeño énfasis en la completitud de potencias y sistemas trigonométricos en $L^2[a, b]$. Finalmente, en la tercera sección, se exponen los dos Teoremas de Müntz. Se culmina la Memoria con el apéndice, que está compuesto por las demostraciones de la Desigualdad de Young, Hölder y Minkowski, y un código en lenguaje de programación de *Matlab* usado para ilustrar numéricamente la lenta convergencia de estos aproximantes (Figura 1.1).

Polinomios de Bernstein y Teorema de Aproximación de Weierstrass

Comenzaremos este capítulo abordando uno de los resultados más importantes en *Teoría de Aproximación*, el cual establece la densidad de los polinomios en el espacio de las funciones continuas definidas en un intervalo compacto $[a, b]$ con respecto a la norma uniforme. El resultado, conocido como *Teorema de Aproximación de Weierstrass*, tiene numerosas repercusiones en el *Análisis Matemático* y especialmente en *Matemática Aplicada* (pues, entre muchas de sus propiedades, las funciones polinómicas son las más fáciles de evaluar).

Teorema 1.0.1 *Sea $f \in \mathcal{C}([a, b])$. Para todo $\epsilon > 0$ existe un polinomio p_n (de grado suficientemente grande) tal que*

$$\|f - p_n\|_\infty = \max_{x \in [a, b]} |f(x) - p_n(x)| \leq \epsilon.$$

Este teorema fue originalmente enunciado y demostrado por *Karl Weierstrass* en 1885 a la edad de 70 años, véase [8]. Desde entonces han habido numerosas generalizaciones del mismo, cabe destacar la establecida por *Marshall Harvey Stone* en 1937, véase [7]. Su generalización se basa en que la razón de dicha densidad está en las propiedades algebraicas de los polinomios, en las propiedades analíticas de las funciones continuas, y en la propiedad topológica de la compacidad de $[a, b]$. Una demostración alternativa a la dada por Weierstrass fue dada veinte años después por el matemático húngaro *L. Fèjér*, a la edad de 19 años, empleando *Series de Fourier*. También existe otra demostración realmente bella que puede deducirse usando herramientas del *Análisis Matemático* conocidas como *Aproximaciones a la Identidad*.

Supongamos que f es una función analítica en $|z| \leq R$. Entonces existe un desarrollo de potencias $\sum_{k=0}^{\infty} a_k z^k$ que converge uniformemente a f en $|z| \leq R$. Es decir, dado $\epsilon > 0$, podemos tomar una cantidad finita de términos suficientemente grande de esta serie de potencias, $p_n(z) = \sum_{k=0}^n a_k z^k \in \mathbb{P}_n$ (siendo \mathbb{P}_n espacio vectorial real de polinomios de grado a lo sumo n), tal que $|f(z) - p_n(z)| \leq \epsilon$ si $|z| \leq R$. En particular, si $x \in (-R, R)$, $|f(x) - p_n(x)| \leq \epsilon$. Las funciones que no son analíticas no admiten tal desarrollo en series de potencias. Sin embargo,

el Teorema de Weierstrass nos permite aproximar uniformemente funciones que son meramente continuas, debilitando considerablemente la hipótesis anterior. Este primer capítulo se ha estructurado en cuatro secciones. En la primera recordaremos algunas cuestiones elementales sobre interpolación polinómica, por un doble fin. Por un lado, por la propia importancia en sí de la interpolación polinómica en procesos numéricos, y por otro lado, porque algunas de las cuestiones que serán presentadas serán necesarias para la demostración del Teorema 1.0.1 que será presentada en la Sección 1.3. La demostración que ofreceremos será la célebre dada por Sergei N. Bernstein en 1912-13 (véase [3]), lo cual hace necesario introducir previamente en la Sección 1.2 los *polinomios de Bernstein* para una función f definida en $[0, 1]$. La particularidad de esta demostración es que construye explícitamente la familia de polinomios aproximantes. El resultado, establecido para el intervalo $[0, 1]$, se deduce inmediatamente tras una transformación lineal a un intervalo compacto genérico $[a, b]$. Si bien es cierto que los polinomios de Bernstein tienen gran interés teórico, como quedará de manifiesto en la Sección 1.3, en la práctica no suelen emplearse debido a que son aproximantes de mala calidad (la convergencia hacia la función f es muy lenta). Sin embargo, poseen diversas propiedades adicionales de interés que serán expuestas en la Sección 1.4, con la que finaliza este primer capítulo. En particular, se demuestra que si la función f posee derivadas continuas hasta orden k en $[a, b]$, entonces las sucesivas derivadas de los polinomios de Bernstein hasta orden k también convergerán uniformemente a las sucesivas derivadas de f (algo que obviamente no se cumple en general con sucesiones funcionales). Si bien hemos consultado diversas referencias bibliográficas, el texto del que nos hemos guiado fundamentalmente es el libro [4] de P. J. Davis. Otras referencias recomendadas son el artículo [5] de A. Pinkus o la conferencia impartida por la profesora Yamilet Quintana Mato, disponible en [6].

1.1. Interpolación Polinómica

Comenzamos esta sección planteando y resolviendo el problema de interpolación polinómica (véase por ejemplo [2]).

Teorema 1.1.1 *Dados $n+1$ puntos $\{x_i\}_{i=0}^n$ distintos entre sí, y $n+1$ ordenadas $\{y_i\}_{i=0}^n$, existe un único polinomio $p_n \in \mathbb{P}_n$ tal que $p_n(x_i) = y_i, \forall i = 0, 1, \dots, n$.*

DEMOSTRACIÓN. La existencia queda probada de la siguiente expresión explícita, conocida como fórmula de Lagrange:

$$p_n(x) = \sum_{i=0}^n y_i l_i(x), \quad l_i(x) = \frac{\pi(x)}{\pi'(x_i)(x - x_i)}, \quad \pi(x) = \prod_{i=0}^n (x - x_i). \quad (1.1)$$

A $\pi \in \mathbb{P}_{n+1}$ se le conoce como el polinomio nodal y a $l_i \in \mathbb{P}_n$, para todo $0 \leq i \leq n$ como el i -ésimo polinomio fundamental de Lagrange. En cuanto a la unicidad,

si $p_1, p_2 \in \mathbb{P}_n$ son soluciones del problema, se sigue que $d = p_1 - p_2 \in \mathbb{P}_n$ se anula en los $n + 1$ nodos $\{x_i\}_{i=0}^n$, concluyendo del Teorema Fundamental del Álgebra que $d \equiv 0$.

□

Al polinomio único que resuelve el problema planteado en el Teorema 1.1.1 se le conoce como el *polinomio interpolador* a f en los nodos $\{x_i\}_{i=0}^n$. Cuando los datos y_i provienen de la evaluación de una función real de variable real f , denotaremos a éste según $p_n(f; x_1, \dots, x_n; x)$. Si la función f posee ciertas condiciones de regularidad, entonces es posible obtener una estimación del error cometido al aproximar la evaluación de la función f por p_n .

Teorema 1.1.2 *En las condiciones del Teorema 1.1.1, si $f \in \mathcal{C}^{(n+1)}(I_t)$, siendo I_t el intervalo más pequeño que contiene a los nodos y al punto $t \in \mathbb{R}$, entonces existe $\xi \in I_t$ tal que $f(t) - p_n(t) = \frac{\pi(t)}{(n+1)!} f^{(n+1)}(\xi)$.*

DEMOSTRACIÓN. El resultado es trivial si $t = x_i$, para algún $i \in \{0, \dots, n\}$. Supongamos pues que t está fijo y verifica $t \neq x_i$, para todo $i \in \{0, \dots, n\}$. Definamos la función $A(x) = E(x) - \frac{\pi(x)}{\pi(t)} E(t)$, para todo $x \in I_t$, siendo $E = f - p_n$. Es inmediato comprobar que $A \in \mathcal{C}^{(n+1)}(I_t)$ y que A tiene $n + 2$ ceros distintos en I_t , esto es, en $\{t, x_0, \dots, x_n\}$. Del Teorema del Valor Medio se sigue que $A^{(j)}$ tiene $n + 2 - j$ ceros distintos en I_t , para todo $j = 0, 1, \dots, n + 1$, y en particular, $A^{(n+1)}$ tiene un cero $\xi \in I_t$. El resultado se concluye así, teniendo en cuenta que $E^{(n+1)} = f^{(n+1)}$ y que $\pi^{(n+1)} = (n + 1)!$.

□

La fórmula explícita de Lagrange presentada en la demostración del Teorema 1.1.1 tiene numerosas aplicaciones, por ejemplo, se emplea en la construcción de *fórmulas de cuadratura* (integración numérica) *de tipo interpolatorio*. Sin embargo, posee un claro inconveniente, y es que no presenta un proceso dinámico para la computación del polinomio interpolador, en el sentido de que una vez computado p_n para el conjunto de nodos $\{x_k\}_{k=0}^n$, si añadimos un nuevo nodo x_{n+1} y deseamos computar p_{n+1} se hace necesario rehacer todos los cálculos. Para solventar este inconveniente presentamos un procedimiento alternativo para el cálculo del polinomio interpolador, conocido como *fórmula de Newton*.

En primer lugar, observamos que si escribimos $p_{n+1} = p_n + C$, siendo C una cierta función polinómica, entonces como $p_n(x_i) = p_{n+1}(x_i)$, para todo $i = 0, \dots, n$, se sigue que $p_{n+1}(f; x_0, \dots, x_{n+1}; x) = p_n(f; x_0, \dots, x_n; x) + a_{n+1}(x - x_0) \cdots (x - x_n)$. Así, de manera dinámica, añadiendo iterativamente un nuevo nodo, podemos expresar el polinomio interpolador en la forma

$$\begin{aligned} p_n(f; x_0, \dots, x_n; x) \\ = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \cdots + a_n(x - x_0) \cdots (x - x_{n-1}). \end{aligned}$$

Las cantidades $a_k = f[x_0, \dots, x_k]$ se conocen como *diferencias divididas de orden k relativas a los nodos x_0, \dots, x_k* , y pueden computarse a partir de la conocida *tabla de Newton*, que se basa en el siguiente

Lema 1.1.1 *Las diferencias divididas de orden superior verifican las siguientes propiedades:*

1. Si (i_0, i_1, \dots, i_n) es una permutación de $(0, 1, \dots, n)$, entonces $f[x_{i_0}, \dots, x_{i_n}] = f[x_0, \dots, x_n]$.
2. $f[x_i] = f(x_i), \forall i = 0, \dots, n$, y

$$f[x_0, \dots, x_n] = \frac{f[x_1, \dots, x_n] - f[x_0, \dots, x_{n-1}]}{x_n - x_0}. \quad (1.2)$$

DEMOSTRACIÓN. De la fórmula de Lagrange (1.1) se sigue que $p_n(x) = a_n x^n + \dots$, donde $a_n = f[x_0, \dots, x_n] = \sum_{i=0}^n \frac{y_i}{\pi'(x_i)}$. Se sigue por tanto que la diferencia dividida correspondiente a una permutación de $(0, 1, \dots, n)$ dará lugar a una reordenación de esa suma finita, demostrando así que las diferencias divididas no dependen del orden en el que se determinen los nodos.

Para demostrar el segundo apartado consideremos los polinomios

$$r(x) = p_{n-1}(f; x_1, \dots, x_n; x) \in \mathbb{P}_{n-1}, \quad s(x) = p_{n-1}(f; x_0, \dots, x_{n-1}; x) \in \mathbb{P}_{n-1}$$

y

$$t(x) = r(x) + \frac{x - x_n}{x_n - x_0} [r(x) - s(x)] \in \mathbb{P}_n. \quad (1.3)$$

Es inmediato comprobar que $t(x_i) = y_i$, para todo $0 \leq i \leq n$, por lo que debe darse $t(x) = p_n(f; x_0, \dots, x_n; x) \in \mathbb{P}_n$, por el Teorema 1.1.1. La propiedad (1.2) se sigue inmediatamente comparando el coeficiente director en ambos lados de (1.3).

□

Si consideramos el conjunto de nodos $\{x_0, \dots, x_n, t\}$ se sigue que

$$p_{n+1}(f; x_0, \dots, x_n, t; x) = p_n(f; x_0, \dots, x_n; x) + f[x_0, \dots, x_n, t](x - x_0) \cdots (x - x_n).$$

Haciendo $x = t$ y del Teorema 1.1.2 se sigue que si $f \in \mathcal{C}^{(n+1)}(I_t)$, entonces

$$f[x_0, \dots, x_n, t] = \frac{f^{(n+1)}(\xi)}{(n+1)!}, \quad \exists \xi \in I_t. \quad (1.4)$$

Esta propiedad se utilizará más adelante en la demostración del Lema 1.1.4.

Un caso muy particular de distribución de nodos es la *distribución equiespaciada*, en la que $x_{i+1} - x_i = h > 0$ (cantidad constante), para todo i .

Definición 1.1.1 *Sea $x_i = x_0 + ih$ con $i = 0, 1, 2, \dots$ y $h > 0$.*

1. Se define el operador diferencia progresiva Δ según $\Delta f(x) = f(x+h) - f(x)$. Haciendo $f(x_i) = f_i$, en la distribución de nodos anterior resulta $\Delta f_i = f_{i+1} - f_i$.
2. Para $r \in \mathbb{N}$ definimos el operador diferencia progresiva de orden r , Δ^r , según $\Delta^r f(x) = \Delta^{r-1} f(x+h) - \Delta^{r-1} f(x)$. En la distribución de nodos anterior se tiene que $\Delta^r f_i = \Delta^{r-1} f_{i+1} - \Delta^{r-1} f_i$.

De esta definición observamos que

$$\Delta f_i = f_{i+1} - f_i, \quad \Delta^2 f_i = f_{i+2} - 2f_{i+1} + f_i, \quad \Delta^3 f_i = f_{i+3} - 3f_{i+2} + 3f_{i+1} - f_i. \quad (1.5)$$

Podemos generalizar esta observación en el siguiente

Lema 1.1.2 *Dada la distribución equiespaciada $x_i = x_0 + ih$ con $i = 0, 1, \dots$ y $h > 0$, entonces*

$$\Delta^n x_k = \sum_{r=0}^n (-1)^{n-r} \binom{n}{r} x_{k+r}. \quad (1.6)$$

DEMOSTRACIÓN. Procedemos por inducción sobre n . Los casos $n = 0, 1, 2$ han sido probados en (1.5). Supongamos que la propiedad es cierta para n . Entonces,

$$\begin{aligned} \Delta^{n+1} x_k &= \Delta(\Delta^n x_k) = \Delta\left(\sum_{r=0}^n (-1)^{n-r} \binom{n}{r} x_{k+r}\right) = \sum_{r=0}^n (-1)^{n-r} \binom{n}{r} \Delta x_{k+r} \\ &= \sum_{r=0}^n (-1)^{n-r} \binom{n}{r} (x_{k+r+1} - x_{k+r}) \\ &= \sum_{r=1}^{n+1} (-1)^{n+1-r} \binom{n}{r-1} x_{k+r} - \sum_{r=0}^n (-1)^{n-r} \binom{n}{r} x_{k+r} \\ &= \sum_{r=1}^n (-1)^{n+1-r} \left[\binom{n}{r-1} + \binom{n}{r} \right] x_{k+r} + \binom{n}{n} x_{n+k+1} - (-1)^n \binom{n}{0} x_k \\ &= \sum_{r=0}^{n+1} (-1)^{n+1-r} \binom{n+1}{r} x_{k+r}. \end{aligned}$$

□

El siguiente resultado establece una relación entre el operador diferencia progresiva de orden k y las diferencias divididas de orden k relativas a una distribución equiespaciada de $k+1$ nodos.

Lema 1.1.3 *Si $\{x_i\}_{i=0}^k$ es una distribución equiespaciada de $k+1$ nodos, siendo $h = x_{i+1} - x_i$ para todo $i = 0, \dots, k-1$, entonces*

$$f[x_0, \dots, x_k] = \frac{\Delta^k f_0}{k! h^k}. \quad (1.7)$$

DEMOSTRACIÓN. El resultado se sigue trivialmente para $n = 0, 1$: $f[x_0] = f(x_0) = \frac{\Delta^0 f_0}{0! h^0}$ y $f[x_0, x_1] = \frac{f_1 - f_0}{h} = \frac{\Delta^1 f_0}{h^1 1!}$. Procedemos mediante inducción sobre k , admitiendo que el resultado es cierto para k y probándolo para $k+1$. Así, de (1.2) y aplicando la hipótesis de inducción tenemos que

$$\begin{aligned} f[x_0, \dots, x_{k+1}] &= \frac{f[x_1, \dots, x_{k+1}] - f[x_0, \dots, x_k]}{x_{k+1} - x_0} = \frac{1}{(k+1)h} \left[\frac{\Delta^k f_1}{k!h^k} - \frac{\Delta^k f_0}{k!h^k} \right] = \frac{\Delta^k (f_1 - f_0)}{(k+1)!h^{k+1}} \\ &= \frac{\Delta^{k+1} f_0}{(k+1)!h^{k+1}}. \end{aligned}$$

□

Concluimos esta sección con una propiedad que se deduce como consecuencia inmediata de los resultados previos y que será necesaria para la demostración del Teorema de Aproximación de Weierstrass en la Sección 1.3.

Lema 1.1.4 *Si $f \in \mathbb{P}_m$, entonces $\Delta^n f(0) = 0, \forall n > m$.*

DEMOSTRACIÓN.

Como $f \in \mathbb{P}_m$, tenemos que $f^{(n)} \equiv 0$, para todo $n > m$. De (1.4) y de (1.7) se sigue así que $\Delta^n f(0) = n!h^n f[x_0, \dots, x_n] = h^n f^{(n)}(\xi) = 0$, como queríamos demostrar.

□

1.2. Polinomios de Bernstein

Como ya hemos comentado, existen diversas demostraciones del Teorema de Aproximación de Weierstrass. La demostración que presentamos en la Sección 1.3, presentada por S. N. Bernstein, es constructiva, es decir, construye explícitamente la familia de polinomios aproximantes. Para ello es necesario previamente introducir en esta sección la familia de polinomios que llevan su nombre.

Definición 1.2.1 *Sea f una función definida en $[0, 1]$ y $n \geq 1$. El n -ésimo polinomio de Bernstein para f se define como*

$$B_n(f; x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}. \quad (1.8)$$

Nótese que $B_n(f; 0) = f(0)$ y $B_n(f; 1) = f(1)$ dado que en ambos casos se anulan todos los términos de la suma, excepto para $k = 0$ en el primer caso y para $k = n$ en el segundo.

En el resultado que demostraremos a continuación se establece una expresión alternativa para los polinomios de Bernstein, que permite conocer explícitamente los coeficientes de los monomios x^t en la expresión de B_n , $0 \leq t \leq n$.

Teorema 1.2.1 *El n -ésimo polinomio de Bernstein para la función f definida en $[0, 1]$ viene dado por $B_n(f; x) = \sum_{k=0}^n \Delta^k f(0) \binom{n}{k} x^k$, donde las diferencias progresivas son computadas en la partición equiespaciada $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\}$.*

DEMOSTRACIÓN.

