El teorema de Weierstrass y la teoría de aproximación

Jerónimo Basa *

23 de julio de 2015

Resumen

En el presente texto se muestra una pequeña reseña sobre la teoría de aproximación de funciones en el área del análisis matemático. Sumado a esto, se manifiesta uno de los teoremas más importantes en este marco: *el teorema de aproximación de Weierstrass*. Las secciones están ordenadas a modo de comenzar con una introducción (que el lector con conocimientos puede obviar tranquilamente) seguido de algunos resultados elementales. Como siempre, la sección de apéndice permite complementar el desarrollo, sin impedir una lectura fluida de la presente exposición.

Índice

*Estudiante de Licenciatura en Matemática por la Universidad Nacional del Litoral

A Federico Font, por sus enormes
cualidades matemáticas
y su virtud de compañerismo
que siempre agradeceré.

Prólogo

El teorema de aproximación de Weierstrass fue originalmente desarrollado por el matemático Karl Weierstrass en su trabajo de 1885. El enunciado, que fue probado por él, puede establecerse de la siguiente manera. Supongamos que f es una función continua en valores complejos definida sobre un intervalo real $[a, b]$. Para cada $\varepsilon > 0$, existe una función polinómica p sobre $\mathbb{C}$ tal que para todo $x \in [a, b]$, tenemos $|f(x) - p(x)| < \varepsilon$, o equivalentemente, la norma uniforme (o supremo) $||f - p||_\infty < \varepsilon$. En otras palabras, el teorema de aproximación afirma que toda función continua definida en un intervalo compacto $[a, b]$ puede ser uniformemente aproximada por una función polinómica. Este teorema ha logrado un gran impacto tanto en la comunidad del análisis matemático (porque los polinomios son las funciones más simples de manejar), como en sus aplicaciones (entre otras en el área del cálculo numérico). Como consecuencia del teorema de aproximación de Weierstrass, uno puede probar que el espacio $\mathcal{C}([a, b], \mathbb{R})$ es separable, como veremos más adelante. En 1937 [9] Marshall Stone generalizó el teorema y luego simplificó la demostración en 1948 [10]. Hoy conocemos estos resultados como el teorema de Stone-Weierstrass. Éste generaliza el teorema de aproximación de Weierstrass en dos maneras. Primero, en lugar de intervalos reales $[a, b]$, se considera un espacio compacto de Hausdorff arbitrario. Y en lugar del álgebra de funciones polinómicas, Stone investigó la aproximación con elementos de otras álgebras más generales, $\mathcal{C}(K, \mathbb{R})$.

Stone comenzó con un espacio compacto de Hausdorff K arbitrario y consideró el álgebra $\mathcal{C}(K, \mathbb{R})$ de funciones continuas reales sobre K, con la topología de la convergencia uniforme. Él quería encontrar otras álgebras de $\mathcal{C}(K, \mathbb{R})$ que sean densas, y notó que la propiedad necesaria que debe satisfacer dicha estructura es que separe puntos. Ahora, podemos enunciar el teorema de la siguiente forma. Sea K un espacio compacto de Hausdorff y $\mathcal{A}$ un álgebra de $\mathcal{C}(K, \mathbb{R})$ que contiene una función no constantemente cero. Entonces $\mathcal{A}$ es denso en $\mathcal{C}(K, \mathbb{R})$ sí, y sólo sí, separa puntos. Esto implica el enunciado original de Weierstrass ya que los polinomios sobre $[a, b]$ forman un álgebra de $\mathcal{C}([a, b], \mathbb{R})$ que contiene las funciones constantes y separa puntos.

En una forma general, podemos decir que la teoría de aproximación se dedica a la descripción de elementos en un espacio topológico X, que pueden ser aproximados por elementos en un subconjunto A de X, dicho de otra manera, la teoría de aproximación permite caraterizar la clausura de A en X. Uno de los resultados relevantes en este aspecto es la densidad de los polinomios trigonométricos en la clase de las funciones reales continuas de período 2π. Tales resultados sirvieron para un famoso ejemplo de Weierstrass de 1861 sobre la existencia de funciones continuas que no son diferenciables en su dominio. La existencia de estas funciones acentúa la necesidad de rigor analítico en la matemática, para una mayor comprensión sobre la naturaleza del conjunto de las funciones continuas, y en consecuencia, influenciar el desarrollo del análisis.

Finalmente, como dato interesante y anecdótico, cuando Weierstrass probó el teorema, tenía 70 años. Por otro lado, veinte años después, otra prueba fue dada por Fejér, a la joven edad de 19 años. Y más sorprendente aún es el hecho de que Fejér era considerado débil en matemáticas en la escuela, y necesitaba de apoyo especial. Su prueba la realiza a través de series de Fourier, obteniendo que el teorema de Weierstrass es equivalente a su versión periódica.

1. Conceptos necesarios y discusión general

Para comenzar, consideramos útil tener en claro ciertas cosas, como las nociones básicas, definiciones, y todo fundamento que nos ayude en el desarrollo del texto. Desde luego, nosotros no hemos demostrado nada nuevo, somos simples narradores intentando hacer llegar al lector los bellos resultados, aplicaciones y, en general, los detalles interesantes que podemos extraer sobre este tema. Ya hemos mencionado en el prólogo algunos puntos sobre este teorema, nos gustaría sin embargo destacar los puntos importantes e ir progresando a medida que avanzamos.

Como primer item a destacar, el teorema trata sobre funciones, pero no de cualquier tipo. Vimos que, en particular, nos interesará trabajar con funciones continuas.

1.1. Continuidad, espacios métricos y polinomios

Definición 1.1. Sean $f : \mathbb{R} \to \mathbb{R}$ y a un número real. Diremos que f es continua en a si, para todo ε mayor que cero, existe un $\delta(a, \varepsilon)$ positivo, tal que si la distancia entre x y a es menor que δ, entonces la distancia entre las respectivas imágenes es menor que ε. Si una función cumple la condición de continuidad para todo elemento en su dominio, decimos que f es continua en todo su dominio.[1]

Muchas de las ideas del análisis (como la continuidad) pueden definirse para espacios más generales llamados *espacios métricos*. Un espacio métrico, que denotaremos por X, es un espacio con dos características. La más intuitiva, es que posee elementos;[2] la segunda, es que existe una *distancia* entre ellos, que llamaremos *la métrica del espacio*. Por ser justamente una *distancia* podemos imaginarnos su naturaleza, como las propiedades que deba cumplir. Resumiremos esto diciendo:

Definición 1.2. Sea X un espacio con elementos en él (ya sean números, funciones, matrices, etcétera) y sean a, b y c elementos distintos de X. Diremos que una métrica (o distancia) es una función $d : X \times X \to \mathbb{R}^+$ que cumple las siguientes cuatro propiedades

- $d(a, a) = 0$.
- $d(a, b) = d(b, a)$.
- $d(a, b) > 0$.
- $d(a, b) \leq d(a, c) + d(c, b)$.

Donde la última propiedad es un conocido principio llamado desigualdad triangular.

Decimos entonces que el par (X, d) es un espacio métrico.[3] Resulta entonces sencillo escribir nuevamente la definición de continuidad en funciones sobre espacios métricos. Como mencionamos, sólo necesitamos una ligera modificación, que es considerar las distancias de los espacios. Formalmente la definición nos dice que si (X, d_X) e (Y, d_Y) son espacios métricos, $f : X \to Y$ es continua en a si

$$\forall \varepsilon > 0, \exists \delta > 0 : \ d_X(x, a) < \delta \ \Rightarrow \ d_Y(f(x), f(a)) < \varepsilon.$$

Si estamos en $\mathbb{R}$, la distancia valor absoluto nos da la definición usual de continuidad.

[1] En general, decimos que f *es continua* a secas.
[2] Nótese que descartamos la opción de considerar el conjunto vacío $\emptyset$.
[3] En caso de que la métrica no sea del todo relevante, nos referiremos al mismo como *espacio métrico X*.

Desde luego, podríamos preguntarnos el porqué nos interesan tanto las funciones continuas. No es necesario recordar que las funciones sirven como modelo de sistemas que vemos en la realidad, y del cual podemos obtener muchas aplicaciones. Pensemos primero en el teorema del valor intermedio. Éste nos dice que si una función continua toma un valor negativo en un punto x_1 y luego toma un valor positivo en otro punto x_2, entonces existe al menos un punto entre x_1 y x_2 donde la función vale cero. Supongamos entonces que usamos una de estas funciones continuas para hallar la temperatura en cada punto de la Tierra (aquí la consideraremos una esfera homogénea). Hallamos las temperaturas en el polo norte y sur, y luego tomamos su diferencia; llamaremos α a esta diferencia. Ahora, comenzaremos desde el polo norte y nos moveremos hacia el polo sur, tomando la temperatura en cada punto a medida que describimos nuestro recorrido,[4] hasta finalmente llegar al extremo sur. Más aún, si estudiamos la diferencia entre puntos *opuestos* (en el sentido de que están en distintos hemisferios) que viven sobre el mismo meridiano, esperamos que al llegar al polo sur la diferencia sea $-\alpha$. El teorema del valor intermedio nos asegura que, examinando las diferencias de temperaturas, deben haber dos puntos opuestos sobre la Tierra con la misma temperatura, pues la diferencia se hace cero en algún momento.

Ahora bien, que una función continua sea útil y agradable, ¿la hace sencilla? En un amplio margen, las funciones continuas son más sencillas que las que no lo son, pero podemos encontrarnos funciones continuas que no lo sean.

Aprovechando que hemos invocado la palabra "sencillas", recordemos que fue mencionada anteriormente, pero en referencia a los polinomios. Esto fue porque hablamos sobre el carácter especial de sencillez que estos poseen. Un polinomio no es más que una expresión de la forma

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0, \qquad a_i \in \mathbb{R}, \quad 0 \leq i \leq n.$$

Es trivial probar que toda función polinómica es una función continua, y por lo tanto omitiremos dicha prueba. Debido a su sencilla expresión, los polinomios son muy sencillos de evaluar numéricamente, lo que los hace una gran herramienta en análisis numérico.