Partiendo de la Definición 1.2.1 se tiene

$$\begin{aligned}
 B_n(f; x) &= \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k \underbrace{(1-x)^{n-k}}_{(1)} \\
 &= \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k \underbrace{\sum_{j=0}^{n-k} \binom{n-k}{j} (-1)^{n-k-j} x^{n-k-j}}_{(2)} \\
 &= \sum_{k=0}^n \sum_{j=0}^{n-k} f\left(\frac{k}{n}\right) \binom{n}{k} \binom{n-k}{j} (-1)^{n-k-j} x^{n-j}.
 \end{aligned}$$

La igualdad (1) = (2) es el desarrollo del Binomio de Newton de la potencia $(1-x)^{n-k}$. Reordenando este sumatorio doble y agrupando los términos que acompañan a x^t , para $t = 0, \dots, n$, obtenemos

$$\begin{aligned}
 B_n(f; x) &= \sum_{t=0}^n x^t \sum_{k=0}^t f\left(\frac{k}{n}\right) \binom{n}{k} \binom{n-k}{n-t} (-1)^{t-k} \\
 &= \sum_{t=0}^n x^t \binom{n}{t} \sum_{k=0}^t f\left(\frac{k}{n}\right) \binom{t}{k} (-1)^{t-k} = \sum_{t=0}^n \Delta^t f(0) \binom{n}{t} x^t,
 \end{aligned}$$

donde en la primera igualdad hemos aplicado la relación $\binom{n}{k} \binom{n-k}{n-t} = \binom{n}{t} \binom{t}{k}$ y en la segunda el Lema 1.1.2 haciendo $k = 0$.

□

La combinación del Lema 1.1.4 y del Teorema 1.2.1 permite probar el siguiente

Lema 1.2.1 Si $B_n(f, x)$ denota el n -ésimo polinomio de Bernstein para la función f definida en $[0, 1]$ y $n \geq 1$, entonces

$$B_n(1; x) \equiv 1, \quad B_n(x; x) = x, \quad B_n(x^2; x) = \frac{n-1}{n}x^2 + \frac{x}{n}. \quad (1.9)$$

DEMOSTRACIÓN.

1. Para $f \equiv 1$ se tiene que $\Delta^0 f(0) = 1$ y $\Delta^n f(0) = 0, \forall n \geq 1$. Por tanto,

$$B_n(1; x) = \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} = \sum_{t=0}^n \Delta^t f(0) \binom{n}{t} x^t = 1.$$

Obsérvese que el primer sumatorio no es otra cosa que el desarrollo del Binomio de Newton de $1 = 1^n = (x + 1 - x)^n$.

2. Para $f(x) = x$, $\Delta^0 f(0) = 0$, $\Delta^1 f(0) = \frac{1}{n}$ y $\Delta^n f(0) = 0, \forall n \geq 2$. Entonces,

$$B_n(x; x) = \sum_{k=0}^n \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k} = \sum_{t=0}^n \Delta^t f(0) \binom{n}{t} x^t = \frac{1}{n} \binom{n}{1} x = x.$$

3. Para $f(x) = x^2$, se tiene que $\Delta^0 f(0) = 0$, $\Delta^1 f(0) = \frac{1}{n^2}$, $\Delta^2 f(0) = \frac{2}{n^2}$ y $\Delta^n f(0) = 0, \forall n \geq 3$. Se sigue finalmente que

$$B_n(x^2; x) = \sum_{t=0}^n \Delta^t f(0) \binom{n}{t} x^t = \frac{1}{n^2} \binom{n}{1} x + \frac{2}{n^2} \binom{n}{2} x^2 = \frac{n-1}{n} x^2 + \frac{x}{n}.$$

□

Los polinomios de Bernstein tienen un importante interés teórico. Constituyen una familia de polinomios explícitos que permiten aproximar uniformemente a toda función $f \in \mathcal{C}([0, 1])$. Sin embargo, en la práctica no se suelen emplear dado que la convergencia hacia la función f es muy lenta. A modo de ejemplo, de (1.9) es inmediato comprobar que para obtener $\|B_n(x^2; x) - x^2\|_\infty < 10^{-6}$ es necesario considerar $n > 250.000$. En efecto,

$$\max_{x \in [0,1]} |B_n(x^2; x) - x^2| = \max_{x \in [0,1]} \frac{1}{n} |x(1-x)| \leq \frac{1}{4n} < 10^{-6} \Leftrightarrow n > 250.000.$$

Esta lenta convergencia queda también de manifiesto en la Figura 1.1.

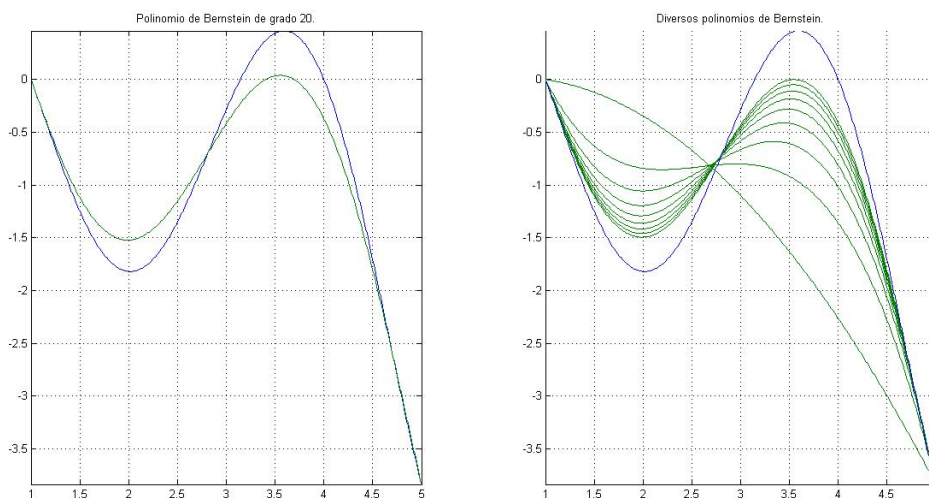


Figura 1.1. Representación gráfica de la función $f(x) = (x-1)(x-4) \cdot \sin(x)$, con $x \in [1, 5]$, y diversos polinomios de Bernstein correspondientes, computados empleando software MATLAB con el código incluido en el Apéndice A.1 de esta memoria. En el gráfico de la izquierda se representan la función f y el polinomio $B_{20}(f; x)$ mientras que en el gráfico de la derecha se representa la función f junto a la familia de polinomios $\{B_{2k}\}_{k=1}^{10}$. Se aprecia claramente la lenta convergencia de los polinomios aproximantes hacia la función elegida.

1.3. Demostración del Teorema de Aproximación de Weierstrass

Para probar el Teorema 1.0.1 comenzaremos probando el siguiente

Teorema 1.3.1 *Sea f una función acotada en $[0, 1]$ y denotemos por $B_n(f; x)$ al polinomio de Bernstein de grado $n \geq 1$ para la función f en $[0, 1]$. Entonces,*

$$\lim_{n \rightarrow \infty} B_n(f; x) = f(x) \tag{1.10}$$

en cualquier punto $x \in [0, 1]$ donde f sea continua. Además, el límite se cumple uniformemente si $f \in \mathcal{C}([0, 1])$.

DEMOSTRACIÓN. Como consecuencia inmediata del Lema 1.2.1 se sigue que

$$\sum_{k=0}^n (k - nx)^2 \binom{n}{k} x^k (1 - x)^{n-k} = nx(1 - x). \tag{1.11}$$

Veamos ahora que para $\delta > 0$ y $0 \leq x \leq 1$ se cumple que

$$\sum_{|\frac{k}{n} - x| \geq \delta} \binom{n}{k} x^k (1 - x)^{n-k} \leq \frac{1}{4n\delta^2} \tag{1.12}$$

(es decir, estamos sumando sobre todos aquellos índices $k = 0, 1, \dots, n$ para los cuales se cumple que $|\frac{k}{n} - x| \geq \delta$, donde las cantidades δ y x se han fijado previamente). En efecto,

$$\begin{aligned} \sum_{|\frac{k}{n} - x| \geq \delta} \binom{n}{k} x^k (1 - x)^{n-k} &\leq \frac{1}{\delta^2} \sum_{|\frac{k}{n} - x| \geq \delta} \left(\frac{k}{n} - x\right)^2 \binom{n}{k} x^k (1 - x)^{n-k} \\ &\leq \frac{1}{\delta^2} \sum_{k=0}^n \left(\frac{k}{n} - x\right)^2 \binom{n}{k} x^k (1 - x)^{n-k} = \frac{nx(1-x)}{\delta^2 n^2} \leq \frac{1}{4n\delta^2}, \end{aligned}$$

donde hemos usado en la última igualdad que $x(1-x) \leq \frac{1}{4}$, para todo $x \in [0, 1]$.

Queda probada, por tanto, la relación (1.12). Dado que $1 = \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k}$,

se sigue $f(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$, y por tanto

$$\begin{aligned} f(x) - B_n(f; x) &= \sum_{k=0}^n [f(x) - f\left(\frac{k}{n}\right)] \binom{n}{k} x^k (1-x)^{n-k} \\ &= \sum_{|\frac{k}{n} - x| < \delta} [f(x) - f\left(\frac{k}{n}\right)] \binom{n}{k} x^k (1-x)^{n-k} \\ &\quad + \sum_{|\frac{k}{n} - x| \geq \delta} [f(x) - f\left(\frac{k}{n}\right)] \binom{n}{k} x^k (1-x)^{n-k}. \end{aligned}$$

Como f está acotada en $[0, 1]$, existe $M > 0$ tal que $|f(x)| \leq M$ para todo $x \in [0, 1]$, y si $\alpha, \beta \in [0, 1]$, siempre se cumple que $|f(\alpha) - f(\beta)| \leq 2M$. Sea $x \in [0, 1]$ un punto de continuidad de f . Entonces, dado $\epsilon > 0$, existe $\delta > 0$ (que depende de ϵ) tal que $|f(x) - f(y)| < \epsilon$, siempre que $|x - y| < \delta$. Así, usando esta cantidad $\delta > 0$ en (1.12) tendremos que

$$\begin{aligned}
|f(x) - B_n(f; x)| &\leq \sum_{|\frac{k}{n} - x| < \delta} |f(x) - f\left(\frac{k}{n}\right)| \binom{n}{k} x^k (1-x)^{n-k} \\
&\quad + \sum_{|\frac{k}{n} - x| \geq \delta} |f(x) - f\left(\frac{k}{n}\right)| \binom{n}{k} x^k (1-x)^{n-k} \\
&\leq \epsilon \sum_{|\frac{k}{n} - x| < \delta} \binom{n}{k} x^k (1-x)^{n-k} \\
&\quad + 2M \sum_{|\frac{k}{n} - x| \geq \delta} \binom{n}{k} x^k (1-x)^{n-k} \\
&\leq \epsilon \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} + \frac{2M}{4n\delta^2} \\
&= \epsilon + \frac{M}{2n\delta^2}.
\end{aligned}$$

Vemos por tanto que $|f(x) - B_n(f; x)| \leq 2\epsilon$, para un n suficientemente grande. Como $\epsilon > 0$ es arbitrario, esto demuestra (1.10). Supongamos ahora que $f \in \mathcal{C}([0, 1])$. Entonces, f es uniformemente continua en $[0, 1]$, por ser éste cerrado y acotado, y dado $\epsilon > 0$, existe $\delta > 0$ tal que $|f(x) - f(y)| < \epsilon$, para todo $x, y \in [0, 1]$ con $|x - y| < \delta$. Tal desigualdad se cumple independientemente de la cantidad x elegida, y la convergencia a $f(x)$ es por tanto uniforme en $[0, 1]$. □

La demostración del Teorema 1.0.1 es consecuencia inmediata del Teorema 1.3.1: DEMOSTRACIÓN. Consideremos la transformación lineal $y = \frac{x-a}{b-a}$, que transforma biyectivamente $[a, b]$ en $[0, 1]$, y la función $g(y) = f(a + (b-a)y) \in \mathcal{C}([0, 1])$. Por el Teorema 1.3.1, dado $\epsilon > 0$, podemos encontrar un polinomio $r(y)$ tal que $|g(y) - r(y)| \leq \epsilon$, para todo $0 \leq y \leq 1$. El resultado se concluye considerando el polinomio $p(x) = r\left(\frac{x-a}{b-a}\right)$ definido en $[a, b]$. □

1.4. Propiedades adicionales sobre los Polinomios de Bernstein

A pesar de que los polinomios de Berstein no se suelen emplear en la práctica en procesos numéricos, tal y como se ha comentado anteriormente, debido a la lenta

convergencia hacia la función f , éstos sí que poseen interesantes propiedades teóricas. Algunas de ellas serán expuestas en esta última sección.

Comenzamos con algunos resultados sobre la aproximación simultánea de funciones y sus derivadas.

Lema 1.4.1 *Sea $p \geq 0$ un número entero. Entonces*

$$B_{n+p}^{(p)}(f; x) = \frac{(n+p)!}{n!} \sum_{t=0}^n \Delta^p f \left(\frac{t}{n+p} \right) \binom{n}{t} x^t (1-x)^{n-t}. \quad (1.13)$$

DEMOSTRACIÓN. Aplicando la fórmula para la derivación del producto de Leibnitz $(uv)^{(p)} = \sum_{j=0}^p \binom{p}{j} u^{(j)} v^{(p-j)}$ a (1.8), obtenemos

$$B_{n+p}^{(p)}(f; x) = \sum_{k=0}^{n+p} f \left(\frac{k}{n+p} \right) \binom{n+p}{k} \sum_{j=0}^p \binom{p}{j} (x^k)^{(j)} [(1-x)^{n+p-k}]^{(p-j)}. \quad (1.14)$$

Como $(x^k)^{(j)} = k! x^{k-j} / (k-j)!$, para todo $k-j \geq 0$ y

$$[(1-x)^{n+p-k}]^{(p-j)} = (-1)^{p-j} (n+p-k)! \frac{(1-x)^{n+j-k}}{(n+j-k)!},$$

para todo $k-j \leq n$, se sigue que (1.14) se convierte en

$$B_{n+p}^{(p)}(f; x) = \sum_{k=0}^{n+p} \sum_{j=0, 0 \leq k-j \leq n}^p f \left(\frac{k}{n+p} \right) \frac{(n+p)!}{(k-j)!(n+j-k)!} \binom{p}{j} (-1)^{p-j} x^{k-j} (1-x)^{n+j-k}. \quad (1.15)$$

Realizamos el cambio de variable $k-j = t$ y vemos que $0 \leq t \leq n$, $j = 0, 1, \dots, p$, corresponde al rango de la suma en (1.15), y por tanto la misma queda como

$$B_{n+p}^{(p)}(f; x) = (n+p)! \sum_{t=0}^n \frac{x^t (1-x)^{n-t}}{t!(n-t)!} \sum_{j=0}^p (-1)^{p-j} \binom{p}{j} f \left(\frac{t+j}{n+p} \right).$$

Por último, (1.13) sigue de (1.6). □

El lema que acabamos de probar nos permitirá demostrar el siguiente resultado de gran importancia: nos afirma que las sucesivas derivadas de los polinomios de Bernstein siguen convergiendo uniformemente a las sucesivas derivadas de la función, siempre que esta sea suficientemente regular en el intervalo.

Teorema 1.4.1 *Sea $f(x) \in \mathcal{C}^p([0, 1])$. Entonces, $\lim_{n \rightarrow \infty} B_n^{(p)}(f; x) = f^{(p)}(x)$ uniformemente en $[0, 1]$.*

DEMOSTRACIÓN. Combinando (1.4) con (1.7), obtenemos $\Delta^n f(x_0) = h^n f^{(n)}(\xi)$, $x_0 < \xi < x_n$, y así $\Delta^p f\left(\frac{t}{n+p}\right) = \frac{1}{(n+p)^p} f^{(p)}(\xi_t)$, para algún ξ_t tal que $\frac{t}{n+p} < \xi_t < \frac{t+p}{n+p}$, $t = 0, 1, \dots, n$. Por tanto, por el Lema 1.4.1,

$$B_{n+p}^{(p)}(f; x) = \frac{(n+p)!}{n!(n+p)^p} \sum_{t=0}^n f^{(p)}(\xi_t) \binom{n}{t} x^t (1-x)^{n-t}.$$

Se sigue ahora que

$$\begin{aligned} \frac{n!(n+p)^p}{(n+p)!} B_{n+p}^{(p)}(f; x) &= \sum_{t=0}^n f^{(p)}\left(\frac{t}{n}\right) \binom{n}{t} x^t (1-x)^{n-t} \\ &\quad + \sum_{t=0}^n \{f^{(p)}(\xi_t) - f^{(p)}\left(\frac{t}{n}\right)\} \binom{n}{t} x^t (1-x)^{n-t}. \end{aligned} \quad (1.16)$$

Ahora bien, dado que $\frac{t}{n+p} \leq \frac{t}{n} \leq \frac{t+p}{n+p}$, $t = 0, \dots, n$, se sigue que $|\xi_t - \frac{t}{n}| < \frac{t+p}{n+p} - \frac{t}{n+p} = \frac{p}{n+p}$. De la continuidad uniforme de $f^{(p)}(x)$, dado $\epsilon > 0$, podemos encontrar $n_0 \in \mathbb{N}$ tal que $|f^{(p)}(\xi_t) - f^{(p)}(t/n)| < \epsilon$, para todo $n \geq n_0$, y para todo $x \in [0, 1]$. Además, $\lim_{n \rightarrow \infty} \frac{n!(n+p)^p}{(n+p)!} = 1$, y por el Teorema 1.3.1, el primer sumando de (1.16) converge uniformemente a $f^{(p)}(x)$.

□

Otra propiedad interesante que presentan los polinomios de Bernstein es que funciones convexas (para las cuales, el segmento de recta que une dos puntos distintos del intervalo de la curva nunca queda por debajo de esta) dan lugar a aproximantes convexos. Esto se puede expresar como el siguiente

Teorema 1.4.2 *Sea f convexa en $[0, 1]$. Entonces,*

$$B_{n-1}(f; x) \geq B_n(f; x), \quad 0 < x < 1, \quad n \geq 2. \quad (1.17)$$

Además, si $f \in \mathcal{C}[0, 1]$, la desigualdad en (1.17) es estricta, a menos que f sea lineal en cada uno de los intervalos $[\frac{j-1}{n-1}, \frac{j}{n-1}]$, $j = 1, 2, \dots, n-1$. En tal caso, $B_{n-1}(f; x) = B_n(f; x)$.

DEMOSTRACIÓN.