Antes de finalizar, veamos algo particular. Podemos decir que lo maravilloso de dichas funciones es que podemos dibujar su gráfica *sin levantar el lápiz*[5], lo que se traduce en la ausencia de "huecos" y "saltos bruzcos". Pero podemos ir aún más lejos y volverlo un poco más elegante.

Definición 1.3. Una función f definida en un intervalo I de $\mathbb{R}$ se dice que es *uniformemente continua* en I si, dado $\varepsilon > 0$, existe $\delta = \delta(\varepsilon) > 0$ tal que, si x y x' verifican que $|x - x'| < \delta$ entonces $|f(x) - f(x')| < \varepsilon$.

A primera vista esta definición no parece diferente a la que ya vimos sobre continuidad. Pero veámoslo más de cerca. Si una función es continua en un intervalo I, eso significa que, para cada x_0 en I, f es cotinua en x_0. Esto a su vez significa que dado $\varepsilon > 0$ existe $\delta > 0$ tal que si $|x - x_0| < \delta$ resulta $|f(x) - f(x_0)| < \varepsilon$. Esto se puede hacer para cada x_0 en I, pero la forma en que depende δ de ε puede cambiar cuando cambia x_0. Lo que pide la definición de continuidad uniforme es que haya una función $\delta(\varepsilon)$ que sirva *para todos los* $x_0 \in I$. Claramente, si una función es uniformemente continua, entonces es continua. El recíproco es cierto sólo en conjuntos compactos. La demostración se encuentra en (A.1).

[4]Siguiendo una geodésica, para ser más precisos.

[5]De hecho, deberíamos pedir que la función sea también derivable para poder hacerlo. Por ejemplo, la función de Weierstrass nunca podríamos dibujarla.

1.2. La norma uniforme

A continuación hablaremos sobre otro detalle mencionado en el prólogo y que veremos mucho a lo largo de la monografía. De una manera muy similar a lo que vimos sobre métrica, decimos que una norma sobre un espacio $\mathcal{C}$ es una función $\rho : \mathcal{C} \to \mathbb{R}^+$ tal que valen las siguientes propiedades:

- $\rho(af) = |a|\rho(f)$ para todo número real a y todo elemento f de $\mathcal{C}$.
- $\rho(f+g) \leq \rho(f) + \rho(g)$ para todo par de elementos distintos f, g de $\mathcal{C}$.
- $\rho(f) = 0 \iff f = 0$.

Un ejemplo muy familiar es la norma Euclídea, la que usamos para hallar la distancia entre dos puntos. En $\mathbb{R}^2$ decimos que si $x = (x_1, x_2)$ entonces su norma $|| \cdot ||_2 : \mathbb{R}^2 \to \mathbb{R}^+$ se define por

$$\|x\|_2 = \sqrt{(x_1^2 + x_2^2)}.$$

Definición 1.4. Para todo número real $p \geq 1$, definimos la norma p de $x \in \mathbb{R}^n$ como

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{\frac{1}{p}} = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

En particular, se define la norma uniforme (o infinito) como el límite de la norma p cuando $p \to \infty$. Lo que equivale a decir que

$$\|x\|_\infty = \text{máx}\{|x_1|, |x_2|, \ldots, |x_n|\}.$$

Ya que las funciones continuas son objeto central de nuestro relato, definamos el espacio formado por todas las funciones continuas que toman valores en un intervalo cerrado, por ejemplo $I = [a, b]$. Denotaremos entonces con $\mathcal{C}[a, b]$ al espacio en cuestión.[6] Por lo tanto, sea $f \in \mathcal{C}[a, b]$, entonces vemos que la norma infinito es el máximo valor absoluto de $f(x)$ sobre $[a, b]$, es decir

$$\|f\|_\infty = \underset{a \leq x \leq b}{\text{máx}} |f(x)|.$$

Puede demostrarse fácilmente que toda norma induce una métrica, y por lo tanto todo espacio normado es un espacio métrico. En particular podemos definir lo siguiente.

Definición 1.5. Sea $\mathcal{C}[a, b]$ el espacio de las funciones continuas en el intervalo $[a, b]$. Entonces la métrica

$$d_\infty(f, g) = \|f - g\|_\infty = \underset{a \leq x \leq b}{\text{máx}} |f(x) - g(x)|$$

define el espacio métrico $(\mathcal{C}[a, b], d_\infty)$.

Más aún, si en lugar de realizar una suma sobre un conjunto finito, integramos de la siguiente manera

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

[6] En principio no estamos diciendo si estas funciones son a valores reales o complejos, pero no será de vital importancia aún.

podemos definir el espacio de las funciones bajo la norma p mediante la anterior integral de Riemann con $f \in \mathcal{C}[a,b]$.

De ahora en adelante utilizaremos estas definiciones cuando hagamos referencia al espacio sobre el que trabajamos y la distancia entre ellos, pues veremos repetidas veces la noción de "hacer chica una diferencia". Gracias a las últimas observaciones, vamos a comprender el sentido con que usaremos dicha frase. Desde luego, nuestro principal objetivo será mostrar que podemos tomar polinomios a una distancia muy chica de toda función continua, es decir, que los polinomios son densos en las funciones continuas.

1.3. ¿Denso en qué sentido?

No podemos dejar pasar por alto un dato interesante. Hemos dicho que podemos probar que un cierto conjunto (el de los polinomios, para ser exactos) es *denso*. Lo que nos lleva a una pregunta, que no la repetiremos, pues dá el nombre a esta sección. Primero, daremos la noción de *entorno*.

Definición 1.6. Sea X un espacio métrico y sea x un elemento. Decimos que una *bola abierta* de centro x es el conjunto de todos los elementos de X que están a una distancia no superior a r (con respecto de x), para un r positivo. Denotamos esto con

$$N_x(r) = \{y \in X : d(x,y) < r\}.$$

Esto es un ejemplo de lo que en topología se conoce como *entorno abierto* (no daremos la definición formal de ello).[7] Supongamos que K es un subconjunto del espacio X. De manera informal, decimos que K es denso en X si, para todo elemento $x \in X$, el elemento está en K, o bien se encuentra "arbitrariamente cerca" de algún elemento de K. Formalmente decimos que K es denso en X, si para cada elemento $x \in X$, todo entorno N_x contiene al menos un elemento de K. El ejemplo más sencillo que conocemos sobre esta idea es la densidad de los números racionales sobre los números reales. Decimos que, para cada número real, siempre podemos encontrar números racionales tan próximos a éste como queramos (véase [2]). Nuestro teorema central nos dice que los polinomios pueden tomarse arbitrariamente cerca a toda función continua. Es decir, las funciones polinómicas son un conjunto denso en el espacio $\mathcal{C}[a,b]$ de funciones continuas en $[a,b]$ con la norma uniforme 1.5.

1.4. Sucesiones y compacidad

Una sucesión (x_n) de números reales es una lista ordenada de números $x_n \in \mathbb{R}$, llamados los términos de la sucesión, donde el subíndice n es un natural. De hecho, si definimos esto formalmente en un contexto más general, vemos que una sucesión no es más que una función.

Definición 1.7. Una sucesión (x_n) de elementos en un conjunto A es una función $f : \mathbb{N} \to A$, donde $x_n = f(n)$.

Veamos un caso particular, para una sucesión de números reales. Si se representan los diversos valores de x_n para una sucesión dada, sobre la recta, se puede tener una idea del comportamiento que ésta tiene. Consideremos, por ejemplo, la sucesión $(x_n) = \frac{1}{n}$. Listando los primeros elementos

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{n}, \dots$$

[7] Nótese que la forma geométrica de una bola abierta depende del espacio y de la métrica.

Geométricamente parece claro que mientras tomamos sucesivos valores de x_n, estos se van acercando a 0. Es decir, cualquier elemento x_n está más cerca de 0 que x_m si $n > m$. Pero esta observación no nos dice nada sobre el comportamiento de la sucesión, pues también es cierto que los sucesivos valores de $\frac{1}{n}$ se van acercando a -1. Ahora bien, esta sucesión tiene una cierta particularidad. Su acercamiento al cero es más especial que su acercamiento al -1. Pues esta sucesión *llega a estar tan cerca de cero como se quiera*. En cambio, con respecto a -1 esto no ocurre, pues en este caso los términos x_n están siempre a distancia mayor que 1. Esta característica especial de que exista un objeto tal que la sucesión *se encuentre tan cerca a él como se quiera* define un tipo particular de sucesiones.

Definición 1.8. Sea (x_n) una sucesión. Diremos que (x_n) converge a x [8] si cualquiera sea el número real $\varepsilon > 0$, hay un número natural N, tal que, para $n \geq N$, se tiene que $|x_n - x| < \varepsilon$. Si una sucesión cumple esta definición, entonces se llama *convergente*.

Si tenemos una sucesión (x_n), una *subsucesión* de ella consiste en elegir algunos de los x_n y formar así una nueva sucesión usando dichos términos. De esta manera, una subsucesión de (x_n) es una lista de la forma

$$x_{n_1}, x_{n_2}, x_{n_3}, \ldots, x_{n_k}, \ldots$$

con

$$n_1 < n_2 < n_3 < \ldots$$

Definición 1.9. Sea $x : \mathbb{N} \to \mathbb{R}$ una sucesión. Una subsucesión de (x_n) es la composición

$$x \circ k : \mathbb{N} \to \mathbb{R}$$

de (x_n) con una función estrictamente creciente $k : \mathbb{N} \to \mathbb{N}$. Indicamos la subsucesión con (x_{n_k}).

Supongamos que estamos en el espacio de las funciones continuas en el intervalo $[a, b]$, y sea (f_n) una sucesión allí. Diremos que la misma *converge uniformemente* si para todo ε positivo y para todo $x_0 \in [a, b]$, existe un $N = N(\varepsilon)$ (que no depende del x_0 tomado), tal que $|f_n(x_0) - f(x_0)| < \varepsilon$ si $n > N$.