Si en (1.8) realizamos el cambio $t = \frac{x}{1-x}$ obtenemos

$$\begin{aligned} &(1-x)^{-n} (B_{n-1}(f; x) - B_n(f; x)) \\ &= (1+t) \sum_{k=0}^{n-1} f\left(\frac{k}{n-1}\right) \binom{n-1}{k} t^k - \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} t^k \\ &= \sum_{k=1}^{n-1} f\left(\frac{k}{n-1}\right) \binom{n-1}{k} t^k + f(0) + \sum_{k=1}^{n-1} f\left(\frac{k-1}{n-1}\right) \binom{n-1}{k-1} t^k + f(1)t^n \\ &\quad - \sum_{k=1}^{n-1} f\left(\frac{k}{n}\right) \binom{n}{k} t^k - f(0) - f(1)t^n = \sum_{k=1}^{n-1} c_k t^k, \end{aligned}$$

donde

$$c_k = \frac{(n-1)!}{(k-1)!(n-k-1)!} \left\{ \frac{1}{k} f\left(\frac{k}{n-1}\right) + \frac{1}{n-k} f\left(\frac{k-1}{n-1}\right) - \frac{n}{k(n-k)} f\left(\frac{k}{n}\right) \right\}. \quad (1.18)$$

Como $\frac{k-1}{n-1} < \frac{k}{n} < \frac{k}{n-1}$, y f es convexa, la cantidad en llaves de (1.18) es mayor

o igual que 0. Entonces, $\sum_{k=1}^{n-1} c_k t^k \geq 0$, y queda demostrado (1.17). Si f es lineal

en cada uno de los intervalos $[\frac{j-1}{n-1}, \frac{j}{n-1}]$, entonces $c_j = 0$, y como $f \in \mathcal{C}[0, 1]$ y es convexa, (1.18) implica que f es lineal en cada intervalo. □

El lema que vamos a probar a continuación lo usaremos después en la demostración del Teorema de Voronovsky, con el que finalizaremos este primer capítulo. Este resultado nos indica que el orden de convergencia de los polinomios de Bernstein es lineal, es decir, si bien es cierto que estos polinomios nos han permitido obtener una demostración constructiva del Teorema de Aproximación de Weierstrass, la convergencia de los aproximantes a la función es lenta.

Lema 1.4.2 *Para todo $x \in [0, 1]$, existe una constante C , independiente de n , tal que $\sum_{|k/n-x| \geq n^{-\frac{1}{4}}} \binom{n}{k} x^k (1-x)^{n-k} \leq \frac{C}{n^{\frac{3}{2}}}$.*

DEMOSTRACIÓN. Para $m \in \mathbb{N}$, consideremos las sumas

$$S_m(x) = \sum_{k=0}^n (k-nx)^m \binom{n}{k} x^k (1-x)^{n-k}. \quad (1.19)$$

Tomando derivada en (1.19), obtenemos

$$\begin{aligned} S'_m(x) &= \sum_{k=0}^n \binom{n}{k} (k-nx)^{m-1} x^{k-1} (1-x)^{n-k-1} [-mnx(1-x) + (k-nx)^2] \\ &= -mnS_{m-1}(x) + \frac{S_{m+1}(x)}{x(1-x)}. \end{aligned}$$

Por tanto, $S_{m+1}(x) = x(1-x) [S'_m(x) + mnS'_{m-1}(x)]$. Por esta relación de recurrencia que acabamos de obtener, podemos deducir que cada $S_m(x)$ es un polinomio en x y n . De los dos primeros apartados del Lema 1.2, y de (1.11), tenemos $S_0(x) = 1$, $S_1(x) = 0$, $S_2(x) = nx(1-x)$. A partir de aquí, con la relación de recurrencia, es fácil ver que S_3 tiene grado 1, S_4 y S_5 tienen grado 2, y S_6 es de grado 3 en n . Por tanto existe una constante C , tal que $|S_6(x)| \leq Cn^3$ para $x \in [0, 1]$. Ya que $|\frac{k}{n} - x| \geq n^{-\frac{1}{4}}$ implica $\frac{(k-nx)^6}{n^{\frac{9}{2}}} \geq 1$, se sigue que

$$\begin{aligned} \sum_{|k/n-x| \geq n^{-\frac{1}{4}}} \binom{n}{k} x^k (1-x)^{n-k} &\leq \frac{1}{n^{\frac{9}{2}}} \sum_{k=0}^n (k-nx)^6 \binom{n}{k} x^k (1-x)^{n-k} \\ &= n^{-\frac{9}{2}} S_6(x) \leq \frac{C}{n^{\frac{3}{2}}}. \end{aligned}$$

□

Teorema 1.4.3 (Voronovsky). Sea f acotada en $[0, 1]$ y $x_0 \in [0, 1]$ tal que existe $f''(x_0)$. Entonces,

$$\lim_{n \rightarrow \infty} n [B_n(f; x_0) - f(x_0)] = \frac{1}{2} x_0 (1 - x_0) f''(x_0). \quad (1.20)$$

DEMOSTRACIÓN. Por el Teorema de Taylor tenemos que

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)(x - x_0)^2}{2} + s(x)(x - x_0)^2,$$

donde $\lim_{x \rightarrow x_0} s(x) = 0$. Realizando el cambio $x = \frac{k}{n}$ obtenemos

$$f\left(\frac{k}{n}\right) = f(x_0) + f'(x_0)\left(\frac{k}{n} - x_0\right) + \frac{f''(x_0)}{2}\left(\frac{k}{n} - x_0\right)^2 + s\left(\frac{k}{n}\right)\left(\frac{k}{n} - x_0\right)^2. \quad (1.21)$$

Multiplicamos ambos lados de (1.21) por $\binom{n}{k} x_0^k (1 - x_0)^{n-k}$ y sumamos desde $k = 0$ hasta $k = n$. De nuevo, utilizando el Lema 1.2.1 y (1.11) se sigue que

$$B_n(f; x_0) = f(x_0) + \frac{x_0(1 - x_0)f''(x_0)}{2n} + \sum_{k=0}^n s\left(\frac{k}{n}\right)\left(\frac{k}{n} - x_0\right)^2 \binom{n}{k} x_0^k (1 - x_0)^{n-k}. \quad (1.22)$$

Denotemos al tercer sumando de (1.22) por S y sea $\epsilon > 0$. Podemos encontrar n suficientemente grande tal que $|x - x_0| < \frac{1}{n^{\frac{1}{4}}}$ implique $|s(x)| < \epsilon$. Así,

$$\begin{aligned} |S| &\leq \sum_{|k/n - x_0| < n^{-\frac{1}{4}}} |s\left(\frac{k}{n}\right)| \left(\frac{k}{n} - x_0\right)^2 \binom{n}{k} x_0^k (1 - x_0)^{n-k} \\ &\quad + \sum_{|k/n - x_0| \geq n^{-\frac{1}{4}}} |s\left(\frac{k}{n}\right)| \left(\frac{k}{n} - x_0\right)^2 \binom{n}{k} x_0^k (1 - x_0)^{n-k}. \end{aligned}$$

Entonces,

$$|S| \leq \epsilon \sum_{k=0}^n \left(\frac{k}{n} - x_0\right)^2 \binom{n}{k} x_0^k (1 - x_0)^{n-k} + M \cdot \sum_{|k/n - x_0| \geq n^{-\frac{1}{4}}} \binom{n}{k} x_0^k (1 - x_0)^{n-k}$$

donde $M = \sup_{0 \leq x \leq 1} s(x)(x - x_0)^2$. Por (1.11) y el Lema 1.4.2, $|S| \leq \frac{\epsilon x_0(1-x_0)}{n} + \frac{MC}{n^{\frac{3}{2}}}$.

Se sigue ahora de (1.22) que

$$\left| n [B_n(f; x_0) - f(x_0)] - \frac{x_0(1 - x_0)}{2} f''(x_0) \right| = |nS| \leq \epsilon x_0(1 - x_0) + \frac{MC}{n^{\frac{1}{2}}}.$$

Como ϵ es arbitrario, queda demostrado (1.20).

□

El problema de mejor aproximación

2.1. ¿Qué entendemos por mejor aproximación?

En el capítulo anterior estudiamos situaciones en las que funciones continuas definidas en un compacto $[a, b]$ pueden ser aproximadas en norma uniforme por polinomios. Se puede sobreentender que por regla general, para obtener aproximaciones más precisas, se necesita un grado mayor de los mismos.

En este capítulo consideraremos también otras normas, no solo la uniforme, y estudiaremos aproximaciones mediante funciones polinómicas en las que el grado se fija de antemano. Como vemos, se trata de un problema de aproximación distinto al abordado en el Capítulo 1. Para entender mejor este problema, consideramos el siguiente ejemplo: ¿De qué maneras podemos aproximar la función x^3 por una línea recta en $[0, 1]$? Para responder este tipo de preguntas se necesita primero definir la noción de cercanía de aproximación y mejor aproximación, para a continuación dar respuestas a preguntas tales como si para una función f definida en $[a, b]$, existe algún elemento de \mathbb{P}_n cuya cercanía a f no sea superada por los demás elementos de \mathbb{P}_n . De ser así, tal elemento sería conocido como la mejor aproximación de f en \mathbb{P}_n .

Ejemplo 2.1.1 *Aproximar $y = x^3$ en $[0, 1]$ por una línea recta $l(x)$, tal que*

- a) $\int_0^1 (x^3 - l(x))^2 dx$ sea mínima.
- b) $\max_{0 \leq x \leq 1} |x^3 - l(x)|$ sea mínima.

Las soluciones únicas a estos dos problemas son a) $l(x) = x$ y b) $l(x) = 6x - 2\sqrt{2}$.

Las investigaciones en este tema se ocupan principalmente de cuatro cuestiones:

1. ¿Bajo qué circunstancias existe una mejor aproximación? Si existe, ¿es única?
2. ¿Cómo pueden caracterizarse las mejores aproximaciones, de una manera analítica o geométrica?
3. ¿Cómo pueden los mejores aproximantes computarse numéricamente?
4. ¿Cuáles son las propiedades asintóticas de la mejor aproximación?

Nuestro primer objetivo será probar resultados en todas esas categorías. Para ello, lo razonable es partir de una estructura matemática abstracta que describa propiedades sobre la mejor aproximación, independientemente del criterio específico de cercanía de aproximación. Dicha estructura nos la proporciona la teoría de espacios normados que abordaremos en la próxima sección.

2.2. Aproximación en espacios normados

Definición 2.2.1 *Un espacio vectorial X se denomina espacio normado, si para cada elemento $x \in X$, se define un número real, que denotamos por $\|x\|$, que satisface las siguientes propiedades:*

1. $\|x\| \geq 0$ (Positividad),
2. $\|x\| = 0$, si y solo si, $x = 0$ (Definido),
3. $\|\alpha x\| = |\alpha| \cdot \|x\|$, para todo escalar α (Homogeneidad),
4. $\|x + y\| \leq \|x\| + \|y\|$ (Desigualdad triangular).

La cantidad $\|x\|$ es conocida como norma de x . Generalmente, se denotará como $(X, \|\cdot\|)$ al espacio vectorial normado. En el caso de que no verifique la propiedad 2, pero sí las restantes, diremos que $\|\cdot\|$ es una seminorma.

Definición 2.2.2 ■ *Diremos que ω es una función peso definida en $[a, b]$ si*

$\omega(x) \geq 0$, con $c_0 = \int_a^b \omega(x) dx < \infty$, $\forall x \in [a, b]$, y $\omega(x) > 0$, $\forall x \in A \subseteq [a, b]$, siendo A un conjunto con medida de Lebesgue positiva.

- *Sea $0 < p < \infty$. La clase de funciones f medibles para las cuales $|f|^p$ es integrable en $[a, b]$ con respecto a la función peso ω es conocida como $L_\omega^p[a, b]$ (si $p = 1$, la clase es designada como $L_\omega[a, b]$ y si $\omega = 1$, se designa como $L^p[a, b]$). Si $f \in L_\omega^p[a, b]$, se define*

$$\|f\|_p = \left(\int_a^b |f(x)|^p \omega(x) dx \right)^{\frac{1}{p}}. \quad (2.1)$$

- *Si $p = \infty$, $L^\infty[a, b]$ se define como el conjunto de las funciones medibles y acotadas. Para $f \in L^\infty[a, b]$ definimos $\|f\|_\infty = \sup_{x \in [a, b]} |f(x)|$.*

Ejemplo 2.2.1 *Es fácil comprobar que $L^p[a, b]$ es un espacio vectorial para $p > 0$, y que si $p \geq 1$, (2.1) cumple las Propiedades 1-3 de norma en la Definición 2.2.1. Sin embargo, la Propiedad 4, conocida en este caso como Desigualdad de Minkowski, no es trivial. Su deducción se ha incluido en el Apéndice A.2 de esta memoria. $(L^p[a, b], \|\cdot\|_p)$ es, por tanto, un espacio vectorial normado con $p \geq 1$.*

Ejemplo 2.2.2 Sea $X = \mathcal{C}([a, b])$ y ω una función peso definida en $[a, b]$. Si consideramos la norma $\|f\|_p = \left(\int_a^b f(x)^p \omega(x) dx\right)^{\frac{1}{p}}$, $1 \leq p < \infty$, obtenemos que $(X, \|\cdot\|_p)$ es un espacio normado.

Ejemplo 2.2.3 Los espacios $\mathbb{R}^n = \{(x_1, \dots, x_n) : x_i \in \mathbb{R}, i = 1, \dots, n\}$ con las normas

$$\|x\|_p = \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}, \text{ con } 1 \leq p < \infty, \quad \text{y} \quad \|x\|_\infty = \sup_{i \in \{1, \dots, n\}} |x_i|$$

son espacios vectoriales normados.

Una consecuencia inmediata de la definición de norma es

$$|\|x\| - \|y\|| \leq \|x - y\|. \quad (2.2)$$

En efecto, $\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\|$, por tanto, $\|x\| - \|y\| \leq \|x - y\|$. De la misma forma, $\|y\| - \|x\| \leq \|y - x\| = \|x - y\|$, y obtenemos así (2.2).

Recordemos también que toda norma induce una distancia:

$$d(x, y) = \|y - x\|. \quad (2.3)$$

En este capítulo trabajaremos especialmente considerando $X = \mathcal{C}([a, b])$ en la Definición 2.2.1.

Las siguientes definiciones serán fundamentales para demostrar la cuestión de existencia y unicidad en un caso particular de problema de mejor aproximación que consideramos justo a continuación.

Definición 2.2.3 Sea X un espacio vectorial. Si x_1 y x_2 son dos elementos distintos de X , entonces el conjunto de todos los elementos de la forma $x = tx_1 + (1-t)x_2$, $0 \leq t \leq 1$, se conoce como el segmento que une a x_1 y x_2 .

Definición 2.2.4 Sea X un espacio vectorial. Un subconjunto $C \subseteq X$ es convexo si C contiene todos los elementos del segmento que une cualesquiera dos elementos de sí mismo. Es decir, si $x_1, x_2 \in C$, entonces $x = tx_1 + (1-t)x_2 \in C$, $0 \leq t \leq 1$.

Definición 2.2.5 Sea $(X, \|\cdot\|)$ un espacio vectorial normado. Se dice que X es estrictamente convexo si para todo $x_1, x_2 \in X$, con $x_1 \neq x_2$, se cumple que si $\|x_1\| = 1$ y $\|x_2\| = 1$, entonces $\left\|\frac{x_1+x_2}{2}\right\| < 1$.

Ejemplo 2.2.4 ($L^\infty[a, b], \|\cdot\|_\infty$) no es estrictamente convexo. Tomemos, por ejemplo, $f(x) = x$, $g(x) = x^2$ y $[a, b] = [0, 1]$. Tenemos que $\|f\|_\infty = \|g\|_\infty = 1$ y

$$\|f + g\|_\infty = \|x + x^2\|_\infty = \sup_{x \in [0, 1]} |x + x^2| = 2.$$

Como $f, g \in \mathcal{C}[0, 1]$, ($\mathcal{C}([a, b]), \|\cdot\|_\infty$) tampoco es estrictamente convexo.

Ejemplo 2.2.5 ($\mathbb{R}, \|\cdot\|$) es estrictamente convexo. Sean $x, y \in \mathbb{R}$ con $\|x\| = \|y\| = 1$ y $x \neq y$. No queda otra opción que $x = 1$ e $y = -1$ o viceversa. Entonces $\|x + y\| = 0$.

Ejemplo 2.2.6 De la Definición 2.2.2, ($L^p[a, b], \|\cdot\|_p$) no es estrictamente convexo para $p = 1$. Si tomamos $f = 2x$, $g = 3x^2$, tenemos que $f, g \in L[0, 1]$ y $\|f\|_1 = \|g\|_1 = 1$. Entonces $\|f + g\|_1 = \left(\int_0^1 |2x + 3x^2| dx\right) = 2$.

Ejemplo 2.2.7 Los espacios ($L^p[a, b], \|\cdot\|_p$) son estrictamente convexos, para $1 < p < \infty$.

DEMOSTRACIÓN. Sean $f, g \in L^p[a, b]$ tales que $\|f\|_p = \|g\|_p = 1$ y $p > 1$. Entonces

$$\begin{aligned} \left\| \frac{f+g}{2} \right\|_p &< \left\| \frac{f+g}{2} \right\|_p^p = \left(\frac{1}{2}\right)^p \int_a^b |(f+g)(x)|^p dx \\ &\leq \left(\frac{1}{2}\right)^p \left[\left(\int_a^b |f(x)|^p dx\right)^{\frac{1}{p}} + \left(\int_a^b |g(x)|^p dx\right)^{\frac{1}{p}} \right]^p = \frac{2^p}{2^p} = 1. \end{aligned}$$

En la última desigualdad se usó la Desigualdad de Minkowski, demostrada en el Apéndice A.2. □

Sean $(X, \|\cdot\|)$ un espacio vectorial normado y x_1, \dots, x_n n elementos de X linealmente independientes. Sea $y \in X$ un elemento adicional. El problema de mejor aproximación lineal consiste en buscar una combinación lineal de x_1, \dots, x_n que aproxime a y de forma que $\|y - (a_1x_1 + \dots + a_nx_n)\|$ sea mínima.

Definición 2.2.6 Una mejor aproximación de y por combinaciones lineales de x_1, \dots, x_n es un elemento $a_1x_1 + \dots + a_nx_n$, para el cual

$$\|y - (a_1x_1 + \dots + a_nx_n)\| \leq \|y - (b_1x_1 + \dots + b_nx_n)\|$$

para cualquier otra combinación de escalares b_1, \dots, b_n .

Teorema 2.2.1 Sea $(X, \|\cdot\|)$ un espacio vectorial normado, $y \in X$ y n elementos linealmente independientes x_1, \dots, x_n de X . El problema de encontrar

$$\min_{a_i} \|y - (a_1x_1 + \dots + a_nx_n)\|$$

tiene solución.

DEMOSTRACIÓN.