Informalmente decimos que una sucesión pertenece a un espacio si cada término de la sucesión está efectivamente en dicho espacio. Un *espacio compacto* confina cada sucesión de puntos en él de tal forma que la sucesión debe acumularse en algún punto del espacio. Podemos pensar que el espacio *apreta* la sucesión tal que logra mantenerla (o al menos mantener una parte) dentro del mismo. A continuación damos la definición rigurosa.

Definición 1.10. Sea X un espacio métrico y $K \subseteq X$. Entonces K es compacto si toda sucesión en K tiene una subsucesión que converge a un elemento de K.

Ejemplo 1.11. El intervalo abierto $I = (0, 1)$ no es compacto. La sucesión $\left(\frac{1}{n}\right)$ en I converge a 0, luego toda subsucesión también converge a 0, pero $0 \notin I$.

Entrando en el ámbito de la topología, puede darse una definición para espacios compactos que es equivalente a la que fue presentada. Más aún, puede probarse que, en $\mathbb{R}$, los conjuntos compactos son conjuntos cerrados y acotados, y viceversa.[9]

Además de la compacidad, nuestra última observación caerá sobre los espacios de Hausdorff. Podemos imaginar a estos como espacio que *separan* puntos distintos en entornos que no se tocan.

[8] O bien, que x es el límite de la sucesión (x_n).

[9] Este resultado, llamado *Teorema de Heine-Borel*, será el que usaremos más adelante.

Definición 1.12. Sea X un espacio métrico (no necesariamente compacto). Entonces X es *espacio de Hausdorff* si para cada par de puntos distintos x, y existen entornos abiertos de cada uno, y que son disjuntos. Formalmente

$$\forall x, y \in X,\ \exists\, N_x \wedge N_y :\ x \in N_x,\ y \in N_y,\ N_x \cap N_y = \emptyset.$$

1.5. Cosas llamadas Álgebras

Veremos rápidamente una forma distinta de interpretar la palabra *álgebra*. Para esto, usaremos la noción puramente intuitiva de una *familia de funciones*, cuyos elementos llamaremos normalmente con las letras f, g. Puede advertirse que lo expuesto a continuación es muy similar a la noción de *espacio vectorial*.

Definición 1.13. Sea $\mathscr{A}$ una familia de funciones reales definidas en un conjunto X. Diremos que $\mathscr{A}$ es un *álgebra* si

1. $f + g \in \mathscr{A}$,
2. $fg \in \mathscr{A}$,
3. $cf \in \mathscr{A}$,

para todo $f \in \mathscr{A}$, $g \in \mathscr{A}$ y para toda constante real c. Es decir, $\mathscr{A}$ es cerrado bajo la suma, el producto y la multiplicación por un escalar.

Ahora que tenemos esta nueva interpretación de un álgebra, veamos dos casos particulares. Esto junto con la noción de espacios de Hausdorff son las herramientas en el teorema general de Stone-Weierstrass. Observaremos que la idea de *separar puntos* como lo vimos en estos últimos espacios, puede también tomarse para álgebras de funciones con la siguiente

Definición 1.14. Sea $\mathscr{A}$ un álgebra de funciones sobre un conjunto X. Decimos que $\mathscr{A}$ *separa puntos*[10] en X si para todo par de puntos distintos $x_1, x_2 \in X$ existe una función $f \in \mathscr{A}$ tal que $f(x_1) \neq f(x_2)$. Si para cada $x \in X$ existe una función $f \in \mathscr{A}$ tal que $f(x) \neq 0$, se dice que $\mathscr{A}$ *no se anula en ningún punto* de X.

En este punto, resulta trivial ver que la familia de polinomios forman un álgebra con estas propiedades sobre un subconjunto de $\mathbb{R}$.

Ejemplo 1.15. Un ejemplo de un álgebra que no separa puntos es el conjunto de todas las funciones pares sobre $[-1, 1]$, pues $f(-x) = f(x)$ para cada función par f.

2. El teorema de aproximación de Weierstrass

El teorema mencionado en el título, es uno de los más importantes a la hora de hablar sobre aproximación de funciones mediante polinomios, que como ya vimos, son muy útiles a nivel práctico. La técnica más conocida en este proceso recibe el nombre de *interpolación polinómica de funciones* que permite aproximar funciones continuas. Muchos estudiantes de nivel básico están acostumbrados a trabajar con esto. De hecho, en un gran porcentaje, se sabe que la primera visión que tenemos de este hecho es el desarrollo por serie de Taylor. Podríamos preguntarnos entonces cuál es la particularidad del teorema de Weierstrass. A modo de mostrar esto, primero presentaremos los dos siguientes resultados.

[10] No confundir la separabilidad de este caso con la definición dada para un conjunto de Hausdorff.

Teorema 2.1 (**Hermite**). *Toda función f continua puede expresarse como serie infinita en términos de polinomios de Hermite*

$$f(x) = \sum_{n=0}^{\infty} A_n H_n(x),$$

donde

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}$$

$$A_k = \frac{1}{2^k k! \sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-x^2} f(x) H_k(x) dx$$

Teorema 2.2 (**Taylor**). *Si $n \geq 0$ es un entero y f una función que es derivable n veces en el intervalo cerrado $[a, x]$ y $n+1$ veces en el intervalo abierto (a, x), entonces se obtiene la aproximación polinómica de Taylor*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(f)$$

donde

$$R_n(f) = \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n dt.$$

Vemos, en estos dos teoremas, que efectivamente una función f continua puede obtenerse por series de polinomios. Pero recordemos el enunciado del teorema de Weierstrass. Éste nos dice que tal aproximación existe siempre que f sea continua en un intervalo $[a, b]$. Si observamos los dos teoremas anteriores, las hipótesis de Weierstrass son más débiles que las presentadas allí.[11] Sin embargo, como desventaja, puede advertirse que este teorema no es en manera constructivo; no nos dice quiénes son los polinomios que participan en tal aproximación.[12]

2.1. Enunciado y demostración

De manera un tanto indirecta, se intentó enunciar el teorema de Weierstrass en su forma más sencilla en el prólogo. Y a pesar de que fue mencionado el carácter no constructivo del mismo, a manera contradictoria, presentaremos una demostración que resultará de la *construcción de polinomios*[13] adecuados. Una demostración realmente bella puede deducirse usando herramientas poderosas en análisis llamadas *aproximaciones a la identidad*.

Comenzaremos notando que

$$\int_0^1 (n+1)(1-y^2)^n \, 2y dy = 1.$$

Como $y \leq 1$ en $[0, 1]$, se sigue que $\int_0^1 (n+1)(1-y^2)^n 2dy > 1$, y hay un $k_n < n+1$ tal que $\int_z^1 k_n(1-y^2)^n 2dy$ es un polinomio $P_n(z)$ tal que $P_n(1) = 0$, $P_n(0) = 1$ y (por simetría de $(1-y^2)^n$ sobre el eje y) $P_n(-1) = 2$.

[11] Nótese que Taylor pide la derivabilidad en el orden de $n+1$ veces para el intervalo abierto. A su vez las condiciones de integrabilidad de Hermite son más fuertes aún.

[12] Más adelante podrá verse que es posible representar estos polinomios.

[13] Adviértase las letras cursivas.

Ahora para b fijo entre 0 y 1, tenemos que $\lim_{n\to\infty}(n+1)b^n = 0$. Se sigue entonces que dado $\varepsilon > 0$ y dado δ con $0 < \delta < 1$, hay un n tal que $(n+1)(1-y^2)^n < \varepsilon/2$ para $\delta \leq y \leq 1$, y entonces $P_n(\delta) < \varepsilon$. De manera similar $P_n(-\delta) > 2-\varepsilon$. Como $P_n(x)$ es una función decreciente en $[-1,1]$, se sigue que $P_n(x)$ está entre 2 y $2-\varepsilon$ sobre $[-1,-\delta]$ y está entre ε y 0 sobre $[\delta, 1]$.

Sea t entre 0 y 1, y sea $Q_n(x) = P_n(x^2 - t)$. Entonces $Q_n(x)$ está entre 2 y $2-\varepsilon$ sobre $[-h,h]$ si $h^2 = t-\delta$, y $Q_n(x)$ está entre 0 y ε sobre $[-s,-\zeta]$ y $[\zeta, s]$ con $\zeta^2 = t+\delta$ y $\zeta < s < 1$.

Como t, s y δ están a nuestra disposición, mediante un apropiado cambio de escala tenemos el siguiente

Lema 2.3. *Sea $0 < a < b < c$ y $m > \varepsilon > 0$. Entonces hay un polinomio $Q_n(x)$ que es creciente en $[-c,0]$ y decreciente en $[0,c]$, y tal que $Q_n(x)$ está entre $m-\varepsilon$ y m sobre $[-a,a]$ y está entre 0 y ε sobre $[-c,-b]$ y $[b,c]$.*

Teorema 2.4 (Weierstrass). *Sea $f(x)$ una función real continua sobre un intervalo compacto $[a,b]$. Dado $\varepsilon > 0$, existe un polinomio $p(x)$ tal que $|f(x)-p(x)| < \varepsilon$ para todo $x \in [a,b]$.*

Demostración. Supongamos sin pérdida de generalidad, que el rango de f es $[0,M]$ sobre $[a,b]$. Es suficiente probar cómo encontrar un polinomio $p_1(x)$ tal que $f - p_1$ tiene rango en $[0, .8M]$. Si podemos hacer esto, lograremos encontrar un polinomio p_2 tal que $f-p_1-p_2$ tiene rango en $[0, .8^2M]$, y por tanto hay un k tal que $.8^kM < \varepsilon$. El polinomio $p = p_1+p_2+\cdots+p_k$ es tal que $|f(x)-p(x)| < \varepsilon$ para todo $x \in [a,b]$. Como f es continua en $[a,b]$, es uniformemente continua, y hay un k tal que si $x, y \in [a,b]$ con $|x-y| < (b-a)/k$ entonces $|f(x)-f(y)| < .1M$. Sea $a = x_0 < x_1 < x_2 < \ldots < x_n = b$ una partición de $[a,b]$ tal que $x_{i+1} - x_i = (b-a)/k$. Consideremos los intervalos $[x_j, x_{j+1}]$ tales que $f(y) \geq \frac{1}{2}M$ si $y \in [x_j, x_{j+1}]$, y sea P el conjunto de los puntos en ese intervalo. Entonces P es la unión finita de intervalos cerrados disjuntos (no tienen puntos finales en común) $I_1, \ldots, I_r$ tal que cada I_i se encuentra *bordeado* por intervalos H_i y J_i de longitud $(b-a)/2k$, con $f(x)$ en el rango $[.4M, .6M]$ sobre J_i y sobre H_i.