Consideremos la norma del error

$$d(a_1, a_2, \dots, a_n) = \| y - (a_1x_1 + a_2x_2 + \dots + a_nx_n) \|$$

como una función real o compleja en las variables a_1, \dots, a_n . Esta función es continua en las variables a_i , como se deduce de lo siguiente:

$$\begin{aligned} & |d(a'_1, a'_2, \dots, a'_n) - d(a_1, a_2, \dots, a_n)| \\ &= | \| y - (a'_1x_1 + a'_2x_2, \dots + a'_nx_n) \| - \| y - (a_1x_1 + a_2x_2, \dots + a_nx_n) \| | \\ &\leq \| (a'_1 - a_1)x_1 + \dots + (a'_n - a_n)x_n \| \\ &\leq |a'_1 - a_1| \cdot \| x_1 \| + \dots + |a'_n - a_n| \cdot \| x_n \| . \end{aligned}$$

La primera desigualdad se cumple por (2.2), y la segunda por la definición de norma. Este resultado nos muestra que $|d(a'_1, a'_2, \dots, a'_n) - d(a_1, a_2, \dots, a_n)|$ se hace tan pequeño como se desee, siempre que los coeficientes a_1, \dots, a_n y a'_1, \dots, a'_n estén lo suficientemente próximos entre sí, dado que los x_1, \dots, x_n son valores fijos. De manera análoga se prueba que la función

$$h(a_1, \dots, a_n) = \| a_1x_1 + \dots + a_nx_n \|$$

es continua en las variables a_i , $i = 1, \dots, n$. Denotemos por S a la superficie esférica $S : |a_1|^2 + |a_2|^2 + \dots + |a_n|^2 = 1$ en \mathbb{R}_n (o \mathbb{C}_n). Como S es cerrada y acotada, h tiene un mínimo en S (por ser h una función continua en las variables a_1, \dots, a_n), al que llamaremos $m > 0$. La posibilidad $m = 0$ no se puede dar, pues si $\| a_1x_1 + \dots + a_nx_n \| = 0$, se sigue, por definición de norma, que $a_1x_1 + \dots + a_nx_n = 0$. Como x_1, \dots, x_n son linealmente independientes, $a_1 = \dots = a_n = 0$, lo que contradice que pertenezcan a S . Ahora, escribiendo $r = (|a_1|^2 + \dots + |a_n|^2)^{\frac{1}{2}}$,

$$h(a_1, \dots, a_n) = r \left\| \frac{a_1}{r}x_1 + \dots + \frac{a_n}{r}x_n \right\| \geq rm.$$

Además,

$$d = \| y - (a_1x_1 + \dots + a_nx_n) \| \geq \| a_1x_1 + \dots + a_nx_n \| - \| y \| \geq rm - \| y \|. \quad (2.4)$$

Esta inecuación implica que $d \rightarrow \infty$, cuando $r \rightarrow \infty$, lo cual quiere decir que podemos centrar nuestra búsqueda del mínimo valor de la función d en una bola cerrada. Más precisamente, si

$$\rho = \inf d(a_1, \dots, a_n), \quad R = \frac{1 + \rho + \| y \|}{m},$$

y $|a_1|^2 + |a_2|^2 + \cdots + |a_n|^2 > R^2$, entonces (2.4) implica que $d(a_1, \dots, a_n) > mR - \|y\| = 1 + \rho > \rho$. De este modo, d alcanza el ínfimo porque está acotada inferiormente y es continua, como vimos al principio de la prueba. Si denotamos por I a todo el espacio al que pertenecen las a 's, y por II el subconjunto de I donde $|a_1|^2 + |a_2|^2 + \cdots + |a_n|^2 \leq R^2$, entonces

$$\inf_I d(a_1, \dots, a_n) = \inf_{II} d(a_1, \dots, a_n) = \min_{II} d(a_1, \dots, a_n),$$

obteniendo así el resultado deseado. □

A las soluciones $a_1x_1 + \cdots + a_nx_n$ se les denomina mejores aproximantes a y . Veremos algunos ejemplos a modo ilustrativo:

Ejemplo 2.2.8 Sea $f \in C([a, b])$ y n un entero fijado. El problema de encontrar

$$\min_{a_0, \dots, a_n} \max_{a \leq x \leq b} |f(x) - (a_0 + a_1x + \cdots + a_nx^n)|$$

tiene solución. Como veremos más adelante, también es única, y se denomina mejor aproximación de Chebyshev de grado menor o igual que n a la función f .

Ejemplo 2.2.9 En las mismas condiciones del ejemplo anterior, si $p \geq 1$ entonces el problema de encontrar

$$\min_{a_0, a_1, \dots, a_n} \int_a^b |f(x) - (a_0 + a_1x + \cdots + a_nx^n)|^p dx$$

tiene solución. Tal problema nos ayuda a encontrar el mejor polinomio aproximante a f en la norma p . La hipótesis inicial se puede debilitar a $f \in L^p[a, b]$.

Ejemplo 2.2.10 Sean x_0, \dots, x_k $k+1$ puntos distintos entre sí, con $k \geq n$. Los siguientes problemas tienen solución:

1.

$$\min_{a_0, a_1, \dots, a_n} \max_{0 \leq i \leq k} |f(x_i) - (a_0 + a_1x_i + \cdots + a_nx_i^n)| \quad y$$

2.

$$\min_{a_0, a_1, \dots, a_n} \max_{0 \leq i \leq k} (f(x_i) - (a_0 + a_1x_i + \cdots + a_nx_i^n))^2.$$

Definición 2.2.7 Para $y; x_1, \dots, x_n$ dados,

$$\min_{a_i} \|y - (a_1x_1 + \cdots + a_nx_n)\| = E_n(y; x_1, \dots, x_n) = E_n(y)$$

es el error de la mejor aproximación cuando y es aproximado por combinaciones lineales de las x 's.

Geoméricamente, puede ser visto como la distancia de y al subespacio generado por x_1, \dots, x_n . Obviamente, se tiene que $E_1(y) \geq E_2(y) \geq E_3(y) \geq \dots$, porque las combinaciones lineales de x_1, \dots, x_n son un caso particular de combinaciones lineales de x_1, \dots, x_n, x_{n+1} . Ya vimos en el Teorema 2.2.1 que el problema de mejor aproximación tiene solución, pero las soluciones no tienen por qué ser únicas. De hecho, el conjunto de soluciones forman un conjunto convexo.

Teorema 2.2.2 *Sea S el conjunto de los mejores aproximantes a y , en las mismas condiciones del Teorema 2.2.1. Entonces S es convexo.*

DEMOSTRACIÓN. Sean x y w dos mejores aproximantes a y , entonces $E_n(y) = \|y - x\| = \|y - w\|$. Supongamos además que $\alpha \geq 0$, $\beta \geq 0$ y $\alpha + \beta = 1$. Entonces

$$\begin{aligned} \|y - (\alpha x + \beta w)\| &= \|\alpha(y - x) + \beta(y - w)\| \leq \alpha \|y - x\| + \beta \|y - w\| \\ &= (\alpha + \beta)E_n(y) = E_n(y). \end{aligned}$$

Como $\alpha x + \beta w$ es una combinación lineal de x_1, \dots, x_n , concluimos que también es un mejor aproximante a y .

□

Corolario 2.2.1 *El conjunto de los mejores aproximantes contiene o un único elemento, o un número infinito de elementos.*

DEMOSTRACIÓN. Como la solución es un conjunto convexo, o tendrá un único elemento, o en caso de tener dos, tendrá también al segmento que los une.

□

Teorema 2.2.3 *Sea $(X, \|\cdot\|)$ un espacio vectorial normado estrictamente convexo. Entonces el problema de mejor aproximación planteado en el Teorema 2.2.1 tiene solución única.*

DEMOSTRACIÓN. Supongamos que x_1 y x_2 son dos mejores aproximantes a y , con $x_1 \neq x_2$. Entonces $\|y - x_1\| = \|y - x_2\| = E_n(y)$. Tenemos que $y - x_1 \neq y - x_2$, y por convexidad estricta, se tiene que $\|(y - x_1) + (y - x_2)\| < 2E_n(y)$. Esto es equivalente a $\|y - \frac{1}{2}(x_1 + x_2)\| < E_n(y)$. Pero esto es un absurdo, pues significaría que el elemento $\frac{1}{2}(x_1 + x_2)$, que es una combinación lineal de x_1, \dots, x_n es mejor aproximante que x_1 y x_2 .

□

Corolario 2.2.2 *Por el Teorema 2.2.3 y el Ejemplo 2.2.7, el problema de mejor aproximación en $(L^p[a, b], \|\cdot\|_p)$ tiene solución única para $1 < p < \infty$. También, del Ejemplo 2.2.4, se sigue que no se puede concluir la unicidad a partir del Teorema 2.2.3, si $p = \infty$.*

El Teorema 2.2.3 presenta una condición suficiente (ser estrictamente convexo), pero no necesaria para que la solución del problema de mejor aproximación sea única, es decir, no es una caracterización.

El teorema que veremos a continuación no pide la condición de convexidad estricta, y funciona para el espacio $X = \mathbb{P}_n$, con n fijada de antemano. La demostración puede verse en [4, Teorema 7.5.6].

Teorema 2.2.4 (Tonelli). *Sea S un conjunto cerrado y acotado en el plano complejo, que contiene más de $n + 1$ puntos y f una función continua en S . Sea*

$$M = \min_{p \in \mathbb{P}_n} \max_{z \in S} |f(z) - p(z)|, \quad (2.5)$$

$p_n(z)$ el polinomio que cumple tal valor para M y $r(z) = f(z) - p_n(z)$. Entonces

1. El número de puntos distintos de S en los que $|r(z)|$ toma su valor máximo es mayor que $n + 1$.
2. La solución de (2.5) es única.

La obtención de la solución del Teorema 2.2.4, conocida como mejor aproximación uniforme (Chebyshev) a funciones continuas, es en general un problema complejo, y solo para los casos $n = 0, 1$ se puede obtener una solución explícita (véase [4, Sección 7.6]). Existe un proceso iterativo, conocido como algoritmo de Remez, que permite obtener una solución aproximada. Este método está inspirado en el siguiente (véase [4, Teorema 7.6.2])

Teorema 2.2.5 (Equioscilación de Chebyshev). *Sea $f \in \mathcal{C}([a, b])$. Supongamos que existe $q_n \in \mathbb{P}_n$ tal que $E_n(f) = \max_{a \leq x \leq b} |f(x) - q_n(x)| = \delta$, y que existen, como mínimo, $n + 2$ puntos $a \leq x_1 \leq \dots \leq x_{n+2} \leq b$, donde $q(x_i) - f(x_i) = \pm\delta$, alternando signo en nodos consecutivos. Entonces $E_n(f) = \delta$, es decir, q es la mejor aproximación uniforme de f en \mathbb{P}_n .*

DEMOSTRACIÓN. Inicialmente, sabemos que $E_n(f) \leq \delta$. Supongamos, por reducción al absurdo, que $E_n(f) < \delta$, es decir, q_n no es la mejor aproximación de f en \mathbb{P}_n . Sea $p \in \mathbb{P}_n$ la mejor aproximación de f , con $p \neq q_n$. Entonces

$$\begin{aligned} q(x_i) - p(x_i) &= (q(x_i) - f(x_i)) - (p(x_i) - f(x_i)) \\ &= \pm\delta - t_i, \quad |t_i| < \delta, \quad \forall 1 \leq i \leq n + 2. \end{aligned}$$

Entonces $q - p \in \mathbb{P}_n$ tiene, como mínimo, $n + 2$ alternancias de signo. Por el Teorema fundamental del Álgebra, tenemos que $q \equiv p$, lo cual es un absurdo.

□

2.3. Aproximación por mínimos cuadrados

En esta sección abordaremos el método de aproximación más comúnmente usado y desarrollado: el método de aproximación por mínimos cuadrados. Este proceso es un caso particular del abordado más generalmente en la sección anterior: cuando la norma proviene de un producto interior. A diferencia del problema de mejor aproximación uniforme a funciones continuas, cuando la norma proviene de un producto interior es posible obtener una solución explícita de la solución. Empezaremos con la definición de producto interior y nociones básicas antes de abordar a este problema de aproximación.

Definición 2.3.1 Sea X un espacio vectorial real X . Una aplicación $\langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{R}$ se dice que es un producto interior si $\forall x_1, x_2, x_3 \in X, \forall \alpha \in \mathbb{R}$ (o $\alpha \in \mathbb{C}$) se cumplen las siguientes propiedades:

- (a) $\langle x_1 + x_2, x_3 \rangle = \langle x_1, x_3 \rangle + \langle x_2, x_3 \rangle$ (Linealidad).
- (b) $\langle x_1, x_2 \rangle = \langle x_2, x_1 \rangle$ (Simetría).
- (c) $\langle \alpha x_1, x_2 \rangle = \alpha \langle x_1, x_2 \rangle$ (Homogeneidad).
- (d) $\langle x, x \rangle \geq 0$. $\langle x, x \rangle = 0$, si y solo si, $x = 0$ (Positividad).

A la cantidad $\langle x_1, x_2 \rangle$ se le llama producto interior de x_1 con x_2 . En el caso de que el cuerpo de escalares sea \mathbb{C} , consideraremos la Simetría Hermitiana: $\langle x_1, x_2 \rangle = \overline{\langle x_2, x_1 \rangle}$. Todo producto interior induce una norma: $\|x\| = \sqrt{\langle x, x \rangle}$, $\forall x \in X$.

Definición 2.3.2 Al espacio $(X, \langle \cdot, \cdot \rangle)$ se le denomina espacio pre-Hilbert. Se dice que X es un espacio de Hilbert, si es pre-Hilbert y es completo, es decir, que toda sucesión de Cauchy de elementos de X converge en X .

Para ver ejemplos de espacios pre-Hilbert, necesitaremos probar primero algunos resultados previos:

Proposición 2.3.1 En todo espacio pre-Hilbert X se cumple la Ley del Paralelogramo:

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2), \quad \forall x, y \in X. \quad (2.6)$$

DEMOSTRACIÓN.

$$\begin{aligned} \|x + y\|^2 + \|x - y\|^2 &= \langle x + y, x + y \rangle + \langle x - y, x - y \rangle \\ &= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 + \|x\|^2 - 2\langle x, y \rangle \\ &= 2(\|x\|^2 + \|y\|^2). \end{aligned}$$

La segunda igualdad se cumple por la propiedad de linealidad del producto interior, y la tercera por la propiedad de simetría.

□

Proposición 2.3.2 (*Desigualdad de Cauchy-Schwarz*). Sea X un espacio pre-Hilbert. Para todo $x, y \in X$ se cumple que $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$. La igualdad se da si y solo si x e y son linealmente dependientes.

DEMOSTRACIÓN. Sean X un espacio pre-Hilbert, $x, y \in X$ y $\lambda \in \mathbb{R}$. Por la positividad del producto interior, sabemos que $0 \leq \|x + \lambda y\|^2$. Por tanto

$$0 \leq \|x + \lambda y\|^2 = \langle x, x \rangle + 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle.$$

Si $y = 0$, la desigualdad se cumple trivialmente. Supongamos que $y \neq 0$ y tomemos $\lambda = -\frac{\langle x, y \rangle}{\langle y, y \rangle}$. Entonces $0 \leq \langle x, x \rangle - \frac{\langle x, y \rangle^2}{\langle y, y \rangle}$. Por tanto

$$\langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle = \|x\|^2 \cdot \|y\|^2 \Rightarrow |\langle x, y \rangle| \leq \|x\| \cdot \|y\|. \quad (2.7)$$

Si x, y fueran linealmente dependientes, entonces se cumple trivialmente la igualdad en la primera desigualdad de (2.7).

□

Nota 2.3.1 En $(L^p[a, b], \|\cdot\|_p)$, la Desigualdad de Cauchy-Schwarz coincide con la Desigualdad de Hölder (véase Anexo A.2).

Ejemplo 2.3.1 $(\mathbb{R}^n, \|\cdot\|_2)$ es un espacio pre-Hilbert con la norma inducida por el producto interior $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$.

Ejemplo 2.3.2 Sean $w_i \geq 0, i = 1, \dots, n$. $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ es un espacio pre-Hilbert con $\langle x, y \rangle = \sum_{i=1}^n w_i x_i y_i$, para todo $x, y \in \mathbb{R}^n$.

Ejemplo 2.3.3 $(\mathcal{C}([a, b]), \|\cdot\|_\omega)$ es un espacio pre-Hilbert con la norma cuadrática pesada

$$\langle f, g \rangle_\omega = \int_a^b f(x)g(x)\omega(x)dx, \quad (2.8)$$

para todo $f, g \in \mathcal{C}([a, b])$ y $\omega > 0$ una función peso definida en $[a, b]$.

Ejemplo 2.3.4 $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$ no es un espacio con producto interior, pues no se cumple la Ley del Paralelogramo: Si tomamos $f(x) = x$ y $g(x) = 1$, entonces no se verifica (2.6). Esto quiere decir que no puede existir un producto interior para el cual $\|\cdot\|_\infty$ sea su norma inducida.

Definición 2.3.3 Sea S un subconjunto de un espacio pre-Hilbert H .

- S es un conjunto linealmente independiente si todos sus elementos son linealmente independientes entre sí, dos a dos.

- $x, y \in H$ son ortogonales, y se denota como $x \perp y$, si $\langle x, y \rangle = \langle y, x \rangle = 0$.
- S es ortonormal si $\langle x, y \rangle = \delta_{x,y}$, $x, y \in S$, donde $\delta_{x,y}$ es el símbolo de la delta de Kronecker que se define como

$$\delta_{x,y} = \begin{cases} 0, & x \neq y, \\ 1, & x = y. \end{cases}$$

Si solamente tenemos que $\langle x, y \rangle = 0$, para $x \neq y$, entonces S es ortogonal.

Teorema 2.3.1 (Teorema de Pitágoras). Sean x_1, \dots, x_n elementos ortogonales entre sí. Entonces,

$$\|x_1 + x_2 + \dots + x_n\|^2 = \|x_1\|^2 + \|x_2\|^2 + \dots + \|x_n\|^2.$$

DEMOSTRACIÓN.

$$\|x_1 + x_2 + \dots + x_n\|^2 = \sum_{i=1}^n \sum_{j=1}^n \langle x_i, x_j \rangle = \sum_{i=1}^n \|x_i\|^2.$$

□

Lema 2.3.1 Cualquier conjunto finito de elementos ortogonales $\{x_1, \dots, x_n\}$, con $x_i \neq 0$, $i = 1, \dots, n$, es linealmente independiente.

DEMOSTRACIÓN. Supongamos que $a_1x_1 + \dots + a_nx_n = 0$, donde existe $k \in \{1, \dots, n\}$ tal que $a_k \neq 0$. Entonces

$$0 = \langle 0, x_k \rangle = \langle a_1x_1 + \dots + a_nx_n, x_k \rangle = a_k \langle x_k, x_k \rangle \Rightarrow \|x_k\| = 0,$$

lo cual es una contradicción, dado que $x_k \neq 0$.

□

Este teorema no es una caracterización. Un conjunto independiente no es necesariamente ortogonal, pero sí es cierto que puede ser ortogonalizado, es decir, podemos encontrar un conjunto ortogonal (u ortonormal) que genera el mismo espacio de combinaciones lineales. En este caso omitimos la demostración por ser un resultado bien conocido.

Teorema 2.3.2 (Gram-Schmidt). Sea x_1, x_2, \dots una sucesión finita o no finita de elementos, tal que cualquier subconjunto finito $\{x_1, \dots, x_k\}$ es siempre linealmente independiente. Entonces, podemos encontrar constantes $a_{i,j}$, tales que los elementos

$$x_i^* = \sum_{j=1}^i a_{ij}x_j, \quad i = 1, 2, \dots,$$

constituyen un conjunto ortogonal u ortonormal.