Ahora aplicamos el lema anterior con $m = \frac{1}{2}M$ y $\varepsilon = M/10r$. Para cada i hay un polinomio $Q_i(x)$ tal que $Q_i(x)$ está entre $\frac{1}{2}M$ y $\frac{1}{2}M - \varepsilon$ sobre I_i, y $Q_i(x)$ está entre 0 y ε sobre el complemento de $(I_i \cup H_i \cup J_i)$ en $[a,b]$. Entonces el polinomio $p_1(x) = Q_1(x)+\cdots+Q_r(x) - 2M/10$ tiene la propiedad de que $f - p_1$ tiene rango en $[0, .8M]$, como debíamos probar.

∎

2.2. Los polinomios de Bernstein

Como ya dijimos, a pesar de que el teorema afirma la existencia de polinomios que convergen uniformemente a una función continua, no es trivial encontrar dichos polinomios. Sin embargo, puede darse una demostración utilizando los *polinomios de Bernstein* que tiene un carácter probabilístico en su desarrollo que se encuentran en [4]. A continuación se presenta una demostración alternativa usando esta técnica, que no es la única, pero sí una de las más conocidas y encontrada en diversa bibliografía.

Definición 2.5. Sea $f : [0,1] \to \mathbb{R}$ una función. Definimos el n-ésimo *polinomio de Bernstein*, asociado a f como

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

para cada $n \in \mathbb{N}$, donde $\binom{n}{k}$ es el coeficiente binomial.[14]

Teorema 2.6. *Sea $f : [0,1] \to \mathbb{R}$ una función continua. Entonces $\{B_n\}_{n=1}^{\infty}$ converge uniformemente a f en $[0,1]$.*

Demostración. Primero necesitamos computar algunos resultados preliminares. Sean $p, q \in \mathbb{R}$ y $n \in \mathbb{N}$. Entonces por el teorema del binomio

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n. \tag{2.1}$$

Derivando ambos miembros con respecto de p obtenemos

$$\sum_{k=0}^{n} \binom{n}{k} k p^{k-1} q^{n-k} = n(p+q)^{n-1}.$$

Entonces resulta

$$\sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} p^k q^{n-k} = p(p+q)^{n-1}. \tag{2.2}$$

Derivando nuevamente

$$\sum_{k=0}^{n} \frac{k^2}{n} \binom{n}{k} p^{k-1} q^{n-k} = p(n-1)(p+q)^{n-2} + (p+q)^{n-1},$$

$$\sum_{k=0}^{n} \frac{k^2}{n^2} \binom{n}{k} p^k q^{n-k} = p^2 \left(1 - \frac{1}{n}\right)(p+q)^{n-2} + \frac{p}{n}(p+q)^{n-1}. \tag{2.3}$$

Sea $x \in [0,1]$ y sean $p = x$ y $q = 1 - x$. Entonces por (2.1)

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = 1. \tag{2.4}$$

También, aplicando (2.2),

$$\sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k} = x, \tag{2.5}$$

y por (2.3),

$$\sum_{k=0}^{n} \frac{k^2}{n^2} \binom{n}{k} x^k (1-x)^{n-k} = x^2 \left(1 - \frac{1}{n}\right) + \frac{x}{n}. \tag{2.6}$$

14

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

De estas últimas ecuaciones, podemos simplificar la siguiente expresión

$$\begin{aligned}
& \sum_{k=0}^{n}\left(\frac{k}{n}-x\right)^2\binom{n}{k}x^k(1-x)^{n-k} \\
= & \sum_{k=0}^{n}\left(\frac{k^2}{n^2}-2x\frac{k}{n}+x^2\right)\binom{n}{k}x^k(1-x)^{n-k} \\
= & \sum_{k=0}^{n}\frac{k^2}{n^2}\binom{n}{k}x^k(1-x)^{n-k}-2x\sum_{k=0}^{n}\frac{k}{n}\binom{n}{k}x^k(1-x)^{n-k}+x^2\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k} \\
= & x^2\left(1-\frac{1}{n}\right)+\frac{x}{n}-2x\cdot x+x^2\cdot 1 \\
= & x^2-\frac{x^2}{n}+\frac{x}{n}-2x^2+x^2 \\
= & x\frac{1-x}{n}.
\end{aligned}$$

Luego

$$\sum_{k=0}^{n}\left(\frac{k}{n}-x\right)^2\binom{n}{k}x^k(1-x)^{n-k}=x\frac{1-x}{n}$$

para $x \in [0,1]$.

Como f es continua en $[0,1]$ y este intervalo es compacto, $f([0,1])$ es compacto (A.3). Entonces existe un número real $M > 0$ tal que

$$|f(x)| < M$$

para todo $x \in [0,1]$. Sea $\varepsilon > 0$. Como f es continua en $[0,1]$, es uniformemente continua allí y existe $\delta > 0$ tal que si $|x-y| < \delta$ entonces $|f(x)-f(y)| < \varepsilon/2$ para todo $x, y \in [0,1]$. Fijado x en $[0,1]$. Ahora observamos, por (6.4) y la definición de polinomios de Bernstein

$$\begin{aligned}
|f(x)-B_n(f)(x)| &= \left|f(x)\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}-\sum_{k=0}^{n}f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}\right| \\
&= \left|\sum_{k=0}^{n}\left(f(x)-f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k}\right| \\
&\leq \sum_{k=0}^{n}\left|f(x)-f\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k}.
\end{aligned}$$

Cuando $\left|x-\frac{k}{n}\right| < \delta$ tenemos

$$\left|f(x)-f\left(\frac{k}{n}\right)\right| < \frac{\varepsilon}{2}.$$

Ahora, si $\left|x-\frac{k}{n}\right| \geq \delta$, entonces

$$\left(x-\frac{k}{n}\right)^2 \geq \delta^2$$

$$\frac{\left(x-\frac{k}{n}\right)^2}{\delta^2} \geq 1$$

y así

$$\left| f(x) - f\left(\frac{k}{n}\right) \right| \leq 2M \leq 2M \frac{\left(x - \frac{k}{n}\right)^2}{\delta^2}.$$

Podemos, por lo tanto, continuar la anterior expresión de manera que

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

porque $x(1-x) < 1$. Ahora podemos elegir $n \in \mathbb{N}$ suficientemente grande tal que $n \geq \frac{4M}{\delta^2 \varepsilon}$. Finalmente

$$|f(x) - B_n(f)(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

como queríamos probar. ■

2.3. ¿Se pueden debilitar las hipótesis?

Realizaremos algunas observaciones respecto de las condiciones necesarias sobre el dominio de las funciones continuas, primero analizando qué podría pasar si no pedimos que el intervalo dado sea cerrado, y luego si el mismo no es acotado. Para ello exhibiremos dos funciones continuas que no pueden ser uniformemente aproximadas en cada caso (por polinomios).

Veamos primero la condición de que el intervalo dado sea cerrado, es obligatoriamente necesaria. Supongamos, sin pérdida de generalidad, que el intervalo dado es de la forma $(a, b]$, y definamos la función $\varphi : (a, b] \to \mathbb{R}$, $\varphi(t) = (t-a)^{-1}$. De esta manera φ es continua en el $(a, b]$, y

$$\lim_{t \to a^+} \varphi(t) = \infty.$$

Por lo tanto, considerando que todo polinomio $p(t)$ tiene continuidad en a, tendremos que

$$||\varphi(t) - p(t)||_\infty = \sup_{t \in (a,b]} |\varphi(t) - p(t)| = \infty.$$

(De hecho deberíamos decir que, dado que el conjunto $A = \{|\varphi(t) - p(t)|, t \in (a, b]\}$ no está acotado superiormente, no existe el supremo del mismo).

Ahora bien, supongamos que el dominio no es un intervalo de longitud finita. Nuevamente, sin pérdida de generalidad, suponemos que el mismo es de la forma $[a, \infty)$, y definimos $\psi : [a, \infty) \to \mathbb{R}$, $\psi(t) = e^{|t|}$. De esta manera, tenemos que, para todo polinomio $p(t)$,

$$||\psi(t) - p(t)||_\infty = \sup_{t \in [a,\infty)} |\psi(t) - p(t)| = \infty,$$

por resultados de cálculo básico. (Cabe la misma aclaración de que el supremo no es infinito, si no que el conjunto dado no es acotado superiormente).

3. Formas prácticas de emplear esta teoría

Más allá de la exposición que se brinda sobre los resultados de éste y otros teoremas en el ámbito de la teoría de aproximación, existen muchas formas de ver aplicaciones de las mismas. De hecho, hay toda una rama de la matemática que se encarga de ello. A modo de presentación, se intentará reflejar lo más posible esta idea.

3.1. Una idea económica

Problemas en los cuales se ven involucrados valores máximos y mínimos (pero no ambos) surgen de manera natural en el ámbito de la *teoría económica*. Supongamos que el valor ν de un objeto de venta (e.g., una reserva de petroleo) se correlaciona con una variable observable s (por ejemplo, los resultados en una perforación muestra). Sea $f(s/\nu)$ la densidad de s, para un ν dado. Supongamos que un comprador, pero no el vendedor, conoce ν ¿Puede un vendedor promedio cobrar al comprador su valor ν? Esto se reduce a resolver la ecuación

$$\nu = \int_S z(s)f(s/\nu)ds,$$

donde $z(s)$ es el precio de carga cuando surge la s resultante. Asumamos que s queda plasmado sobre un espacio métrico compacto S.

Si el vendedor ofrece al comprador un conjunto de funciones precio $\{z_1, \dots z_n\}$, y deja que el comprador elija la que más le gusta (i.e., la que minimiza el precio esperado), el vendedor observará que, en promedio

$$p(\nu) = \min_{1 \le i \le n} \int_S z_i(s)f(s/\nu)ds.$$

Esto requiere que $p(\nu)$ no sea mayor que el valor ν del comprador; de esta manera podrá estar dispuesto a participar bajo este esquema.

$$R = \left\{ \int z(s)f(s/\cdot)ds : z \in \mathcal{C}(S) \right\}$$

entonces el vendedor puede cobrar, con precisión, al valor del comprador cuando la identidad está en R. Obviamente, el vendedor puede estar arbitrariamente cerca si $\overline{R_m} = \mathcal{C}[0, \overline{\nu}]$, donde el valor cae en $[0, \overline{\nu}]$. Debemos notar que $\mathbf{1} \in R$[15] ya que $f(\cdot/\nu)$ es una densidad.

Ejemplo 3.1. Supongamos que una variable aleatoria s tiene una *función distribución acumulativa* s^ν con $s \in [0, 1]$. A un agente económico que conoce ν se le ofrece una lista de pagos $\{z_i\}$. Este agente escoge el precio con el menor valor esperado

$$p(\nu) = \min_{1 \le i \le n} \int_0^1 z_i(s)\nu s^{\nu-1}ds.$$

¿Es el conjunto de tales cargos denso en $\mathcal{C}[0, 1]$? Esto es, si el valor del agente de un objeto de venta es $\pi(\nu)$, ¿hay una lista $\{z_i(s)\}$ que aproxima los cargos al valor del agente?

La respuesta es *sí*. Consideremos

$$A = \left\{ f : f(\nu) = \int_0^1 z(s)\nu s^{\nu-1}ds,\ z \in \mathcal{C}([0, 1]) \right\}.$$

[15] Con $\mathbf{1}$ nos referimos a la función idénticamente igual a 1.

Notemos que A contiene a $\mathbf{1}$ (usando $z = 1$), $f(\nu) = \nu/(\nu + 1)$ (para $z(s) = s$), y $g(\nu) = \nu/(\nu + 2)$ (para $z(s) = s^2$). Como se verá más adelante, f y g cumplen las condiciones de (4.3) por lo que $\overline{A_m} = \mathcal{C}[0, 1]$.

3.2. Usando la aproximación para ejercitar la mente

Veamos que además de las demostraciones, que incluyen de por sí mucho esfuerzo mental, existen diversas formas de enfocar un estudio breve sobre estas teorías de aproximación. La mayoría de los libros de análisis matemático incluyen al menos uno de estos puntos. No he de seguir plan exacto en esta parte, sin embargo creo útil incluirla.

Ejercicio 3.2. Sea f continua en el intervalo $[0, 1]$ con la propiedad de que

$$\int_0^1 f(x)x^n dx = 0 \quad \forall\, n \in \mathbb{N},$$

probar que $f(x) = 0$ en $[0, 1]$

Probaremos que $\int_0^1 f^2(x)\,\mathrm{d}x = 0$ (de esta manera, puesto que $f^2(x) \geq 0$, deberá ser $f^2(x) = 0$ y por lo tanto $f(x) = 0$). Gracias al teorema de Weierstrass, sabemos que existe una sucesión P_n de polinomios que aproxima uniformemente a f. Luego

$$\int_0^1 f^2(x)dx = \int_0^1 f(x)\cdot f(x)dx = \int_0^1 f(x)\left(\lim_{n\to\infty} P_n(x)\right)dx.$$

La convergencia uniforme de P_n nos permite intercambiar el límite con la integral, obteniendo

$$\int_0^1 f^2(x)dx = \lim_{n\to\infty}\int_0^1 f(x)P_n(x)dx = \lim_{n\to\infty}\int_0^1 f(x)\left(\sum_{i=0}^{n} a_i x^i\right)dx.$$

Finalmente, por la linealidad de la integral,

$$\int_0^1 f^2(x)dx = \lim_{n\to\infty}\sum_{i=0}^{n} a_i \int_0^1 f(x)x^i dx = 0.$$

Ejercicio 3.3. Realizando una pequeña modificación al enunciado del ejercicio anterior, podemos obtener otro similar (y tal vez más bello aún).

Sea $f : [0, 1] \to \mathbb{R}$ una función continua tal que

$$\int_0^1 f(x)e^{nx}dx = 0$$

para todo n natural. Probar que entonces $f(x) = 0$ para $0 \leq x \leq 1$.

Para probarlo, basta comprobar que la familia de funciones e^{nx} forma un álgebra que separa puntos y no se anula. De esta manera, vía el teorema de Stone-Weierstrass, se puede repetir los pasos seguidos en el ejercicio anterior.

Ejercicio 3.4. Mostrar un ejemplo de una función continua $f : (-1, 1) \to \mathbb{R}$ que no puede ser aproximada uniformemente por polinomios.

Vamos a afirmar que nuestra función

$$f(x) = \frac{1}{x+1}$$

que es continua en el $(-1,1)$ no puede ser aproximada uniformemente por polinomios. Vamos a asumir que esto no es cierto. Sea entonces $\varepsilon = 1$ y por definición, existe un polinomio $p(x)$ tal que

$$|f(x) - p(x)| \leq 1$$

para todo $x \in (-1,1)$. Como $p(x)$ está acotado en $(-1,1)$, se sigue que existe una constante $M > 0$ tal que $|p(x)| \leq M$ para todo $x \in (-1,1)$. Pero entonces

$$|f(x)| = |p(x) + f(x) - p(x)| \leq |p(x)| + |f(x) - p(x)| \leq M + 1$$

para todo $x \in (-1,1)$. Pero esto contradice el hecho de que $f(x)$ es no acotada en $(-1,1)$.

El siguiente ejercicio establece una comparación entre distintas aproximaciones mediante polinomios: la de Weierstrass y de Taylor (de hecho, da un contraejemplo de por qué los polinomios de Taylor no son compatibles con el teorema de aproximación de Weierstrass).

Ejercicio 3.5. Sea f definida por

$$f(x) = \begin{cases} e^{-1/x^2}, & \text{si } x \neq 0 \\ 0, & \text{si } x = 0 \end{cases}.$$

Probar que los polinomios de Taylor de f alrededor de cero no convergen a f, excepto en cero.

La función $f(x)$ es infinitamente diferenciable en $x = 0$, y tiene todas sus derivadas iguales a cero. En consecuencia, la serie de Taylor de $f(x)$ alrededor de cero es idénticamente igual a cero. Sin embargo, $f(x)$ no es la función cero, y por lo tanto no es igual a su serie de Taylor alrededor del origen.

Ejercicio 3.6. Sea $P_0 = 0$, y definamos, para $n = 1, 2, \ldots,$

$$P_{n+1}(x) = P_n(x) + \frac{x^2 - P_n^2(x)}{2}.$$

Probar que

$$\lim_{n\to\infty} P_n(x) = |x|,$$

uniformemente en $[-1,1]$.

Al parecer, este es un caso en donde *vemos* quienes son los polinomios necesarios para aproximar dicha función continua. El resultado se sigue de la identidad

$$|x| - P_{n+1}(x) = [|x| - P_n(x)]\left[1 - \frac{|x| + P_n(x)}{2}\right]$$

para probar que $0 \leq P_n(x) \leq P_{n+1}(x) \leq |x|$ si $|x| \leq 1$ y que

$$|x| - P_n(x) \leq |x|\left(1 - \frac{|x|}{2}\right)^n \leq \frac{2}{n+1}$$

si $|x| \leq 1$. Haciendo $n \to \infty$ se sigue el resultado deseado.

Ejercicio 3.7. Sea f una función (al menos una vez) continuamente diferenciable en $[a,b]$. Probar que hay una sucesión de polinomios $\{P_n\}$ tal que $P_n(x) \to f(x)$ y $P_n'(x) \to f'(x)$ uniformemente en $[a,b]$.

Por ser f (al menos una vez) continuablemente diferenciable, f' es una función continua. Por el teorema de Weierstrass, existe una sucesión de polinomios $Q_n(x)$ tal que la aproxima uniformemente en $[a, b]$. Sea

$$P_n(x) = f(a) + \int_a^x Q_n(t)dt.$$

Como $P_n(a) = f(a)$ y $P_n'(x) = Q_n(x)$ se tiene que

$$\begin{aligned} f(x) - P_n(x) &= (f(x) - f(a)) - (P_n(x) - P_n(a)) \\ &= \int_a^x f'(t)dt - \int_a^x P_n'(t)dt \\ &= \int_a^x (f'(t) - Q_n(t))\, dt. \end{aligned}$$

Si $|f'(t) - Q_n(t)| < \varepsilon$ para todo t, entonces el valor absoluto de esta integral es, a lo sumo, $(b - a)\varepsilon$, que nos deduce el resultado deseado.

4. Generalizar y deducir

En la presente sección se expondrán dos cosas interesantes (que pueden advertirse fácilmente). Primero, como fue mencionado en el prólogo, mostraremos los resultados de Stone acerca de su generalización del teorema de Weierstrass, para lo cual usaremos lo visto en 1.5. A continuación, veremos cómo se puede obtener el teorema de Weierstrass como corolario directo de otros teoremas (no necesariamente más generales) a partir de ideas que se irán desarrollando en el proceso. Para una lectura directa, omitiremos algunos resultados en un primer plano, pero haremos referencia a ellos, pues nuestro interés general recae en el gran teorema central.

4.1. La generalización de Stone

Antes de ver de manera directa el resultado que deseamos probar, es necesario ir combinando los conocimientos que ya hemos discutido.