Supongamos que el espacio pre-Hilbert es $\mathcal{C}([a, b])$ con el producto interior $\langle \cdot, \cdot \rangle_\omega$ dado en (2.8), y la sucesión de elementos independientes es la familia de los monomios $\{1, x, x^2, \dots, x^n\}$. Aplicando el proceso de ortogonalización de Gram-Schmidt, obtenemos una familia de polinomios $\{p_i\}$, $i \geq 0$, verificando que:

- el grado de p_i es i , $i = 0, 1, \dots, n$.
- $\langle p_i, p_j \rangle = k_{i,j} \delta_{i,j}$, $i, j = 0, 1, \dots, n$, $k_{i,j} > 0$.

A esta familia $\{p_i\}_{i=0}^n$ se le conoce como la familia de polinomios ortogonales u ortonormales con respecto a la función peso ω .

Si el proceso se repite para todo $n \geq 0$, obtenemos una familia de polinomios ortogonales, que es única salvo constante multiplicativa, la cual se puede elegir de manera que la familia sea mónica u ortonormal.

Algunas propiedades de polinomios ortogonales, que añadiremos a continuación, no serán demostradas, pues excede en la brevedad intencionada para este capítulo y ya han sido abordadas a lo largo del Grado en Matemáticas.

Proposición 2.3.3 *Todo polinomio ortogonal de grado n con respecto a ω en $[a, b]$ posee n ceros reales, distintos y situados en (a, b) .*

Proposición 2.3.4 *La familia $\{p_n\}_{n=0}^\infty$ de polinomios ortogonales mónicos puede computarse recursivamente mediante la ley de recurrencia a tres términos:*

$$p_{n+1} = (x - \alpha_n)p_n(x) - \beta_n p_{n-1}(x), \quad \forall n \geq 0$$

siendo $p_{-1} \equiv 0$, $p_0 \equiv 1$ y $\forall n \geq 0$, $\forall m \geq 1$,

$$\alpha_n = \frac{\langle xp_n(x), p_n(x) \rangle_\omega}{\langle p_n(x), p_n(x) \rangle_\omega}, \quad \beta_m = \frac{\langle p_m(x), p_m(x) \rangle_\omega}{\langle p_{m-1}(x), p_{m-1}(x) \rangle_\omega} > 0.$$

De la ley de recurrencia para la familia de polinomios mónicos se obtiene la correspondiente para los polinomios ortonormales.

Proposición 2.3.5 *La familia $\{P_n\}_{n=0}^\infty$ de polinomios ortonormales puede computarse recursivamente mediante la ley de recurrencia a tres términos:*

$$P_{n+1}(x) = (\gamma_n x - \delta_n) P_n(x) - \eta_n P_{n-1}(x),$$

siendo

$$\gamma_n = \sqrt{\frac{\langle p_n, p_n \rangle_\omega}{\langle p_{n+1}, p_{n+1} \rangle_\omega}}, \quad \delta_n = \frac{\langle xp_n, p_n \rangle_\omega}{\sqrt{\langle p_n, p_n \rangle_\omega \langle p_{n+1}, p_{n+1} \rangle_\omega}},$$

$$\eta_n = \frac{\langle p_{n-1}, p_{n-1} \rangle_\omega}{\sqrt{\langle p_n, p_n \rangle_\omega \langle p_{n+1}, p_{n+1} \rangle_\omega}}.$$

Obsérvese que de las Proposiciones 2.3.4-2.3.5 podemos computar eficientemente una familia de polinomios ortogonales con respecto a ω en $[a, b]$ sin recurrir al proceso de ortogonalización de Gram-Schmidt, simplemente a partir de la familia de momentos $c_k = \int_a^b x^k \omega(x) dx$, $k \geq 0$.

Proposición 2.3.6 Si $\alpha \in \mathbb{R}$, $[a, b] = [-\alpha, \alpha]$ y $\omega(x) = \omega(-x)$, entonces p_{2n} solo tiene potencias pares y p_{2n-1} solo potencias impares.

Tras todas estas consideraciones previas, pasamos a resolver el problema de mejor aproximación en mínimos cuadrados. Situémonos en el espacio pre-Hilbert $X = \mathcal{C}([a, b])$, con el producto interior $\langle f, g \rangle_\omega = \int_a^b f(x)g(x)\omega(x)dx$, y denotemos $\|f\|_\omega = \sqrt{\langle f, f \rangle_\omega}$. Sean $\mathbb{E} = \text{span}\{f_1, f_2, \dots, f_n\}$ un subespacio vectorial de $\mathcal{C}([a, b])$ y $f \in \mathcal{C}([a, b])$. El problema consiste en buscar $\varphi = \sum_{j=1}^n c_j f_j \in \mathbb{E}$ que minimice $E[\varphi] = \|f - \varphi\|_\omega$.

$$\begin{aligned} E^2[\varphi] &= \|f - \varphi\|_\omega^2 = \langle f - \varphi, f - \varphi \rangle_\omega = \|f\|_\omega^2 + \|\varphi\|_\omega^2 - 2\langle f, \varphi \rangle_\omega \\ &= \int_a^b f^2(t)\omega(t)dt + \int_a^b \left(\sum_{j=1}^n c_j f_j(t) \right)^2 \omega(t)dt - 2 \int_a^b \left(\sum_{j=1}^n c_j f_j(t) \right) f(t)\omega(t)dt \\ &= \int_a^b f^2(t)\omega(t)dt + \int_a^b \left(\sum_{j=1}^n c_j f_j(t) \right)^2 \omega(t)dt - 2 \sum_{j=1}^n c_j \int_a^b f_j(t)f(t)\omega(t)dt. \end{aligned}$$

Al ser $E[\varphi] = E\left[\sum_{j=1}^n c_j f_j\right]$ una función de n variables reales $\{c_1, \dots, c_n\}$, para encontrar los posibles puntos críticos debemos resolver $\nabla E = 0$. Esto es equivalente a buscar los puntos críticos de E^2 . Así,

$$\frac{\partial E^2[\varphi]}{\partial c_i} = -2 \int_a^b f_i(t)f(t)\omega(t)dt + 2 \int_a^b \left(\sum_{j=1}^n c_j f_j(t) \right) f_i(t)\omega(t)dt$$

Por tanto, se tiene que $\frac{\partial E^2[\varphi]}{\partial c_i} = 0$, si y solo si $\sum_{j=1}^n \langle f_i, f_j \rangle_\omega c_j = \langle f_i, f \rangle_\omega$. Tenemos entonces que el (único) punto crítico será la solución $C = (c_1 \dots c_n)^T$ del sistema lineal

$$G_n \cdot C = B, \tag{2.9}$$

donde

$$G_n = \begin{pmatrix} \langle f_1, f_1 \rangle_\omega & \langle f_2, f_1 \rangle_\omega & \cdots & \langle f_n, f_1 \rangle_\omega \\ \langle f_1, f_2 \rangle_\omega & \langle f_2, f_2 \rangle_\omega & \cdots & \langle f_n, f_2 \rangle_\omega \\ \vdots & \vdots & \ddots & \vdots \\ \langle f_1, f_n \rangle_\omega & \langle f_2, f_n \rangle_\omega & \cdots & \langle f_n, f_n \rangle_\omega \end{pmatrix}, \quad B = \begin{pmatrix} \langle f_1, f \rangle_\omega \\ \langle f_2, f \rangle_\omega \\ \vdots \\ \langle f_n, f \rangle_\omega \end{pmatrix}.$$

A este sistema se le conoce como el sistema de ecuaciones normales, y a G_n como la Matriz de Gram (de orden n), que es real, simétrica y definida positiva:

$$\begin{aligned} X^T \cdot G_n \cdot X &= (x_1, \dots, x_n) G_n \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \langle f_i, f_j \rangle_\omega \\ &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i f_i, x_j f_j \rangle_\omega = \left\| \sum_{k=1}^n x_k f_k \right\|_\omega^2 \geq 0. \end{aligned}$$

Además,

$$\left\| \sum_{k=1}^n x_k f_k \right\|_\omega^2 = 0 \Leftrightarrow \left\| \sum_{k=1}^n x_k f_k \right\| = 0 \Leftrightarrow x_k = 0, \forall k = 1, \dots, n \Leftrightarrow X \equiv 0,$$

ya que $\{f_i\}_{i=1}^n$ es una familia de funciones linealmente independientes. Se concluye entonces que $\det(G_n) \neq 0$, y el sistema (2.9) tiene solución única, algo que ya sabíamos por ser $(X, \langle \cdot, \cdot \rangle)$ un espacio estrictamente convexo. Por tanto, la función $E^2[\varphi]$ admite un único punto crítico. Veamos que es mínimo. La matriz Hessiana de $E^2[\varphi]$ será

$$H(E^2[\varphi]) = \begin{pmatrix} \frac{\partial^2 E^2[\varphi]}{\partial c_1^2} & \frac{\partial^2 E^2[\varphi]}{\partial c_1 \partial c_2} & \dots & \frac{\partial^2 E^2[\varphi]}{\partial c_1 \partial c_n} \\ \frac{\partial^2 E^2[\varphi]}{\partial c_2 \partial c_1} & \frac{\partial^2 E^2[\varphi]}{\partial c_2^2} & \dots & \frac{\partial^2 E^2[\varphi]}{\partial c_2 \partial c_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E^2[\varphi]}{\partial c_n \partial c_1} & \frac{\partial^2 E^2[\varphi]}{\partial c_n \partial c_2} & \dots & \frac{\partial^2 E^2[\varphi]}{\partial c_n^2} \end{pmatrix}.$$

Como $\frac{\partial^2 E^2[\varphi]}{\partial c_i \partial c_j} = 2\langle f_i, f_j \rangle_\omega$, se tiene que $H(E^2[\varphi]) = 2 \cdot G_n$, que es definida positiva, y por tanto, C es un mínimo. Obsérvese que todo este desarrollo es igualmente válido si hubiéramos partido de un espacio pre-Hilbert general $(X, \langle \cdot, \cdot \rangle)$ en general. Sin embargo, nuestro principal interés será considerar el caso particular $X = \mathcal{C}([a, b])$ y $\langle f, g \rangle = \int_a^b f(x)g(x)\omega(x)dx$. Vemos por tanto que, a diferencia de lo que ocurría en la sección anterior, el problema de mejor aproximación queda totalmente caracterizado cuando la norma del espacio vectorial normado proviene de un producto interior.

El proceso anterior presenta, sin embargo, dos inconvenientes:

- El sistema podría estar mal condicionado, es decir, pequeños errores en los datos podrían generar grandes errores en la solución.

Por ejemplo, si $w \equiv 1$, en $[a, b] = [0, 1]$, $E = \mathbb{P}_n$ y $f_i = t^{i-1}$, $1 \leq i \leq n+1$, se cumple

$$\langle f_i, f_j \rangle_\omega = \int_0^1 t^{i+j-2} dt = \frac{1}{i+j-1}.$$

La matriz de Gram que resulta se conoce como matriz de Hilbert, que es el prototipo de matriz mal condicionada. A modo de ejemplo, si $n = 10$, entonces $\text{cond}(G_{10}) \approx 1'6 \cdot 10^{13}$.

- El proceso no es dinámico. Cada coeficiente c_i depende del número de elementos de la base. Si extendemos el subespacio E con el fin de mejorar la aproximación, no nos servirán todos los cálculos realizados previamente.

Para solucionar estos inconvenientes, podemos ortogonalizar la base de partida $\{f_1, \dots, f_n\}$ por el proceso de ortogonalización de Gram-Schmidt. Entonces la matriz de Gram sería diagonal y la solución del sistema de ecuaciones normales es por tanto inmediata:

$$c_j = \frac{\langle f_j, f \rangle_\omega}{\langle f_j, f_j \rangle_\omega}, \forall j = 1, \dots, n.$$

Por tanto, la solución al problema inicial de mejor aproximación viene dada explícitamente por

$$\varphi(t) = \sum_{j=1}^n \frac{\langle f_j, f \rangle_\omega}{\langle f_j, f_j \rangle_\omega} f_j(t).$$

Los coeficientes no dependen de n , sino de f y la base $\{f_1, \dots, f_n\}$, por lo que se soluciona el inconveniente del dinamismo del proceso. Si la base fuera ortonormal, a los coeficientes $c_j = \langle f_j, f \rangle_\omega$ se les denomina Coeficientes de Fourier de f con respecto a la base ortonormal $\{f_1, \dots, f_n\}$.

En el caso particular de $X = \mathcal{C}([a, b])$, $\langle f, g \rangle_\omega = \int_a^b f(x)g(x)\omega(x)dx$, partiendo del sistema linealmente independiente de monomios $\{1, x, x^2, \dots, x^n\}$, obtendríamos a partir de la ley de recurrencia a tres términos la correspondiente familia de polinomios ortonormales, y de ahí, al computar los coeficientes de Fourier, obtendríamos directamente la solución a nuestro problema. Continuamos con una expresión del error cometido en la aproximación:

Teorema 2.3.3 Sean x_1, x_2, \dots, x_n elementos linealmente independientes. Si

$$E_n(y) = \min_{a_i} \| y - (a_1x_1 + a_2x_2 + \dots + a_nx_n) \|,$$

entonces,

$$E_n^2(y) = \frac{g(x_1, x_2, \dots, x_n, y)}{g(x_1, x_2, \dots, x_n)},$$

denotando como $g(x_1, \dots, x_n)$ al determinante de la matriz de Gram para los elementos x_1, \dots, x_n .

Se omite la demostración por excederse del temario necesario para esta memoria. Puede encontrarse en [4, Teorema 8.7.4].

Proposición 2.3.7 Sean el espacio pre-Hilbert $(X, \langle \cdot, \cdot \rangle_\omega)$, $\mathbb{E} = \text{span}\{f_1, \dots, f_n\}$ un subespacio vectorial de X y $f \in X$. Si $\varphi = \sum_{j=1}^n c_j f_j \in \mathbb{E}$ es la solución al

problema de mejor aproximación de f , entonces el error de esta aproximación es

$$\|f - \varphi\|_{\omega} = \sqrt{\|f\|_{\omega}^2 - \sum_{k=1}^n c_k^2 \|f_k\|_{\omega}^2}.$$

Veremos a continuación una propiedad minimal importante de las series truncadas de Fourier.

Teorema 2.3.4 Sea $(X, \langle \cdot, \cdot \rangle)$ un espacio pre-Hilbert, x_1^*, x_2^*, \dots un sistema ortonormal e $y \in X$ un elemento arbitrario. Entonces

$$\left\| y - \sum_{i=1}^N \langle y, x_i^* \rangle x_i^* \right\| \leq \left\| y - \sum_{i=1}^N a_i x_i^* \right\| \quad (2.10)$$

para cualesquiera elementos a_1, a_2, \dots, a_N .

DEMOSTRACIÓN.

$$\begin{aligned} \left\| y - \sum_{i=1}^N a_i x_i^* \right\|^2 &= \langle y - \sum_{i=1}^N a_i x_i^*, y - \sum_{i=1}^N a_i x_i^* \rangle \\ &= \langle y, y \rangle - \sum_{i=1}^N a_i \langle x_i^*, y \rangle - \sum_{i=1}^N \bar{a}_i \langle y, x_i^* \rangle + \sum_{i,j=1}^N a_i \bar{a}_j \langle x_i^*, x_j^* \rangle \\ &= \langle y, y \rangle - \sum_{i=1}^N a_i \langle x_i^*, y \rangle - \sum_{i=1}^N \bar{a}_i \langle y, x_i^* \rangle + \sum_{i=1}^N |a_i|^2 \\ &\quad + \sum_{i=1}^N \langle x_i^*, y \rangle \langle y, x_i^* \rangle - \sum_{i=1}^N \langle x_i^*, y \rangle \langle y, x_i^* \rangle \\ &= \langle y, y \rangle - \sum_{i=1}^N |\langle y, x_i^* \rangle|^2 + \sum_{i=1}^N |a_i - \langle y, x_i^* \rangle|^2. \end{aligned}$$

Ya que los dos primeros términos no dependen de las "a's", el mínimo se alcanza cuando $a_i^* = \langle y, x_i^* \rangle$, $i = 1, 2, \dots, N$. □

Del Teorema 2.3.4 podemos deducir los siguientes corolarios con los que finalizamos este capítulo.

Corolario 2.3.1

$$\min_{a_i} \left\| y - \sum_{i=1}^N a_i x_i^* \right\|^2 = \|y\|^2 - \sum_{i=1}^N |\langle y, x_i^* \rangle|^2.$$

Ya que el valor mínimo del teorema anterior es mayor o igual que 0, tenemos el siguiente

Corolario 2.3.2 (Desigualdad de Bessel) Si $\{x_i^*\}_{i=1}^N$ es un sistema ortonormal, entonces

$$\sum_{i=1}^N |\langle y, x_i^* \rangle|^2 \leq \|y\|^2. \quad (2.11)$$

Teoría de la clausura y la completitud

3.1. Teoría de la clausura y la completitud

En este último capítulo abordaremos los conceptos de clausura y completitud en espacios con producto interior y en espacios vectoriales normados.

Definición 3.1.1 *Un sistema de elementos x_1, x_2, \dots , finito o no finito de un espacio normado $(X, \|\cdot\|)$, se dice que es cerrado si cada elemento $x \in X$ puede ser aproximado tanto como se desee por combinaciones lineales finitas de los x_i . Esto es, dado $x \in X$ y $\epsilon > 0$, podemos encontrar constantes a_1, \dots, a_n tales que*

$$\|x - (a_1x_1 + a_2x_2 + \dots + a_nx_n)\| \leq \epsilon.$$

Ejemplo 3.1.1 *Cualquier conjunto de n elementos linealmente independientes en \mathbb{R}^n o \mathbb{C}^n es cerrado. Además, la aproximación sería perfecta, pues con combinaciones lineales de los elementos de una base de \mathbb{R}^n o \mathbb{C}^n se obtendrían los elementos buscados.*

Ejemplo 3.1.2 *El Teorema de Aproximación de Weierstrass (Teorema 1.0.1) nos dice que $\{1, x, x^2, \dots\}$ es cerrado en el espacio normado $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$.*

Ejemplo 3.1.3 *Sea X el conjunto de funciones analíticas que son continuas en $|z| \leq 1$, con la norma $\|f\| = \max_{|z| \leq 1} |f(x)|$. Las potencias $\{z, z^2, z^3, \dots\}$ no son cerradas en X . Si lo fueran, dado $\epsilon > 0$, podríamos encontrar constantes a_1, \dots, a_n , tales que*

$$\max_{|z| \leq 1} |1 - (a_1z + a_2z^2 + \dots + a_nz^n)| \leq \epsilon.$$

Si tomamos $z = 0$, tendríamos que $1 \leq \epsilon$, lo que es una contradicción.

Definición 3.1.2 ■ *Sea $(X, \|\cdot\|)$ un espacio vectorial normado. Una sucesión de elementos $\{x_n\}$ se dice que converge a un elemento $x \in X$, si $\lim_{n \rightarrow \infty} \|x - x_n\| = 0$. También se puede denotar como $\lim_{n \rightarrow \infty} x_n = x$.*

- *Un espacio X se dice que es completo si toda sucesión de Cauchy converge en X .*

Una sucesión convergente no puede converger a dos elementos distintos. Supongamos que así sea y que $\lim_{n \rightarrow \infty} \|x_n - x\| = \lim_{n \rightarrow \infty} \|x_n - y\| = 0$. Entonces, por (2.3) y la Definición 2.2.1(4) tenemos que $0 \leq \|x - y\| \leq \|x - x_n\| + \|x_n - y\|$. Tomando $n \rightarrow \infty$, tenemos que $\lim_{n \rightarrow \infty} \|x - y\| = 0$, y por tanto $x = y$.

Ejemplo 3.1.4 *Sea $(\mathcal{C}([-1, 1]), \|\cdot\|_2)$ con la función peso $\omega \equiv 1$ y $f_n(x) = \left[\frac{n}{1+n^4x^2}\right]^{\frac{1}{2}}$. Entonces, $\|0 - f_n\|^2 = \int_{-1}^1 \frac{n}{1+n^4x^2} dx = \frac{2}{n} \arctan n^2 \rightarrow 0$. Sin embargo, $f_n(0) \rightarrow \infty$.*

En el espacio $X = \mathcal{C}([a, b])$ podemos establecer una relación entre convergencia en $\|\cdot\|_2$ y $\|\cdot\|_\infty$.

Lema 3.1.1 *Sea $X = \mathcal{C}([a, b])$ con las normas $\|\cdot\|_2$ y $\|\cdot\|_\infty$, y consideramos la función peso ω . Si una sucesión $\{f_n\}$ converge a f respecto de la norma $\|\cdot\|_\infty$, converge también a f con la norma $\|\cdot\|_2$. El recíproco no es cierto.*

DEMOSTRACIÓN. Obsérvese que

$$\|f\|_2 \leq \left[\int_a^b \|f\|_\infty^2 \omega(x) dx \right]^{\frac{1}{2}} = \|f\|_\infty \cdot \sqrt{c_0} < \infty, \quad c_0 = \int_a^b \omega(x) dx < \infty.$$

Supongamos que $f_n \rightarrow f$ en norma $\|\cdot\|_\infty$. Entonces, $\forall \epsilon > 0$ existe $N = N(\epsilon)$, tal que $\forall n \in \mathbb{N}$, $\|f_n - f\|_\infty < \frac{\epsilon}{\sqrt{c_0}}$. Para ϵ y $N \in \mathbb{N}$ tomados: $\|f_n - f\|_2 \leq \|f_n - f\|_\infty \cdot \sqrt{c_0} < \epsilon$. Tomando $n \rightarrow \infty$, obtenemos $f_n \rightarrow f$ en norma $\|\cdot\|_2$. Veamos que el recíproco es falso. Tomamos $f_n = x^n$, $x \in [0, 1]$ y $\omega \equiv 1$. Puntualmente,

$$f_n \xrightarrow{n \rightarrow \infty} f = \begin{cases} 0, & x \in [0, 1), \\ 1, & x = 1. \end{cases}$$

Por tanto, $\|f_n - f\|_2^2 = \int_0^1 (f_n(x) - f(x))^2 dx = \frac{1}{2n+1} \xrightarrow{n \rightarrow \infty} 0$. Sin embargo, $\lim_{n \rightarrow \infty} \|f_n - f\|_\infty \neq 0$.

□

Para el siguiente resultado, que es consecuencia de los resultados obtenidos al final del Capítulo 2, supondríamos que nuestro espacio vectorial normado es realmente pre-Hilbert.

Teorema 3.1.1 *Sea x_1^*, x_2^*, \dots una sucesión de elementos ortonormales de un espacio pre-Hilbert $(X, \langle \cdot, \cdot \rangle)$ que podría consistir en un número finito de elementos. Consideremos las siguientes afirmaciones:*

- (A) El conjunto $\{x_1^*, x_2^*, \dots\}$ es cerrado.
 (B) Las series de Fourier de cualquier elemento $y \in X$ convergen en norma a y . Es decir, $\lim_{n \rightarrow \infty} \left\| y - \sum_{k=1}^n \langle y, x_k^* \rangle x_k^* \right\| = 0$.
 (C) Se cumple la identidad de Parseval. Esto significa, para cualquier $y \in X$,

$$\|y\|^2 = \langle y, y \rangle = \sum_{n=1}^{\infty} |\langle y, x_n^* \rangle|^2. \quad (3.1)$$

Debemos notar que la Identidad de Parseval (3.1) es una generalización del Teorema de Pitágoras (Teorema 2.3.1).

- (C') Se cumple la identidad de Parseval extendida. Esto es, para cualesquiera $x, y \in X$, $\langle x, y \rangle = \sum_{n=1}^{\infty} \langle x, x_n^* \rangle \langle x_n^*, y \rangle$.
 (D) No puede existir un sistema ortonormal estrictamente más grande que contenga a x_1^*, x_2^*, \dots .
 (E) La sucesión $\{x_1^*, x_2^*, \dots\}$ es completa. Es decir, si $y \in X$ y $\langle y, x_k^* \rangle = 0$, $k = 1, 2, \dots$, entonces $y = 0$.
 (F) Cualquier elemento de X está determinado de manera única por sus coeficientes de Fourier. Esto es, si $\langle \omega, x_k^* \rangle = \langle y, x_k^* \rangle$, $k = 1, 2, \dots$, entonces $\omega = y$.

Entonces se cumple que

$$A \Leftrightarrow B \Leftrightarrow C \Leftrightarrow C' \Rightarrow D \Leftrightarrow E \Leftrightarrow F. \quad (3.2)$$

Si X es completo, entonces todas las afirmaciones son equivalentes:

$$A \Leftrightarrow B \Leftrightarrow C \Leftrightarrow C' \Leftrightarrow D \Leftrightarrow E \Leftrightarrow F. \quad (3.3)$$

DEMOSTRACIÓN. Supongamos que se da (A). Entonces, por (2.10),

$$\left\| y - \sum_{k=1}^n \langle y, x_k^* \rangle x_k^* \right\| \leq \left\| y - \sum_{k=1}^n a_k x_k^* \right\|.$$

Se sigue de (A) que la segunda expresión puede ser acotada por ϵ , por tanto se da (B). El recíproco también es cierto, pues podemos aproximar cada elemento de X por sus coeficientes de Fourier, y por tanto, el conjunto de los x_k^* es cerrado. Por ortogonalidad,

$$\left\langle x - \sum_{k=1}^n \langle x, x_k^* \rangle x_k^*, y - \sum_{k=1}^n \langle y, x_k^* \rangle x_k^* \right\rangle = \langle x, y \rangle - \sum_{k=1}^n \langle x, x_k^* \rangle \langle x_k^*, y \rangle, \quad x, y \in X.$$

Por la Desigualdad de Cauchy-Schwarz (Proposición 2.3.2),

$$\left| \langle x, y \rangle - \sum_{k=1}^n \langle x, x_k^* \rangle \langle x_k^*, y \rangle \right| \leq \left\| x - \sum_{k=1}^n \langle x, x_k^* \rangle x_k^* \right\| \cdot \left\| y - \sum_{k=1}^n \langle y, x_k^* \rangle x_k^* \right\|.$$

Si se cumple (B), ambos miembros a la derecha de la desigualdad anterior tienden a 0, por tanto se cumple (C'). Si además tomamos $x = y$ en (C'), este implica (C). Por el Corolario 2.3.1,

$$0 \leq \left\| y - \sum_{k=1}^n \langle y, x_k^* \rangle x_k^* \right\|^2 = \left\| y \right\|^2 - \sum_{k=1}^n |\langle y, x_k^* \rangle|^2.$$

Entonces, hemos probado $(C) \Rightarrow (B)$, y por tanto, $(A) \Leftrightarrow (B) \Leftrightarrow (C) \Leftrightarrow (C')$. Supongamos que se da (A) y que $x_1^*, x_2^*, \dots, \omega$, ($\omega \neq x_i^*$) también es un sistema ortonormal. Este sistema también es un cerrado en X . Como $A \Rightarrow C'$, tenemos las siguientes ecuaciones:

$$\| \omega \|^2 = \sum_{k=1}^{\infty} |\langle \omega, x_k^* \rangle|^2 + \langle \omega, \omega \rangle, \quad \| \omega \|^2 = \sum_{k=1}^{\infty} |\langle \omega, x_k^* \rangle|^2.$$

Por tanto $\langle \omega, \omega \rangle = 0$, y esto es una contradicción, pues $\| \omega \| = 1$. Por tanto se da (D).

Supongamos que existe $y \in X$, $y \neq 0$ tal que $\langle y, x_k^* \rangle = 0$, $k = 1, 2, \dots$. Entonces $x_1^*, x_2^*, \dots, \frac{y}{\|y\|}$ sería un sistema ortonormal estrictamente más grande que x_1^*, x_2^*, \dots . Por tanto, $(D) \Leftrightarrow (E)$.

Supongamos que $\langle \omega, x_k^* \rangle = \langle y, x_k^* \rangle$, $k = 1, 2, \dots$. Entonces $\langle \omega - y, x_k^* \rangle = 0$, $k = 1, 2, \dots$. Si se cumple (E), entonces $\omega - y = 0$, por tanto $(E) \Rightarrow (F)$. Si (E) fuera falso, podríamos encontrar $z \neq 0$ con $\langle z, x_k^* \rangle = 0$, $k = 1, 2, \dots$. Para cualquier y , $\langle y, x_k^* \rangle = \langle y + z, x_k^* \rangle$, $k = 1, 2, \dots$. Entonces y , $y + z$ serían dos elementos distintos con los mismos coeficientes de Fourier, por lo que (F) sería falso. Por tanto $(F) \Rightarrow (E)$. Esto completa la cadena de implicaciones (3.2).

Ahora asumamos que X es completo. Probaremos la cadena de implicaciones (3.3) demostrando que $(F) \Rightarrow (B)$. Sea $\omega \in X$. Denotemos $s_n = \sum_{k=1}^n \langle \omega, x_k^* \rangle x_k^*$.

Para $n > m$, tenemos que $s_n - s_m = \sum_{k=m+1}^n \langle \omega, x_k^* \rangle x_k^*$, por lo que $\| s_n - s_m \|^2 =$

$\sum_{k=m+1}^n |\langle \omega, x_k^* \rangle|^2$. Por (2.11), $\sum_{k=m+1}^n |\langle \omega, x_k^* \rangle|^2 < \infty$. Como la suma anterior con-

verge, dado $\epsilon > 0$ podemos encontrar $N(\epsilon)$ tal que $\sum_{k=m+1}^n |\langle \omega, x_k^* \rangle|^2 < \epsilon$, para todo

$m, n \geq N(\epsilon)$. Por tanto s_n es una sucesión de Cauchy, y como asumimos que X es completo, s_n converge a algún elemento $s \in X$:

$$\lim_{n \rightarrow \infty} \| s_n - s \| = 0. \quad (3.4)$$

Sea v un entero positivo fijado y $n \geq v$. Entonces $\langle s - s_n, x_v^* \rangle = \langle s, x_v^* \rangle - \langle s_n, x_v^* \rangle = \langle s, x_v^* \rangle - \langle \omega, x_v^* \rangle$. De nuevo, por la Desigualdad de Cauchy-Schwarz (Proposición 2.3.2),

$$|\langle s, x_v^* \rangle - \langle \omega, x_v^* \rangle| = |\langle s - s_n, x_v^* \rangle| \leq \|s - s_n\| \cdot \|x_v^*\| = \|s - s_n\|. \quad (3.5)$$

En vista de (3.4) y (3.5) tenemos que $\langle s, x_v^* \rangle = \langle \omega, x_v^* \rangle$, $v = 1, 2, \dots$. Por (F), esto implica que $s = \omega$, por lo que (3.4) queda como $\lim_{n \rightarrow \infty} \left\| \omega - \sum_{k=1}^n \langle \omega, x_k^* \rangle x_k^* \right\| = 0$. Pero esto es exactamente (B). □

La propiedad de completitud dada en el Teorema 3.1.1(E) se puede definir para un conjunto de elementos (recuérdase la Definición 3.1.2).

Definición 3.1.3 Sea X un espacio con producto interior. Un conjunto S de elementos de X se dice que es completo si $\langle y, x \rangle = 0$, $\forall x \in S \Rightarrow y = 0$.

Recordemos de la Definición 3.1.1 que si X es un espacio con producto interior, entonces una sucesión $\{x_k\} \in X$ es cerrada si cada elemento $x \in X$ puede aproximarse tanto como se desee por combinaciones lineales finitas de esta.

Ejemplo 3.1.5 Sea $X = \mathcal{C}([a, b])$, ω una función peso definida en $[a, b]$, y el producto interior inducido $\langle f, g \rangle = \int_a^b f(x)g(x)\omega(x)dx$. Dada $f \in \mathcal{C}([a, b])$ y $\epsilon > 0$, podemos encontrar constantes a_k , tales que $|f(x) - (a_0 + \dots + a_n x^n)| \leq \epsilon$, $a \leq x \leq b$. Por integración,

$$\int_a^b (f(x) - (a_0 + \dots + a_n x^n))^2 \omega(x) dx \leq \epsilon^2 c_0, \quad c_0 = \int_a^b \omega(x) dx < \infty.$$

Se sigue de esto que las potencias $\{1, x, x^2, \dots\}$ son cerradas en $(X, \langle \cdot, \cdot \rangle)$.

Teorema 3.1.2 (Riesz) Sean X un espacio completo con producto interior, a_k constantes tales que $\sum_{k=1}^{\infty} |a_k|^2 < \infty$ y $\{x_k^*\}$ una sucesión ortonormal completa. Entonces, existe $y \in X$ tal que

$$\langle y, x_k^* \rangle = a_k, \quad k = 1, 2, \dots \quad (3.6)$$

DEMOSTRACIÓN. Consideremos los elementos $s_n = \sum_{k=1}^n a_k x_k^*$. Tenemos que

$\|s_n - s_m\|^2 = \sum_{k=m+1}^n |a_k|^2$. Ya que $\sum_{k=1}^{\infty} |a_k|^2 < \infty$, $\{s_n\}$ es una sucesión de Cauchy, y existe y tal que $\lim_{n \rightarrow \infty} \|y - s_n\| = 0$. Si fijamos k y $n \geq k$,

$$\|y - s_n\| = \|y - s_n\| \cdot \|x_k^*\| \geq |\langle y - s_n, x_k^* \rangle| = |\langle y, x_k^* \rangle - a_k|.$$

Tomando $n \rightarrow \infty$, obtenemos (3.6).

□

Definición 3.1.4 Sea X un espacio vectorial dado. Se dice que L es un funcional lineal sobre X , si es una aplicación lineal de X a \mathbb{R} o \mathbb{C} . Es decir, que cumple las siguientes propiedades:

- $L(x + y) = L(x) + L(y), \quad \forall x, y \in X.$
- $L(\alpha x) = \alpha L(x), \quad \forall x \in X \text{ y } \alpha \in \mathbb{R} \text{ o } \alpha \in \mathbb{C}.$

Definición 3.1.5 Sea X un espacio dado y L un funcional definido sobre X . Se dice que L es un funcional acotado, si existe una constante $M > 0$ tal que

$$|L(x)| \leq M \cdot \|x\|, \quad \forall x \in X. \quad (3.7)$$

Definición 3.1.6 Sea X un espacio normado y L un funcional lineal acotado, definido sobre X . Entonces, $\|L\|$ es el valor mínimo para el que se cumple (3.7). Obviamente, tenemos que $|L(x)| \leq \|L\| \cdot \|x\|, \forall x \in X$, y para cada $\epsilon > 0$, podemos encontrar $x_0 \in X$ tal que $|L(x_0)| > (\|L\| - \epsilon) \cdot \|x_0\|$. De la desigualdad anterior se puede extraer una fórmula alternativa (y más sencilla) para $\|L\|$:

$$\|L\| = \sup_{x \in X} \frac{|L(x)|}{\|x\|}, \quad \forall x \in X.$$

Definición 3.1.7 El espacio normado de los funcionales lineales acotados definidos en un espacio normado X es conocido como el espacio normado conjugado de X y se denota como X^* .

Definición 3.1.8 Se dice que una sucesión $\{x_k\}$ de elementos de X es completa, si $L(x_k) = 0, \forall k = 1, 2, \dots, L \in X^*$ implica $L \equiv 0$.

Teorema 3.1.3 Sea X un espacio normado e Y un subespacio vectorial de X . Sea $x_0 \in X \setminus Y$ y supongamos que $d = \inf_{y \in Y} \|y - x_0\| > 0$. Entonces podemos encontrar un funcional acotado $L \in X^*$, tal que

$$L(x) = 0, \quad L(x_0) = 1, \quad \|L\| = \frac{1}{d}, \quad x \in Y.$$

Teorema 3.1.4 (Lauricella). Sea X un espacio normado y $\{x_n\}$ un sistema cerrado en X . Entonces, otro sistema $\{y_n\}$ es cerrado en X , si y solo si, es cerrado en $\{x_n\}$, es decir, si cada elemento de $\{x_n\}$ puede ser aproximado arbitrariamente cerca por combinaciones lineales de los y_n .

Omitiremos las demostraciones de estos dos últimos teoremas por exceder al contenido necesario para la Memoria. Ambas pueden encontrarse en [4, Teorema 11.1.6] y [4, Teorema 11.1.8], respectivamente.

Teorema 3.1.5 (Banach) *Sea X un espacio normado (real o complejo). Una sucesión de elementos $\{x_k\}$ es cerrada, si y solo si, es completa.*

DEMOSTRACIÓN. Supongamos primero que $\{x_k\}$ es cerrada. Sea $L \in X^*$, con $L(x_k) = 0$, $k = 1, 2, \dots$. Dado $x \in X$, podemos aproximarlos arbitrariamente cerca por combinaciones finitas de los x_k : $\|x - (a_1x_1 + a_2x_2 + \dots + a_nx_n)\| \leq \epsilon$, para algunos coeficientes a_k . Entonces,

$$\begin{aligned} |L(x)| &= |L[x - (a_1x_1 + a_2x_2 + \dots + a_nx_n)]| \\ &\leq \|L\| \cdot \|x - (a_1x_1 + a_2x_2 + \dots + a_nx_n)\| \leq \|L\| \cdot \epsilon. \end{aligned}$$

Tomemos $\epsilon \rightarrow 0$ y obtendremos $L(x) = 0$. Ya que x es arbitrario, $L \equiv 0$. Recíprocamente, supongamos que $L(x_k) = 0$, $k = 1, 2, \dots$, implica que $L \equiv 0$. Sea $x_0 \in X$ e Y el subespacio generado por las combinaciones lineales finitas de los x_k . Necesitamos probar que $d = \inf_{y \in Y} \|x_0 - y\| = 0$. Supongamos lo contrario.