Definición 4.1. Si $\mathscr{A}$ es un álgebra de funciones sobre un conjunto X, con la propiedad de que cuando $f_n \in \mathscr{A}$ para n natural y $f_n \rightrightarrows f$,[16] implican $f \in \mathscr{A}$, decimos que $\mathscr{A}$ es *uniformemente cerrado*. Más aún, si $\mathfrak{B}$ es el conjunto de todas las funciones que son el límite de sucesiones uniformemente convergentes de $\mathscr{A}$, entonces $\mathfrak{B}$ se llama la *clausura uniforme* de $\mathscr{A}$.

Por ejemplo, el conjunto de todos los polinomios forman un álgebra, y el teorema de Weierstrass nos dice que el conjunto de las funciones continuas sobre $[a, b]$ es la clausura uniforme del conjunto de polinomios sobre $[a, b]$. En particular, si $\mathfrak{B}$ es la clausura uniforme de un álgebra $\mathscr{A}$,entonces $\mathfrak{B}$ mismo es un álgebra uniformemente cerrada (este hecho se sigue de las propiedades de la convergencia uniforme para sucesiones en un álgebra).

Daremos a continuación la demostración más general. Como en muchos casos, esta parte será más bien técnica, sólo debemos seguir los pasos. Los resultados parciales que van apareciendo pueden seguirse inmediatamente de la referencia indicada.

[16] Esta doble flecha la usaremos para denotar convergencia uniforme.

Teorema 4.2 (Stone-Weierstrass). *Sea $\mathscr{A}$ un álgebra de funciones continuas en un conjunto compacto de Hausdorff K. Si $\mathscr{A}$ cumple 1.14 sobre K, entonces la clausura uniforme $\mathfrak{B}$ de $\mathscr{A}$ consiste en todas las funciones reales continuas sobre K.*[17]

Demostración. La demostración se divide en cuatro pasos. Primero probaremos que si $f \in \mathfrak{B}$, entonces $|f| \in \mathfrak{B}$. Sea

$$a = \sup_{x \in K} |f(x)|$$

y $\varepsilon > 0$ dado. Por A.5 existen números reales $c_1, \ldots, c_n$ tal que

$$\left| \sum_{i=1}^{n} c_i y^i - |y| \right| < \varepsilon, \qquad -a \leq y \leq a. \tag{4.1}$$

Como $\mathfrak{B}$ es un álgebra, la función

$$g = \sum_{i=1}^{n} c_i f^i$$

es un miembro de $\mathfrak{B}$. Luego, por (4.1) y por definición de a, se tiene

$$|g(x) - |f(x)|| < \varepsilon, \qquad \forall x \in K.$$

Por ser $\mathfrak{B}$ uniformemente cerrado, se sigue que $|f| \in \mathfrak{B}$.

A continuación veremos que si $f \in \mathfrak{B}$ y $g \in \mathfrak{B}$, entonces $\text{máx}(f, g) \in \mathfrak{B}$ y $\text{mín}(f, g) \in \mathfrak{B}$.[18] Esto es claro si consideramos el resultado anterior y las identidades

$$\text{máx}(f, g) = \frac{f+g}{2} + \frac{|f-g|}{2},$$

$$\text{mín}(f, g) = \frac{f+g}{2} - \frac{|f-g|}{2}.$$

Claramente esto puede extenderse a cualquier conjunto finito de funciones, es decir, si $f_1, \ldots, f_n \in \mathfrak{B}$ entonces $\text{máx}(f_1, \ldots, f_n) \in \mathfrak{B}$ y $\text{mín}(f_1, \ldots, f_n) \in \mathfrak{B}$.

Como tercer paso probaremos lo siguiente. Dada una función real f, continua en K, un punto x en K y $\varepsilon > 0$, existe una función $g_x \in \mathfrak{B}$ tal que $g_x(x) = f(x)$ y

$$g_x(t) > f(t) - \varepsilon, \qquad \forall t \in K. \tag{4.2}$$

Como $\mathscr{A} \subset \mathfrak{B}$ y $\mathscr{A}$ satisface las hipótesis de A.6, también así lo hace $\mathfrak{B}$. Por lo tanto, para cada $y \in K$, podemos encontrar una función $h_y \in \mathfrak{B}$ tal que

$$h_y(x) = f(x), \qquad h_y(y) = f(y). \tag{4.3}$$

Por la continuidad de h_y existe un conjunto abierto J_y, que contiene a y, tal que

$$h_y(t) > f(t) - \varepsilon. \tag{4.4}$$

[17]Dicho de otro modo, $\mathscr{A}$ es denso en $\mathcal{C}(K)$.

[18]Por $\text{máx}(f, g)$ nos referimos a la función h definida como

$$h(x) = \begin{cases} f(x), & \text{si } f(x) \geq g(x), \\ g(x), & \text{si } f(x) < g(x) \end{cases}.$$

Una definición similar se sigue para $\text{mín}(f, g)$.

Por ser K compacto, existe un conjunto finito de puntos $y_1, \ldots, y_n$ tales que

$$K \subset \bigcup_{i=1}^{n} J_{y_i}. \tag{4.5}$$

Llamemos

$$g_x = \text{máx}(h_{y_1}, \ldots, h_{y_n}).$$

Por lo visto anteriormente, $g_x \in \mathfrak{B}$ y seguido de (4.3), (4.4) y (4.5) tenemos que g_x satisface (4.2).

Nuestro último gran paso será mostrar que, dada una función real f, continua en K, y un $\varepsilon > 0$, existe una función $h \in \mathfrak{B}$ tal que

$$|h(x) - f(x)| < \varepsilon, \qquad \forall x \in K. \tag{4.6}$$

Como $\mathfrak{B}$ es uniformemente cerrado, esta última afirmación es equivalente a la conclusión del teorema central. Para ver esto, consideremos la función g_x, para cada $x \in K$, que construimos en el paso anterior. Por la continuidad de g_x, existe un conjunto abierto V_x que contiene a x, tal que

$$g_x(t) < f(t) + \varepsilon \quad \forall t \in V_x. \tag{4.7}$$

Por ser K compacto, existe un conjunto finito de puntos $x_1, \ldots, x_m$ tal que

$$K \subset \bigcup_{i=1}^{m} V_{x_i}. \tag{4.8}$$

Sea

$$h = \text{mín}(g_{x_1}, \ldots, g_{x_n}). \tag{4.9}$$

En nuestro segundo paso, vimos que $h \in \mathfrak{B}$ y (4.2) implica que

$$h(t) > f(t) - \varepsilon \quad \forall t \in K, \tag{4.10}$$

más aún, por (4.7), (4.8) y (4.9) se sigue que

$$h(t) < f(t) + \varepsilon \quad \forall t \in K. \tag{4.11}$$

Finalmente, (4.6) resulta de (4.10) y (4.11). ■

Corolario 4.3. *Supongamos que A es un subespacio lineal de $\mathcal{C}[a,b]$ que contiene la función **1** y dos funciones f y g que satisfacen*

- *f es estrictamente creciente,*
- *$\frac{g(x)-g(y)}{f(x)-f(y)}$ es estrictamente creciente en $x \neq y$, para todo y.*

Entonces $\overline{A_m} = \mathcal{C}[a,b]$.

4.2. Nuestro tema central como corolario

En este punto, vamos a mostrar cómo las herramientas vistas anteriormente, junto con algunas ideas nuevas, permiten obtener nuestro teorema central como resultado directo de otros más generales. Existen diversas formas de obtener el teorema de Weierstrass como corolario de otros desarrollos;[19] nosotros aquí presentamos uno de ellos.

[19] Por ejemplo, véase *corolario 1.11* [3].

4.2.1. Un poco de Fourier

Consideremos las sumas parciales de la serie de Fourier como

$$S_n f(x) = \sum_{k=-n}^{n} c_k e^{ikx}$$

donde $c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt$ son los coeficientes de Fourier. Ahora podemos escribir

$$\begin{aligned} S_n f(x) &= \sum_{k=-n}^{n} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)e^{-iky}dy e^{ikx} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)D_n(x-y)dy \qquad \text{donde} \quad D_n(t) = \sum_{k=-n}^{n} e^{ikt}. \end{aligned}$$

A D_n se le llama *núcleo de Dirichlet.*

Definición 4.4. Dadas dos funciones integrables f y g, definimos su *convolución* h de la siguiente manera

$$h(x) = (f * g)(x) = \int f(y)g(x-y)dy.$$

Gracias a que la convolución es conmutativa (A.7), podemos decir que

$$S_n f(x) = f * D_n(x) = D_n * f(x)$$

y resolviendo la serie para D_n, obtenemos que

$$D_n(t) = \frac{\operatorname{sen}(n+\frac{1}{2})t}{\operatorname{sen}(\frac{t}{2})} \qquad \text{(A.8)}$$

y además la propiedad de que

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(t)dt = 1.$$

4.2.2. El teorema de Fejér

Lo que venimos desarrollando hasta ahora tiene una explicación. Nos gustaría mostrar que toda función continua puede ser aproximada por polinomios trigonométricos de manera uniforme en el intervalo $[-\pi,\pi]$. En un arrebato de audacia, uno podría pensar que las sumas parciales $S_n f$ de la serie de Fourier son un buen candidato para ello. Desafortunadamente hay funciones continuas para las que su suma parcial $S_n f(x)$ no converge a $f(x)$.[20]

Por lo tanto, en lugar de $S_n f$ vamos a considerar un método más factible, llamado *sumabilidad de Cesàro*, que consiste en tomar el límite de la media aritmética de las sumas parciales. Definamos

$$\sigma_N f(x) = \frac{1}{N+1}\sum_{n=0}^{N} S_n f(x)$$

[20] No mostraremos un contraejemplo de este hecho, pues se tratan de funciones muy complejas.

Si usamos la convolución, inmediatamente se sigue que

$$\begin{aligned}\sigma_N f(x) = K_N * f(x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(x-y)f(y)dy \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(t)f(x-t)dt\end{aligned}$$

donde K_N (llamado *núcleo de Fejér*) es

$$K_N(t) = \frac{1}{N+1}\sum_{n=0}^{N} D_n(t).$$

Es hora de probar el teorema, haciendo uso de (A.9) para las propiedades de K_N.