Entonces, por el Teorema 3.1.3 podemos encontrar $L \in X^*$, tal que $L(y) = 0$, $y \in Y$, y $L(x_0) = 1$. Si $L(x_k) = 0$, $k = 1, 2, \dots$, por completitud tenemos que $L \equiv 0$. Pero esto contradice que $L(x_0) = 1$.

□

Ejemplo 3.1.6 *Sea X un espacio con producto interior, y completo. Cualquier $L \in X^*$ puede ser expresado como $L(x) = \langle x, x_0 \rangle$, para algún $x_0 \in X$ (véase [1, Sección 16]). Por tanto, la definición de completitud de $\{x_k\}$ es que $\langle x_k, x \rangle = 0$, $k = 1, 2, \dots$, implica $x = 0$. En este caso, el teorema anterior nos da la equivalencia entre la propiedad (A) y (E) del Teorema 3.1.1.*

Ejemplo 3.1.7 *Sea $X = C([a, b])$ con la norma $\|f\| = \max_{a \leq x \leq b} |f(x)|$. Por el Teorema de Weierstrass (Teorema 1.0.1), la sucesión de potencias $\{1, x, x^2, \dots\}$ es cerrada. Para cualquier función $g(x) \in C([a, b])$, el funcional lineal $L(f) = \int_a^b f(x)g(x)\omega(x)dx$ pertenece a X^* , siendo ω una función peso en $[a, b]$. Por tanto, por el Teorema 3.1.5, esta sucesión es completa y $\int_a^b x^n g(x)\omega(x)dx = 0$, $n = 0, 1, 2, \dots$ implica que $g \equiv 0$.*

3.2. Completitud de las potencias y sistemas trigonométricos en $L^2[a, b]$

Teorema 3.2.1 *Las potencias $\{1, x, x^2, \dots\}$ son completas en $L^2[a, b]$.*

DEMOSTRACIÓN. Sea $f \in L^2[a, b]$ y supongamos que

$$\int_a^b x^n f(x) dx = 0, \quad n = 0, 1, 2, \dots$$

Denotemos $F(x) = \int_a^x f(t) dt$. Entonces $F(x) \in \mathcal{C}([a, b])$. En particular, $F(a) = 0$ y $F(b) = 0$. Integrando por partes:

$$0 = \int_a^b x^n f(x) dx = x^n F(x) \Big|_a^b - n \int_a^b x^{n-1} F(x) dx.$$

Se sigue que $\int_a^b x^{n-1} F(x) dx = 0$, $n = 1, 2, \dots$. Por el Ejemplo 3.1.7, las potencias son completas en $\mathcal{C}([a, b])$ con la norma $\|f\| = \max_{a \leq x \leq b} |f(x)|$, y esto implica que $F \equiv 0$. Por tanto $f(x) = 0$ en casi todo punto.

□

Corolario 3.2.1 *La sucesión de polinomios ortonormales en $[a, b]$, $p_n^*(x) = k_n x^n + \dots$, $k_n > 0$, cumple todas las propiedades del Teorema 3.1.1, para $f \in L^2[a, b]$.*

Definición 3.2.1 *Se dice que T es un polinomio trigonométrico, si puede ser expresado como combinaciones lineales finitas de $\sin(nx)$ y $\cos(nx)$, con $n \in \mathbb{N}$, $x \in \mathbb{R}$. Si $T(x) = a_0 + \sum_{k=0}^n a_k \cos(kx) + b_k \sin(kx)$, y $|a_n| + |b_n| > 0$, entonces se dirá que T es un polinomio trigonométrico de grado exacto n .*

Lema 3.2.1 *Sea $f \in \mathcal{C}([-\pi, \pi])$ y supongamos que*

$$\int_{-\pi}^{\pi} f(x) \cos(nx) dx = 0, \quad n = 0, 1, \dots, \quad (3.8)$$

y

$$\int_{-\pi}^{\pi} f(x) \sin(nx) dx = 0, \quad n = 1, 2, \dots. \quad (3.9)$$

Entonces, $f \equiv 0$.

DEMOSTRACIÓN. Sea T_n un polinomio trigonométrico arbitrario. Se sigue de (3.8) y (3.9) que $\int_{-\pi}^{\pi} T_n(x) f(x) dx = 0$. Supongamos que $f \not\equiv 0$. Entonces existe un punto interior x_0 de $[-\pi, \pi]$ para el cual $f(x_0) \neq 0$. Supongamos, sin pérdida de generalidad, que $f(x_0) = m > 0$. Por continuidad, podemos encontrar un intervalo $I : x_0 - \delta < x < x_0 + \delta$ contenido en $(-\pi, \pi)$, tal que $f(x) \geq \frac{m}{2}$. Consideremos la función trigonométrica $t(x) = 1 - \cos \delta + \cos(x - x_0)$. Para $x_0 - \delta < x < x_0 + \delta$, $\cos(x - x_0) > \cos \delta$, entonces $t(x) > 1$. Para los casos

$x = x_0 \pm \delta$, $t(x) = 1$. En el resto de $[-\pi, \pi]$, $-1 \leq \cos(x - x_0) < \cos \delta$, por lo que

$$-\cos \delta \leq t(x) < 1, \text{ y por tanto, } |t(x)| < 1.$$

Ahora consideremos $T_n(x) = [t(x)]^n$, polinomio trigonométrico de grado exacto n . Es claro que

$$T_n(x) > 1 \text{ en } I_1 : x_0 - \delta < x < x_0 + \delta,$$

$$T_n(x) = 1, \text{ para } x = x_0 \pm \delta,$$

$$|T_n(x)| < 1 \text{ en } I_2 : [-\pi, \pi] \setminus [x_0 - \delta, x_0 + \delta],$$

pero

$$0 = \int_{-\pi}^{\pi} f(x)T_n(x)dx = \int_{I_1} f(x)T_n(x)dx + \int_{I_2} f(x)T_n(x)dx.$$

Por tanto $\int_{I_1} = -\int_{I_2}$. Ahora, $\left| \int_{I_2} f(x)T_n(x)dx \right| \leq \int_{I_2} |f(x)| dx$, por tanto está acotada para $n \rightarrow \infty$. Ya que $t(x) > 1$ en I_1 , $t(x) \geq 1 + \epsilon$ en

$$I_3 : x_0 - \frac{\delta}{2} \leq x \leq x_0 + \frac{\delta}{2}.$$

Por tanto, $T_n(x) = [t(x)]^n \geq (1 + \epsilon)^n$ en I_3 , y

$$\int_{I_1} f(x)T_n(x)dx \geq \frac{m}{2} \int_{I_3} T_n(x)dx \geq \frac{m}{2} (1 + \epsilon)^n.$$

Esto es una contradicción, ya que $\int_{I_1} \rightarrow \infty$, mientras que \int_{I_2} está acotada. Por tanto, la asunción $f \not\equiv 0$ es falsa.

□

Teorema 3.2.2 *El sistema de funciones $\cos(nx)$, $n = 0, 1, \dots$, $\sin(nx)$, $n = 1, 2, \dots$ es completo en $L^2[-\pi, \pi]$.*

DEMOSTRACIÓN. Sea $f \in L^2[-\pi, \pi]$. Debemos probar que las condiciones (3.8) y (3.9) implican $f(x) = 0$, en casi todo punto. Esto implica completitud por el Ejemplo 3.1.6. Sabemos que la función $F(x) = \int_{-\pi}^x f(t)dt$ pertenece a $\mathcal{C}([-\pi, \pi])$. Además, $F(-\pi) = 0$ y $F(\pi) = 0$. El último sigue de (3.8) con $n = 0$. Si T_n constituye un polinomio trigonométrico arbitrario, entonces $\int_{-\pi}^{\pi} f(x)T_n(x)dx = 0$.

Pero

$$\int_{-\pi}^{\pi} f(x)T_n(x)dx = F(x)T(x) \Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f(x)T'_n(x)dx.$$

Por tanto, $\int_{-\pi}^{\pi} f(x)T'_n(x)dx = 0$ para todas las derivadas T' de polinomios trigonométricos. En particular, $\int_{-\pi}^{\pi} F(x)\sin(nx)dx = \int_{-\pi}^{\pi} F(x)\cos(nx)dx = 0$, $n = 1, 2, \dots$. Consideremos ahora

$$G(x) = F(x) - c, \quad c = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(x)dx.$$

Es obvio que

$$\int_{-\pi}^{\pi} G(x)\sin(nx)dx = 0, \quad n = 1, 2, \dots,$$

$$\int_{-\pi}^{\pi} G(x)\cos(mx)dx = 0, \quad m = 0, 1, \dots$$

Por el Lema 3.2.1, $G \equiv 0$, por tanto $F(x) = c$. Pero $F(\pi) = 0$. Entonces, $F \equiv 0$, y $f(x) = 0$, en casi todo punto. □

3.3. Teorema de la Clausura de Müntz

Supongamos que tenemos una sucesión de potencias $\{x^{p_k}\}$. ¿Bajo qué circunstancias pueden ser aproximadas funciones continuas, o de $L^2[a, b]$, por combinaciones lineales de estas potencias? Müntz dio un método brillante para solucionar esta cuestión, que es una aplicación del Teorema 2.3.3. A continuación, veremos dos lemas necesarios para terminar esta memoria con el Teorema de la Clausura de Müntz.

Lema 3.1. (Cauchy). Si

$$D_n = \begin{vmatrix} \frac{1}{a_1+b_1} & \frac{1}{a_1+b_2} & \cdots & \frac{1}{a_1+b_n} \\ \vdots & & & \vdots \\ \frac{1}{a_n+b_1} & \frac{1}{a_n+b_2} & \cdots & \frac{1}{a_n+b_n} \end{vmatrix}, \quad \text{entonces } D_n = \frac{\prod_{i>j}^n (a_i - a_j)(b_i - b_j)}{\prod_{i,j=1}^n (a_i + b_j)}.$$

DEMOSTRACIÓN. Consideremos las a_i y los b_j como $2n$ elementos independientes y calculemos su denominador común, una vez hayamos expandido el determi-

nante D_n . El común denominador buscado sería $\prod_{i,j=1}^n (a_i + b_j)$. Cada término

individual tiene grado -1 , por tanto D_n será de grado $-n$. El denominador común tiene n^2 factores, por tanto tiene grado n^2 . Se sigue entonces que el numerador debe ser un polinomio de grado $n^2 - n$ en las variables a_i y b_j . Si $a_i = a_j$ para algunos $i \neq j$, entonces coincidirían la fila i y j , por tanto $D_n = 0$. Lo mismo ocurriría con las columnas i y j si $b_i = b_j$. Supongamos entonces

lo contrario. Se sigue que el numerador debería contener un factor de la forma $\prod_{i>j}^n (a_i - a_j) \cdot \prod_{i>j}^n (b_i - b_j)$. Cada producto contiene $\frac{n(n-1)}{2}$ factores, por tanto el producto de ambos tiene $n(n-1)$. El grado es por tanto correcto y obtenemos

$$D_n = c_n \frac{\prod_{i>j}^n (a_i - a_j) (b_i - b_j)}{\prod_{i,j=1}^n (a_i + b_j)}, \quad (3.10)$$

siendo c_n una constante independiente de las a_i y los b_j . Probemos que $c_n = 1$. Nótese que

$$a_n D_n = \begin{vmatrix} \frac{1}{a_1+b_1} & \cdots & \frac{1}{a_1+b_n} \\ \vdots & & \vdots \\ \frac{a_n}{a_n+b_1} & \cdots & \frac{a_n}{a_n+b_n} \end{vmatrix}.$$

Entonces,

$$\lim_{a_n \rightarrow \infty} a_n D_n = \begin{vmatrix} \frac{1}{a_1+b_1} & \cdots & \frac{1}{a_1+b_n} \\ \vdots & & \vdots \\ \frac{1}{a_n+b_1} & \cdots & \frac{1}{a_n+b_n} \\ 1 & \cdots & 1 \end{vmatrix}.$$

También,

$$\lim_{b_n \rightarrow \infty} \lim_{a_n \rightarrow \infty} a_n D_n = \begin{vmatrix} \frac{1}{a_1+b_1} & \cdots & \frac{1}{a_1+b_{n-1}} & 0 \\ \vdots & & \vdots & \vdots \\ \frac{1}{a_{n-1}+b_1} & \cdots & \frac{1}{a_{n-1}+b_{n-1}} & 0 \\ 1 & \cdots & 1 & 1 \end{vmatrix} = D_{n-1}.$$

Entonces, $\lim_{b_n \rightarrow \infty} \lim_{a_n \rightarrow \infty} \frac{a_n D_n}{D_{n-1}} = 1$. Pero de (3.10), tenemos que

$$\frac{a_n D_n}{D_{n-1}} = \frac{c_n}{c_{n-1}} \frac{a_n \prod_{j=1}^{n-1} (a_n - a_j) \cdot \prod_{j=1}^{n-1} (b_n - b_j)}{\prod_{j=1}^n (a_n + b_j) \cdot \prod_{j=1}^{n-1} (b_n + a_j)}.$$

Tomando límites, $\lim_{b_n \rightarrow \infty} \lim_{a_n \rightarrow \infty} \frac{a_n D_n}{D_{n-1}} = 1 = \frac{c_n}{c_{n-1}}$. Por tanto, $c_n = c_{n-1}$. Se comprueba fácilmente que $c_1 = 1$. Por tanto, $c_n = 1$. □

Lema 3.2. *Sea $p_i \neq p_j$. Entonces, asumiendo que $p_i, q > -\frac{1}{2}$, tenemos que*

$$E_n^2(x^q) = \min_{a_k} \int_0^1 |x^q - a_1 x^{p_1} - a_2 x^{p_2} - \cdots - a_n x^{p_n}|^2 dx = \frac{1}{2q+1} \prod_{i=1}^n \left\{ \frac{p_i - q}{p_i + q + 1} \right\}^2.$$

DEMOSTRACIÓN. Del Teorema 2.3.3, tenemos que $E_n^2(x^q) = \frac{g(x^q, x^{p_1}, x^{p_2}, \dots, x^{p_n})}{g(x^{p_1}, x^{p_2}, \dots, x^{p_n})}$.

Ahora, $\langle x^\alpha, x^\beta \rangle = \int_0^1 x^\alpha x^\beta = \frac{1}{\alpha + \beta + 1}$. Entonces,

$$g(x^{p_1}, \dots, x^{p_n}) = \left| \begin{array}{cccc} \frac{1}{p_1+p_1+1} & \cdots & \frac{1}{p_1+p_n+1} \\ \vdots & & \vdots \\ \frac{1}{p_n+p_1+1} & \cdots & \frac{1}{p_n+p_n+1} \end{array} \right| = \frac{\prod_{i>j}^n (p_i - p_j)^2}{\prod_{i,j=1}^n (p_i + p_j + 1)}.$$

Para $g(x^q, x^{p_1}, x^{p_2}, \dots, x^{p_n})$ tenemos una expresión similar. Entonces,

$$\begin{aligned} \frac{g(x^q, x^{p_1}, \dots, x^{p_n})}{g(x^{p_1}, \dots, x^{p_n})} &= \frac{(q-p_1)^2 (q-p_2)^2 \cdots (q-p_n)^2}{(q+p_1+1)(q+p_2+1) \cdots (q+p_n+1)(q+q+1)(p_1+1+q) \cdots (p_n+1+q)} \\ &= \frac{1}{2q+1} \prod_{i=1}^n \frac{(q-p_i)^2}{(q+p_i+1)^2}. \end{aligned}$$

□

Teorema 3.3.1 (Müntz). *Sea $\{x^p\}$ una sucesión no finita compuesta de diferentes potencias, con $p > -\frac{1}{2}$. Entonces, $\{x^p\}$ es cerrada en $L^2(0, 1)$, si y solo si, los exponentes $\{p\}$ contienen una sucesión $\{p_i\}$ tal que,*

$$\lim_{i \rightarrow \infty} p_i = -\frac{1}{2}, \quad \sum_{i=1}^{\infty} \left(p_i + \frac{1}{2} \right) = \infty, \quad (3.11)$$

o

$$\lim_{i \rightarrow \infty} p_i = p, \quad -\frac{1}{2} < p < \infty, \quad (3.12)$$

o

$$\lim_{i \rightarrow \infty} p_i = \infty, \quad p_i \neq 0, \quad \sum_{i=1}^{\infty} \frac{1}{p_i} = \infty. \quad (3.13)$$

DEMOSTRACIÓN. Nótese que la condición $p > -\frac{1}{2}$ asegura que $x^p \in L^2(0, 1)$. Por el Teorema 3.2.1, sabemos que las potencias $\{1, x, x^2, \dots\}$ son cerradas en $L^2(0, 1)$. Por tanto, $\{x^p\}$ será cerrada en $L^2(0, 1)$ si y solo si, cada potencia x^q , $q = 0, 1, \dots$ es aproximable por combinaciones lineales de los x^p . Esto último sigue del Teorema 3.1.4. Entonces, para cada q , tenemos que encontrar una sucesión p_1, p_2, \dots , tal que $\lim_{n \rightarrow \infty} \min_{a_i} \int_0^1 |x^q - a_1 x^{p_1} - a_2 x^{p_2} - \dots - a_n x^{p_n}|^2 dx = 0$. De acuerdo al Lema 3.2, para cada q tenemos que ser capaces de encontrar una sucesión $\{p_k\}$ tal que

$$\lim_{n \rightarrow \infty} \prod_{i=1}^n \left(\frac{p_i - q}{p_i + q + 1} \right)^2 = 0. \quad (3.14)$$

Empecemos por ver la implicación de derecha a izquierda. Supongamos que p tiene un límite finito $p \neq -\frac{1}{2}$. Entonces, $\lim_{i \rightarrow \infty} \frac{p_i - q}{p_i + q + 1} = \frac{p - q}{p + q + 1}$. Ya que

$p > -\frac{1}{2}$ y $q \geq 0$, es fácil ver que $-1 < \frac{p-q}{p+q+1} < 1$. Dado $\epsilon > 0$, tenemos que para $i \geq n_\epsilon$, $\left| \frac{p_i - q}{p_i + q + 1} \right| < 1 - \epsilon$. Por tanto, se cumple (3.14). Supongamos que podemos obtener una sucesión $\{p_i\}$, tal que se cumpla (3.11). Escribiremos

$$\prod_{i=1}^n \left(\frac{p_i - q}{p_i + q + 1} \right)^2 = \frac{\prod_{i=1}^n \left(1 - \frac{p_i + \frac{1}{2}}{q + \frac{1}{2}} \right)^2}{\prod_{i=1}^n \left(1 + \frac{p_i + \frac{1}{2}}{q + \frac{1}{2}} \right)^2}. \quad (3.15)$$

Ya que $\lim_{i \rightarrow \infty} p_i = -\frac{1}{2}$, tenemos que $0 < \frac{p_i + \frac{1}{2}}{q + \frac{1}{2}} < 1$. Por tanto, $0 < 1 - \frac{p_i + \frac{1}{2}}{q + \frac{1}{2}} < 1$, y el numerador de la parte derecha en (3.15) se mantiene acotado. Además, ya que $\sum_{i=1}^{\infty} \frac{p_i + \frac{1}{2}}{q + \frac{1}{2}} = \infty$, el denominador diverge a $+\infty$. Por tanto, se vuelve a cumplir la condición (3.14). Finalmente, supongamos que se cumple (3.13). Escribimos

$$\prod_{i=1}^n \left(\frac{p_i - q}{p_i + q + 1} \right)^2 = \frac{\prod_{i=1}^n \left(1 - \frac{q}{p_i} \right)^2}{\prod_{i=1}^n \left(1 + \frac{q+1}{p_i} \right)^2}. \quad (3.16)$$

Debido a la hipótesis, el denominador diverge a $+\infty$. Por tanto, se cumple (3.14). Ahora veamos la implicación de izquierda a derecha (la condición de necesidad). Supongamos que el conjunto de exponentes $\{p\}$ no contiene ninguna sucesión $\{p_i\}$, tal que se cumpla alguna de las ecuaciones (3.11)-(3.13). Entonces, $\{p\}$ tiene que ser

1. Una sucesión $\{p_i\}$ con $\lim_{i \rightarrow \infty} p_i = -\frac{1}{2}$, $\sum_{i=1}^{\infty} \left(p_i + \frac{1}{2} \right) < \infty$, o
2. Una sucesión $\{p_i\}$ con $\lim_{i \rightarrow \infty} p_i = \infty$, $\sum_{i=1}^{\infty} \frac{1}{p_i} < \infty$, o
3. Una sucesión $\{p_i\}$ que puede ser dividida en dos subsucesiones $\{r_i\}$, $\{s_i\}$, una del tipo 1, y otra del tipo 2.