Teorema 4.5 (Fejér). *Sea f una función continua 2π-periódica. Entonces la media $\sigma_N f$ converge uniformemente a f en la norma uniforme. Es decir*

$$\max_{x\in\mathbb{R}} |\sigma_N f(x) - f(x)| \to 0, \quad \textit{cuando} \quad N \to \infty.$$

Demostración. Dado $\varepsilon > 0$ queremos ver que existe $M = M(\varepsilon)$ tal que, para $N \geq M$ se tiene

$$|\sigma_N f(x) - f(x)| < \varepsilon,$$

para todo x. Ahora vamos a escribir

$$\begin{aligned}\sigma_N f(x) - f(x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(t)f(x-t)dt - f(x) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(t)[f(x-t)-f(x)]dt, \qquad (4.12)\end{aligned}$$

donde (4.12) utiliza la propiedad *3* de K_N.

Por ser f continua en $[-\pi,\pi]$ es uniformemente continua allí. Como f es 2π-periódica, f es uniformemente continua en $\mathbb{R}$. Esto significa que existe un $\delta > 0$ tal que

$$|f(x-t) - f(x)| < \frac{\varepsilon}{4} \quad \text{para} \quad |t| \leq \delta \quad \text{y para todo} \quad x \in \mathbb{R}.$$

Dividamos la integral en dos partes

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(t)[f(x-t)-f(x)]dt = I_N(x) + II_N(x)$$

donde

$$I_N(x) = \frac{1}{2\pi}\int_{-\delta}^{\delta} K_N(t)[f(x-t)-f(x)]dt,$$

$$II_N(x) = \frac{1}{2\pi}\int_{[-\pi,\pi]\setminus[-\delta,\delta]} K_N(t)[f(x-t)-f(x)]dt.$$

Vamos a hacer una estimación para I_N que sirva para todo N. Haremos

$$\begin{aligned}|I_N(x)| &\leq \frac{1}{2\pi}\int_{-\delta}^{\delta} |K_N(t)|\,|[f(x-t)-f(x)]|dt \\ &\leq \frac{1}{2\pi}\int_{-\delta}^{\delta} |K_N(t)|\,\frac{\varepsilon}{4}dt \\ &\leq \frac{1}{2\pi}\int_{-\delta}^{\delta} K_N(t)dt\,\frac{\varepsilon}{4} = \frac{\varepsilon}{4}\end{aligned}$$

usando las propiedades de K_N. Como este resultado se sigue para todo N, podemos elegir N suficientemente grande para estimar el segundo término $II_N(x)$.

Ahora usamos la última propiedad de K_N y de esta manera poder estimar la integral en $x \in [-\pi, -\delta] \cup [\delta, \pi]$. Haremos uso de la cota

$$|f(x-t) - f(x)| \leq |f(x-t)| + |f(x)| \leq 2M,$$

siendo

$$M = \max_{x \in [-\pi,\pi] \setminus [-\delta,\delta]} \{|f(x)|\}.$$

Luego, resulta

$$\begin{aligned} |II_N(x)| &\leq 2M \frac{1}{2\pi} \int_{[-\pi,\pi]\setminus[-\delta,\delta]} \frac{1}{N+1} \left(\frac{2}{1-\cos(\delta)} \right) dt \\ &\leq \frac{1}{N+1} \left(\frac{4M}{1-\cos(\delta)} \right). \end{aligned}$$

Ahora bien, como $\frac{1}{N+1} \to 0$ cuando $N \to \infty$, podemos elegir N_0 tal que para $N \geq N_0$ la cantidad

$$\frac{1}{N+1} \left(\frac{4M}{1-\cos(\delta)} \right)$$

se haga menor que $\varepsilon/4$. Esto nos dice que para $N \geq N_0$ ambas cantidades, $|I_N(x)|$ y $|II_N(x)|$ son menores que $\varepsilon/4$ para todo x. Finalmente

$$\max_{x \in \mathbb{R}} |\sigma_N f(x) - f(x)| < \frac{\varepsilon}{2} \quad \text{para} \quad N \geq N_0. \qquad \blacksquare$$

Ahora bien, ¿qué hacemos con todo esto? Como dijimos, queremos mostrar que el teorema de Weierstrass se puede obtener de los resultados mencionados.

Corolario 4.6 (Teorema de aproximación de Weierstrass). *Sea f continua en un intervalo $[a,b]$. Entonces f puede ser aproximada uniformemente por polinomios en $[a,b]$.*

Demostración. Claramente, lo que deseamos probar es que, dado $\varepsilon > 0$ existe un polinomio P (que sólo depende de ε) tal que

$$\max_{x \in [a,b]} |f(x) - P(x)| < \varepsilon.$$

Consideremos el caso en que $[a,b] = [-\frac{\pi}{2}, \frac{\pi}{2}]$. Ahora vamos a extender la función f a una función continua F en $[-\pi, \pi]$ tal que $F(x) = f(x)$ en $[-\frac{\pi}{2}, \frac{\pi}{2}]$ y $F(-\pi) = F(\pi) = 0$. Entonces podemos extender F a una función continua 2π-periódica en todo $\mathbb{R}$.

Dado $\varepsilon > 0$, por el teorema de Fejér podemos encontrar un polinomio trigonométrico

$$T(x) = a_0 + \sum_{k=1}^{N} [a_k \cos(kx) + b_k \operatorname{sen}(kx)]$$

de manera que

$$\max_{x \in \mathbb{R}} |F(x) - T(x)| < \frac{\varepsilon}{2}.$$

Más aún, la serie de Taylor para coseno y seno convergen uniformemente en todo intervalo compacto. Entonces podemos encontrar un polinomio P tal que

$$\max_{|x| \leq \pi/2} |T(x) - P(x)| < \frac{\varepsilon}{2}.$$

Combinando las dos últimas ideas, junto con el hecho de que $f = F$ en $[-\frac{\pi}{2}, \frac{\pi}{2}]$, resulta

$$\max_{|x|\leq\pi/2} |f(x) - P(x)| = \max_{|x|\leq\pi/2} |F(x) - P(x)| < \varepsilon. \qquad \blacksquare$$

A. Resultados necesarios

Teorema A.1 (**Heine**). *Sea $f : S \to T$ una función definida entre dos espacios métricos (S, d_S) y (T, d_T). Sea A un subconjunto compacto*[21] *de S y supongamos que f es continua en A. Entonces f es uniformemente continua en A.*

Demostración. Dado $\varepsilon > 0$, a cada punto $a \in A$ se le puede asociar una bola $B_S(a; r)$ centrada en a y de radio r en el espacio S, con r dependiendo de a, tal que

$$d_T(f(x), f(a)) < \frac{\varepsilon}{2} \qquad \text{siempre que } x \in B_S(a; r) \cap A.$$

Consideremos la colección de las bolas $B_S(a; r/2)$ de radio $r/2$. Recubren a A y, como A es compacto, basta un número finito de ellas para recubrir a A, es decir

$$A \subseteq \bigcup_{k=1}^{m} B_S\left(a_k; \frac{r_k}{2}\right).$$

En cualquiera de las bolas de doble radio, $B_S(a_k, r_k)$ se tiene

$$d_T(f(x), f(a)) < \frac{\varepsilon}{2} \qquad \text{siempre que } x \in B_S(a_k; r_k) \cap A.$$

Sea δ el menor de los números $r_1/2, \ldots, r_m/2$. Probaremos que este δ satisface la definición de continuidad uniforme.

En efecto, consideremos dos puntos de A, por ejemplo x y p, con $d_S(x, p) < \delta$. En virtud de la anterior discusión existirá una bola $B_S(a_k; r_k/2)$ que contenga a x, luego

$$d_T(f(x), f(a)) < \frac{\varepsilon}{2}.$$

Por la desigualdad triangular tenemos que

$$d_S(p, a_k) \leq d_S(p, x) + d_S(x, a_k) < \delta + \frac{r_k}{2} \leq \frac{r_k}{2} + \frac{r_k}{2} = r_k.$$

Por lo tanto, $p \in B_S(a_k; r_k) \cap S$, entonces tenemos también que

$$d_T(f(p), f(a_k)) < \frac{\varepsilon}{2}.$$

Finalmente

$$d_T(f(x), f(p)) \leq d_T(f(x), f(a_k)) + d_T(f(ak), f(p)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \qquad \blacksquare$$

Lema A.2. *Sea $f : S \to T$ una función de un espacio métrico (S, d_S) en otro (T, d_T). Entonces f es continua en S si, y sólo si, para cada conjunto abierto Y de T, la antiimagen $f^{-1}(Y)$ es abierta en S.*

Teorema A.3. *Sea $f : S \to T$ una función de un espacio métrico (S, d_S) en otro (T, d_T). Si f es continua en un subconjunto compacto X de S, entonces la imagen $f(X)$ es un subconjunto compacto de T.*

Demostración. Sea F una familia de conjuntos A que es un cubrimiento por abiertos de $f(X)$, es decir $f(X) \subseteq \bigcup_{A \in F} A$. Como f es continua sobre el subespacio métrico (X, d_S) podemos aplicar el lema anterior para concluir que cada uno de los conjuntos

[21] En el sentido de (B.1)

$f^{-1}(A)$ es abierto en (X, d_S). Los conjuntos $f^{-1}(A)$ forman un cubrimiento abierto de X y, como X es compacto, podemos extraer un subcubrimiento finito de X; sea entonces

$$X \subseteq f^{-1}(A_1) \cup \cdots \cup f^{-1}(A_p).$$

Entonces aplicando f a ambos miembros

$$f(X) \subseteq f[f^{-1}(A_1) \cup \cdots \cup f^{-1}(A_p)] = f[f^{-1}(A_1)] \cup \cdots \cup f[f^{-1}(A_p)] \subseteq A_1 \cup \cdots \cup A_p$$

luego $f(X)$ es compacto. ■

Teorema A.4. *Sea $\mathfrak{B}$ la clausura uniforme de un álgebra $\mathscr{A}$ de funciones acotadas. Entonces $\mathfrak{B}$ es un álgebra uniformemente cerrado.*

Demostración. Si $f \in \mathfrak{B}$ y $g \in \mathfrak{B}$ existen sucesiones uniformemente convergentes f_n, g_n tales que $f_n \rightrightarrows f$, $g_n \rightrightarrows g$ y $f_n \in \mathscr{A}$, $g_n \in \mathscr{A}$. Como estamos trabajando con funciones acotadas, es fácil ver que

$$f_n + g_n \rightrightarrows f + g, \quad f_n g_n \rightrightarrows fg, \quad cf_n \rightrightarrows cf,$$

donde c es una constante real. Por lo tanto, $f + g \in \mathfrak{B}$, $fg \in \mathfrak{B}$ y $cf \in \mathfrak{B}$. Luego, $\mathfrak{B}$ es un álgebra uniformemente cerrado ■

Corolario A.5. *Para cada intervalo $[-a, a]$ existe una sucesión de polinomios reales P_n tal que $P_n(0) = 0$ y tal que*

$$\lim_{n\to\infty} P_n(x) = |x|$$

uniformemente en $[-a, a]$.

Demostración. Por el teorema de Weierstrass, existe una sucesión $\{P_n^*\}$ de polinomios reales que convergen a $|x|$ uniformemente en $[-a, a]$. En particular, $P_n^*(0) \overset{n\to\infty}{\rightarrow} 0$. Los polinomios

$$P_n(x) = P_n^*(x) - P_n^*(0) \qquad n \in \mathbb{N}$$

tienen la propiedad requerida. ■

Teorema A.6. *Supongamos que $\mathscr{A}$ es un álgebra de funciones en un conjunto X que cumple 1.14. Supongamos también que x_1, x_2 son puntos distintos de X y c_1, c_2 son dos constantes reales. Entonces $\mathscr{A}$ contiene una función f tal que*

$$f(x_1) = c_1, \qquad f(x_2) = c_2.$$

Demostración. Nuestras suposiciones muestran que $\mathscr{A}$ tiene funciones g, h y k tales que

$$g(x_1) \neq g(x_2), \quad h(x_1) \neq 0, \quad k(x_2) \neq 0.$$

Hagamos

$$u = gk - g(x_1)k, \qquad v = gh - g(x_2)h.$$

Entonces $u \in \mathscr{A}$, $v \in \mathscr{A}$, $u(x_1) = v(x_2) = 0$, $u(x_2) \neq 0$ y $v(x_1) \neq 0$. Luego la f que buscamos es

$$f = \frac{c_1 v}{v(x_1)} + \frac{c_2 u}{u(x_2)}.$$ ■

Propiedad A.7. *La convolución de funciones f y g ambas 2π-periódicas es conmutativa.*

Demostración. Para ver esto, primero notemos que para una función integrable 2π-periódica tenemos

$$\int_{-\pi}^{\pi} F(t)dt = \int_{a-\pi}^{a+\pi} F(t)dt$$

para todo a. La conmutatividad se sigue de la definición de $f * g$ con el cambio de variables $t = x - y$ (con $dt = -dy$) obteniendo

$$\begin{aligned} 2\pi f * g(x) &= \int_{-\pi}^{\pi} f(y)g(x-y)dy = \int_{x+\pi}^{x-\pi} f(x-t)g(t)(-1)dt \\ &= \int_{x-\pi}^{x+\pi} f(x-t)g(t)dt = \int_{-\pi}^{\pi} g(t)f(x-t)dt = 2\pi g * f(x) \end{aligned}$$

donde el último paso utiliza el hecho de que las funciones son 2π-periódicas. ■

Propiedad A.8.

$$D_n(t) = \frac{\sin(n+\frac{1}{2})t}{\sin(\frac{t}{2})}.$$

Demostración. Considerando que

$$\sum_{k=0}^{n} e^{ikt} = \frac{e^{i(n+1)t} - 1}{e^{it} - 1},$$

tenemos que,

$$\sum_{k=-n}^{-1} e^{ikt} = \sum_{k=1}^{n} e^{-ikt} = \frac{e^{-i(n+1)t} - 1}{e^{-it} - 1} - 1,$$

donde la segunda suma se puede simplicar a

$$\frac{e^{-int} - 1}{1 - e^{it}}.$$

Entonces

$$D_n(t) = \frac{e^{i(n+1)t} - e^{-int}}{e^{it} - 1}.$$

Multiplicando numerador y denominador por $e^{-it/2}$ tenemos

$$D_n(t) = \frac{e^{i(n+1/2)t} - e^{-i(n+1/2)t}}{e^{it/2} - e^{-it/2}},$$

que es el resultado deseado. ■

Lema A.9. *El núcleo de Fejér cumple las siguientes propiedades*

1. *Para* $[-\pi, \pi]$ *se tiene*

$$\begin{aligned} K_N(x) &= \frac{1}{N+1}\frac{1-\cos(N+1)x}{1-\cos(x)} \\ &= \frac{1}{N+1}\left(\frac{\operatorname{sen}\left(\frac{N+1}{2}\right)x}{\operatorname{sen}\left(\frac{x}{2}\right)}\right)^2 \end{aligned}$$

   si x *no es un entero múltiplo de* 2π. *Y* $K_N(0) = N + 1$.

2. $K_N(x) \geq 0$ *para todo* $x \geq 0$.

3.

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(t)dt = 1.$$

4.

$$K_N(x) \leq \frac{1}{N+1}\left(\frac{2}{1-\cos(\delta)}\right) \quad \textit{para} \quad 0 < \delta \leq x \leq \pi.$$

Demostración. Escribiendo nuevamente el núcleo de Dirichlet

$$D_n(x) = \frac{\operatorname{sen}(n+\frac{1}{2})x}{\operatorname{sen}(\frac{x}{2})} = \frac{\operatorname{sen}(\frac{x}{2})\operatorname{sen}(n+\frac{1}{2})x}{\operatorname{sen}^2(\frac{x}{2})}.$$

Observando que $2\operatorname{sen}(a)\operatorname{sen}(b) = \cos(a-b) - cos(a+b)$ y aplicando esto con $a = (n+1/2)x$, $b = x/2$ se tiene

$$D_n(x) = \frac{\cos(nx) - \cos(n+1)x}{2\operatorname{sen}^2(\frac{x}{2})}.$$

Entonces

$$\begin{aligned} K_N(x) &= \frac{1}{N+1}\sum_{n=0}^{N} D_n(x) \\ &= \frac{1}{N+1}\sum_{n=0}^{N} \frac{\cos(nx) - \cos(n+1)x}{2\operatorname{sen}^2(\frac{x}{2})} \\ &= \frac{1}{N+1}\left(\frac{1-\cos(N+1)x}{2\operatorname{sen}^2(\frac{x}{2})}\right) \end{aligned}$$

Ahora usamos el hecho de que $\cos(2a) = \cos^2(a) - \operatorname{sen}^2(a) = 1 - 2\operatorname{sen}^2(a)$, por lo que $2\operatorname{sen}^2(a) = 1 - \cos(2a)$. Si usamos esto para $a = \frac{x}{2}$ obtenemos la primera fórmula para $K_N(x)$, y si lo usamos con $a = (N+1)\frac{x}{2}$ obtenemos la segunda fórmula.

Las propiedades *2* y *4* se siguen inmediatamente de dichas fórmulas.

La propiedad *3* se sigue de $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(t)dt = 1$. ∎

B. Definiciones equivalentes

Definición B.1. Sea X un espacio métrico. Diremos que la familia $\{A_i\}$ es un *cubrimiento* por abiertos de X si cada A_i es un conjunto abierto, y si

$$X \subseteq \bigcup_{i\in\mathbb{N}} A_i.$$

Un espacio métrico se llama *compacto* si, para toda familia $\{A_i\}$ que forme un cubrimiento por abiertos de X, existe un subcubrimiento finito de $\{A_i\}$ que también recubre a X. Es decir existe un conjunto finito de índices tal que

$$X \subseteq \bigcup_{i=1}^{n} A_i.$$

Definición B.2. Sean (X, d_X) e (Y, d_Y) espacios métricos y sea $f : X \to Y$ una función de X en Y. Entonces f es *continua* en un punto p de X si para cada $\varepsilon > 0$ existe un $\delta > 0$ tal que

$$d_Y(f(x), f(p)) < \varepsilon \qquad \text{siempre que} \qquad d_X(x, p) < \delta.$$

Esta definición puede expresarse en términos de *entornos*: una función f es continua en p si, y sólo si, para cada $\varepsilon > 0$, existe un $\delta > 0$ tal que

$$f(N_p(\delta)) \subseteq N_{f(p)}(\varepsilon).$$

Referencias

[1] Tom M Apostol. *Mathematical analysis.* Addison Wesley Publishing Company, 1974.

[2] John William Scott Cassels. *An introduction to Diophantine approximation*, volume 1957. University Press Cambridge, 1957.

[3] Javier Duoandikoetxea. *Fourier analysis*, volume 29. American mathematical society, 2000.

[4] Willliam Feller. *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons, 2008.

[5] LaRita Barnwell Hipp. *The Weierstrass Approximation Theorem.* PhD thesis, University of South Carolina, 2013.

[6] Ricardo Noriega. Cálculo diferencial e integral. *Editorial Docencia, BS AS*, 1991.

[7] Walter Rudin. *Principles of mathematical analysis*, volume 3. McGraw-Hill New York, 1964.

[8] Eugene Schenkman. The weierstrass approximation theorem. *American Mathematical Monthly*, pages 65–66, 1972.

[9] Marshall H Stone. Applications of the theory of boolean rings to general topology. *Transactions of the American Mathematical Society*, 41(3):375–481, 1937.

[10] Marshall H Stone. The generalized weierstrass approximation theorem. *Mathematics Magazine*, 21(5):237–254, 1948.

[11] Karl Weierstrass. Über die analytische darstellbarkeit sogenannter willkürlicher functionen einer reellen veränderlichen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, 2:633–639, 1885.