En el caso 1, nos centramos en (3.15). Elegimos, sin pérdida de generalidad, $q \neq p_1, p_2, \dots$. El numerador converge a 0, si y solo si, se anula algún factor. Esto es imposible, por tanto el numerador tiene un límite positivo. Esto significa que (3.14) no se cumple y x^q no puede ser aproximado arbitrariamente cerca por x^{p_1}, x^{p_2}, \dots . En el caso 2, nos situamos en (3.16). Si $q \neq p_1, p_2, \dots$, entonces el numerador y el denominador del lado derecho de la ecuación convergen a un número no nulo. Por tanto, otra vez, la aproximación arbitrariamente cerca de $\{x^q\}$ es imposible. En el caso 3, tomamos $q \neq r_1, r_2, \dots; s_1, s_2, \dots$. Tomamos $\prod_{i=1}^{\infty} \left(\frac{r_i - q}{r_i + q + 1} \right)^2 = a$, $\prod_{i=1}^{\infty} \left(\frac{s_i - q}{s_i + q + 1} \right)^2 = b$, donde sabemos que $0 < a < \infty$

y $0 < b < \infty$. Entonces, $\prod_{i=1}^{\infty} \left(\frac{p_i - q}{p_i + q + 1} \right)^2 = ab$. Otra vez, no se cumple (3.14).

Esto concluye la demostración.

□

Finalizamos la Memoria enunciando el segundo teorema de la clausura de Müntz. Omitiremos también esta demostración por excederse del contenido necesario para la misma. La demostración se puede encontrar en [4, Teorema 11.3.4].

Teorema 3.3.2 (Müntz). *Sea $\{p\}$ una sucesión de números distintos y no negativos. Para que $\{x^p\}$ sea una sucesión cerrada en $\mathcal{C}([0, 1])$, es suficiente que una de las p sea igual a 0, y $\{p\}$ contenga una subsucesión $\{p_i\}$, tal que*

$$\lim_{i \rightarrow \infty} p_i = \infty, \quad y \quad \sum_{i=1, p_i \neq 0}^{\infty} \frac{1}{p_i} = \infty, \quad o$$

$$\lim_{i \rightarrow \infty} p_i = p, \quad 0 < p < \infty.$$

A

Apéndice

A.1. Programación en Matlab: computación de polinomios de Bernstein

```
function ber(x,f,a,b,n)

%Transformación lineal del intervalo [a,b] al [0,1]:
g=subs(f,x,a+(b-a)*x);

%Vector simbólico donde almacenar los diversos
polinomios de Bernstein:
Bern=[];

syms k
%aux=floor(sqrt(n)); Alternativa
aux=floor(n/4);
for i=n:-aux:1
    B=simplify(symsum(nchoosek(i,k)*x^k *(1-
x)^(i-k)*subs(g,x,k/i),k,0,i));
    t=subs(B,x,(x-a)/(b-a));
    Bern=[Bern t];
    clear B t
end

u=[a:0.01:b]; v=double(subs(f,u));
w=double(subs(Bern(1),u));

subplot(1,2,1), plot(u,v,u,w), axis([a b min(v)
max(v)]), grid on, title ('Polinomio de Bernstein de
grado n.');
```

```
subplot(1,2,2), hold on
for i=2:length(Bern)
    j=double(subs(Bern(i),u));
    plot(u,v,u,j), axis([a b min(v) max(v)]), grid
on, title ('Diversos polinomios de Bernstein.');
```

```
clear j
end
hold off
end
```

Figura A.1.

A.2. Desigualdades de Young, Hölder y Minkowski

Para probar la Desigualdad de Minkowski (que permite probar que $(L^p(\mu), \|\cdot\|_p)$ es un espacio vectorial normado estrictamente convexo), necesitamos dos notables desigualdades previas:

Proposición A.2.1 (*Desigualdad de Young*). Sean $p, q > 1$ tales que $\frac{1}{p} + \frac{1}{q} = 1$, y $a, b \geq 0$. Entonces, $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$.

DEMOSTRACIÓN. Si $a = 0$ o $b = 0$, el lado izquierdo de la desigualdad es 0 y el derecho no negativo, entonces se cumple trivialmente. Por otro lado, por la convexidad de la función e^x , se tiene que

$$e^{\alpha x + \beta y} \leq \alpha e^x + \beta e^y, \quad (\text{A.1})$$

para $x, y \in \mathbb{R}$ y $\alpha, \beta \geq 0$, con $\alpha + \beta = 1$. Entonces, sustituyendo

$$\alpha = \frac{1}{p}, \quad \beta = \frac{1}{q}, \quad x = p \ln(a), \quad y = q \ln(b)$$

en (A.1), se obtiene el resultado. □

Para probar la próxima desigualdad necesitamos el siguiente

Lema A.2.1 Sean $x_1, x_2 \geq 0$ y $a, b > 0$, tales que $a + b = 1$. Entonces

$$x_1^a x_2^b \leq ax_1 + bx_2. \quad (\text{A.2})$$

La igualdad se cumple si y solo si $x_1 = x_2$.

DEMOSTRACIÓN. Sean $t > 1$, $m < 1$ y $f(t) = t^m$. Por el Teorema del Valor Medio, tenemos que $f(t) = f(1) + (t-1)f'(\xi)$, $1 < \xi < t$. Por tanto $t^m = 1 + (t-1)m\xi^{m-1}$. Como $\xi^{m-1} < 1$,

$$t^m < 1 + m(t-1). \quad (\text{A.3})$$

Supongamos que $x_1 > x_2 > 0$ (En el caso de que $x_1 = 0$ o $x_2 = 0$ la desigualdad se cumple trivialmente.). Elegimos $t = \frac{x_1}{x_2} > 1$, $m = a$, tal que $1-m = b$. De (A.3) obtenemos $\left(\frac{x_1}{x_2}\right)^a < 1 + a\left(\frac{x_1}{x_2} - 1\right)$, que equivale a $x_1^a x_2^{1-a} < x_2 + a(x_1 - x_2) = ax_1 + (1-a)x_2$. Por tanto, $x_1^a x_2^b < ax_1 + bx_2$. Si $x_2 > x_1 > 0$, hubiéramos llegado al mismo resultado intercambiando los roles de x_1 e x_2 . Por otro lado, si $x_1 = x_2$,

$$x_1^a x_2^b = x_1^{a+b} = x_1 = ax_1 + bx_1.$$

□

Proposición A.2.2 (*Desigualdad de Hölder*). Sean $f, g : \mathbb{R} \rightarrow \mathbb{R}$ y $p, q > 1$ tales que $\frac{1}{p} + \frac{1}{q} = 1$. Entonces

$$\int_a^b |f(x)||g(x)|dx \leq \left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} \left(\int_a^b |g(x)|^q dx \right)^{\frac{1}{q}}. \quad (\text{A.4})$$

DEMOSTRACIÓN. Por el Lema A.2.1 denotamos $x_1 = \frac{|f(x)|^p}{\int_a^b |f(x)|^p dx}$, $x_2 = \frac{|g(x)|^q}{\int_a^b |g(x)|^q dx}$, $a = \frac{1}{p}$, $b = 1 - \frac{1}{p} = \frac{1}{q}$. Sustituyendo en (A.2) obtenemos

$$\left(\frac{|f(x)|^p}{\int_a^b |f(x)|^p dx} \right)^{\frac{1}{p}} \left(\frac{|g(x)|^q}{\int_a^b |g(x)|^q dx} \right)^{\frac{1}{q}} \leq \frac{1}{p} \frac{|f(x)|^p}{\int_a^b |f(x)|^p dx} + \frac{1}{q} \frac{|g(x)|^q}{\int_a^b |g(x)|^q dx}.$$

multiplicando por $\left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} \cdot \left(\int_a^b |g(x)|^q dx \right)^{\frac{1}{q}}$ y tomando integrales en los numeradores, tenemos que

$$\int_a^b |f(x)g(x)|dx \leq \left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} \left(\int_a^b |g(x)|^q dx \right)^{\frac{1}{q}}.$$

Como $\left| \int_a^b f(x)g(x)dx \right| \leq \int_a^b |f(x)g(x)|dx$, se sigue (A.4).

□

Proposición A.2.3 (*Desigualdad de Minkowski*). Sean $f, g : \mathbb{C} \rightarrow \mathbb{C}$ y $p \in [1, +\infty)$. Entonces

$$\left(\int_a^b |f(x) + g(x)|^p dx \right)^{\frac{1}{p}} \leq \left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} + \left(\int_a^b |g(x)|^p dx \right)^{\frac{1}{p}}. \quad (\text{A.5})$$

DEMOSTRACIÓN. En el caso de $p = 1$ la demostración es la conocida desigualdad triangular:

$$\int_a^b |f(x) + g(x)|dx \leq \int_a^b (|f(x)| + |g(x)|)dx = \int_a^b |f(x)|dx + \int_a^b |g(x)|dx.$$

Si $p > 1$ y al menos uno de los sumandos del lado derecho de (A.5) es infinito, la desigualdad también se cumple. Supongamos que ambos lados de la ecuación son finitos. Definimos un conjunto auxiliar $Y = \{x \in [a, b] : |f(x)| \geq |g(x)|\}$. Entonces para todo $x \in Y$, $|f(x) + g(x)| \leq 2|f(x)|$, y para todo $x \in X \setminus Y$, $|f(x) + g(x)| \leq 2|g(x)|$. Estas cotas se usarán para demostrar que $|f(x) + g(x)|^p$ es integrable.

$$\begin{aligned} \int_a^b |f(x) + g(x)|^p dx &= \int_Y |f(x) + g(x)|^p dx + \int_{[a,b] \setminus Y} |f(x) + g(x)|^p dx \\ &\leq 2^p \int_Y |f(x)|^p dx + 2^p \int_{[a,b] \setminus Y} |g(x)|^p dx \leq 2^p \int_a^b |f(x)|^p dx + 2^p \int_a^b |g(x)|^p dx < +\infty. \end{aligned}$$

Si $\int_a^b |f(x) + g(x)|^p dx = 0$, entonces se cumple la desigualdad. Si $\int_a^b |f(x) + g(x)|^p dx > 0$, escribimos $|f(x) + g(x)|^p$ como

$$\begin{aligned} |f(x) + g(x)|^p &= |f(x) + g(x)| |f(x) + g(x)|^{p-1} \\ &\leq |f(x)| |f(x) + g(x)|^{p-1} + |g(x)| |f(x) + g(x)|^{p-1}, \end{aligned}$$

y luego integramos a ambos lados y aplicamos la Desigualdad de Hölder:

$$\begin{aligned} \int_a^b |f(x) + g(x)|^p dx \\ \leq \left[\left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} + \left(\int_a^b |g(x)|^p dx \right)^{\frac{1}{p}} \right] \cdot \left(\int_a^b |f(x) + g(x)|^{q(p-1)} dx \right)^{\frac{1}{q}}. \end{aligned}$$

De la condición de $\frac{1}{p} + \frac{1}{q} = 1$, se sigue que $q(p-1) = p$. Si dividimos a ambos lados de (A.2) entre $\left(\int_a^b |f(x) + g(x)|^p dx \right)^{\frac{1}{q}}$ y sustituimos $1 - \frac{1}{q}$ por $\frac{1}{p}$, se sigue (A.5).

□

Bibliografía

- [1] N.I. AKHIEZER and I.M. GLAZMAN. *Theory of Linear Operators in Hilber space*, Dover Publications, INC, New York, 1993.
- [2] ATKINSON, K. E. *An Introduction to Numerical Analysis* Wiley, 1989.
- [3] BERNSTEIN, S. N. Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités. *Comm. Soc. Math. Kharkow*, vol. 13(1-2) (1912-13).
- [4] DAVIS, P. J. *Interpolation and Approximation*. Dover Publications, 1975.
- [5] PINKUS, A. Weierstrass and Approximation Theory. *J. Approx. Theory*, vol. 107 (2000), pp. 1–66.
- [6] QUINTANA MATO, Y. *El Teorema de Aproximación de Weierstrass*. Seminario Gamma. [Mayo de 2021] Disponible en: <https://www.youtube.com/watch?v=C6b6Ncz-sTU>.
- [7] STONE, M. H. Applications of the theory of Boolean rings to general topology. *Trans. Amer. Math. Soc.*, vol. 41 (1937), pp. 375–481.
- [8] WEIERSTRASS, K. Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, 1885, vol. 2, pp. 633–639.

Some Topics on Approximation Theory

Eduardo Javier Águila Hernández

Facultad de Ciencias • Sección de Matemáticas

Universidad de La Laguna

alu0101156934@ull.edu.es

Abstract

This memory has been structured into three chapters and an appendix. In the first chapter the Bernstein polynomials are introduced, the famous Weierstrass Approximation Theorem is proved using them, and some additional properties are established. In the second chapter, the problem of best approximation in normed spaces is addressed, with special emphasis on the case of Pre-Hilbert spaces. Likewise, without much extension, orthogonal polynomials are analyzed, and the Least Squares Approximation problem is solved. Finally, in the third chapter, the closure and its consequences are investigated, the completeness of trigonometric powers and systems in $L^2[a, b]$ is briefly analyzed, and it ends with Müntz's theorems. Finally, in the Appendix, a Matlab code that computes Bernstein polynomials is presented and the Young, Hölder and Minkowski Inequalities are proved.

1. Introduction

In this memory, topics on Approximation Theory are studied. It begins by introducing interpolation theory and defining the Bernstein polynomials to then prove the Weierstrass Approximation Theorem. Likewise, some additional properties about the Bernstein Polynomials are given.

In the second chapter, the problem of best approximation in normed spaces and pre-Hilbert or Hilbert spaces is addressed, with various results to highlight such as Tonelli's or Chebyshev's Equioscillation Theorems. Then, the best approximation problem in Least Squares is analyzed, deducing the Normal Equations and defining the Gram Matrix, with previously seen properties of orthogonal polynomials.

In the third and last chapter, the theory of Functional Analysis and theory of closure and completeness are introduced. A section on the completeness of powers and trigonometric systems in $L^2[a, b]$ was also established.

Finally, in the Appendix, there are the proofs of the Young, Hölder and Minkowski Inequalities, and the code in *Matlab* programming language for the computation of Bernstein polynomials.

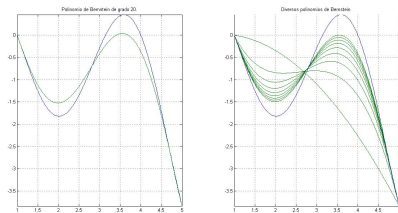


Figure 1: Graph on the slow convergence of Bernstein polynomials to the function $f(x) = (x - 1)(x - 4) \cdot \sin x$, $x \in [1, 5]$.

2. Outline

Several results are noteworthy. In the author's opinion, the most important one is the proof of the Weierstrass Approximation Theorem, since it establishes that polynomials are dense in the space of continuous functions defined on a compact of the real line. It is a very important result with a rather aesthetic proof, making use of the Bernstein polynomials.

There are many more outstanding results, like the Tonelli's and Chebyshev's Equioscillation Theorems; the Normal Equations or

examples of best approximation in the second chapter; the Riesz Theorem or the Müntz Theorems in the third chapter, which establish necessary and sufficient conditions for a non-finite sequence of powers of the variable x to be closed in $L^2[0, 1]$. Nevertheless, I will end this section with a curious and simple, but not easy figure from [4, Page 257], which elegantly differentiates the four definitions of closed and complete spaces and sequences, that might be interesting.

	Closed	Complete
Space	A subspace is closed if it contains all its limit points.	A space is complete if every Cauchy sequence has a limit in the space.
Sequence	A sequence is closed if every element of the space can be approximated arbitrarily closely by finite linear combinations of the elements of the sequence.	A sequence $\{x_n\}$ is complete if $L(x_n) = 0$, $n = 0, 1, \dots$, $L \in X^*$, implies $L = 0$.

3. Sections And Organization

In the following table it is illustrated all the information into distinct sections, such as background, objectives, and methodology.

Topics	Summary
Background	Basic prior knowledges of numerical and mathematical analysis is required.
Objectives	Prove the Weierstrass Approximation Theorem, study the best approximation problem in normed and Hilbert spaces, including the problem in Least Squares and arriving at the two Müntz Theorems.
Methodology	The references are being studied (especially the book [4]) and at the same time the outstanding results are being transferred to the Memory.

References

- [1] N.I. AKHIEZER and I.M. GLAZMAN. *Theory of Linear Operators in Hilber space*, Dover Publications, INC, New York, 1993.
- [2] ATKINSON, K. E. *An Introduction to Numerical Analysis* Wiley, 1989.
- [3] BERNSTEIN, S. N. Démonstration du théorème de Weierstrass fondée e sur le calcul des probabilités. *Comm. Soc. Math. Kharkow*, vol. 13(1-2) (1912-13).
- [4] DAVIS, P. J. *Interpolation and Approximation*. Dover Publications, 1975.
- [5] PINKUS, A. Weierstrass and Approximation Theory. *J. Approx. Theory*, vol. 107 (2000), pp. 1–66.
- [6] QUINTANA MATO, Y. *El Teorema de Aproximación de Weierstrass*. Seminario Gamma. [Mayo de 2021] Disponible en: <https://www.youtube.com/watch?v=C6b6Ncz-sTU>.
- [7] STONE, M. H. Applications of the theory of Boolean rings to general topology. *Trans. Amer. Math. Soc.*, vol. 41 (1937), pp. 375–481.
- [8] WEIERSTRASS, K. Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, 1885, vol. 2, pp. 633–639.