Métodos Analíticos en Estadística
Tema 6: Aproximación por mínimos cuadrados

1 Aproximación de funciones

Dada una función $f \in C[a,b]$, el objetivo de la aproximación polinómica de funciones es encontrar un polinomio p de *grado n prefijado* de tal manera que

$$\| f - p \| \leq \| f - q \|, \qquad \forall q \in \Pi_n. \tag{1.1}$$

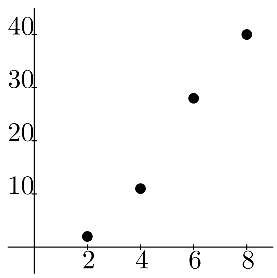
A dicho polinomio p se le llama *mejor aproximación de f en Π_n respecto a la norma* $\| \cdot \|$.

Esto se aplica principalmente para resolver dos tipos de problemas. El primer problema surge cuando tenemos una función f explícitamente pero por su complejidad se desea un tipo de función más simple que dé aproximadamente los mismos resultados. El segundo tipo de problemas aparece en muchas aplicaciones experimentales cuando obtenemos un conjunto de datos dados y buscamos una función que se ajuste lo mejor posible a dichos datos.

Por ejemplo, en un experimento obtenemos los siguientes datos de un proceso

i	t_i	y_i
1	2	2
2	4	11
3	6	28
4	8	40

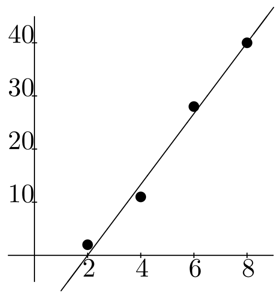
En principio podríamos pensar en utilizar la interpolación polinómica, obteniendo un polinomio de grado 3 que pasaría exactamente por dichos puntos. Pero si los representamos obtenemos



lo que nos indica que es razonable suponer que los datos realmente representan una recta y que aparecen unas pequeñas oscilaciones debidas a los errores intrínsecos a todas las mediciones experimentales. El intentar interpolar lo que estaríamos haciendo sería introducir unas oscilaciones en la función observada que en realidad no tiene.

Lo que hace la teoría de aproximación es buscar la mejor recta ($p \in \Pi_1$) que refleje el comportamiento de la función aproximada, lo que matemáticamente quiere decir que la distancia entre la recta buscada y los datos sea la menor posible, o, lo que es lo mismo, su norma. Esto es lo que significa la ecuación (1.1).

La aproximación de mínimos cuadrados que veremos nos dirá que esta recta es $y = 6.55x - 12.5$, representándose como



lo que nos dará una visión probablemente más realista que si usamos interpolación. Además, el polinomio aproximante es más simple y menos costoso computacionalmente que el interpolatorio.

A partir de esta idea se construye la teoría general de la aproximación numérica, que generaliza al caso de que queramos aproximar con un polinomio de grado prefijado n.

Dependiendo de la norma que usemos para medir las distancias en el espacio de las funciones continuas $C[a,b]$ se tendrán varios tipos de aproximación:

- *Aproximación polinómica uniforme:* la norma es la norma infinito o del máximo:

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

- *Aproximación por mínimos cuadrados:* la norma elegida es la norma 2, respecto a una *función peso* $w(x) \geq 0$

$$\| f \|_{2,w} = \left(\int_a^b f^2(x) w(x)\, dx \right)^{1/2},$$

  que es la más utilizada en la práctica. Tiene la gran ventaja de que es una norma que proviene de un producto escalar:

$$\| f \|_{2,w} = \sqrt{\langle f, f \rangle_w}, \qquad \langle f, g \rangle_w = \int_a^b f(x) g(x) w(x)\, dx.$$

Observación.- El problema de aproximación de hecho es aún más general que esto: si tenemos una función real de variable real f y una clase de funciones aproximantes Φ, el problema de aproximación consiste en buscar un aproximante $\hat{\phi} \in \Phi$ de tal manera que

$$\| f - \hat{\phi} \| \leq \| f - \phi \|, \qquad \forall \phi \in \Phi.$$

El que se elija de forma natural como clase de aproximantes Φ el conjunto de los polinomios de coeficientes reales Π viene motivado por el conocido **Teorema de Weierstrass**: *toda función continua $f(x)$ en un intervalo cerrado $[a,b]$ puede aproximarse uniformemente por polinomios*, es decir, para cualquier $\varepsilon > 0$ existe un polinomio $p(x)$ para el que

$$|f(x) - p(x)| \leq \varepsilon, \qquad a \leq x \leq b.$$

Hay otras maneras de elegir el conjunto de aproximantes Φ: el conjunto de funciones spline, el conjunto de polinomios trigonométricos y el conjunto de funciones racionales (cocientes de polinomios) son los más habituales, pero nosotros sólo nos vamos a centrar en la aproximación polinómica.

2 Aproximación por mínimos cuadrados

Consideremos el espacio vectorial real de las funciones reales continuas en un intervalo cerrado $f \in C[a,b]$. En dicho espacio definimos el *producto interior*:

$$\langle f, g \rangle_w := \int_a^b f(x) g(x) w(x) dx, \qquad f, g \in C[a,b].$$

fijada una función $w(x)$ llamada *función peso* que está definida en $[a,b]$ y que verifica:

- $w(x) \geq 0$, $\forall x \in [a,b]$.
- $\displaystyle\int_a^b w(x)\, dx < \infty$, o sea, es integrable en $[a,b]$.

Este producto es un producto interior en $C[a,b]$ pues verifica:

1. $\langle \alpha f, g \rangle_w = \alpha \langle f, g \rangle_w$ para cualquier escalar α.

2. $\langle f, g\rangle_w = \langle g, f\rangle_w$.
3. $\langle f_1 + f_2, g\rangle_w = \langle f_1, g\rangle_w + \langle f_2, g\rangle_w$.
4. $\langle f, f\rangle_w \geq 0$ para toda $f \in C[a,b]$ y $\langle f, f\rangle_w = 0$ si y sólo si $f(x) = 0, \forall x \in [a,b]$.

Además, podemos definir a partir de este producto la *norma euclídea o norma asociada al producto interior*:

$$\| f \|_{2,w} = \sqrt{\langle f, f\rangle_w} = \sqrt{\int_a^b (f(x))^2 w(x) dx}$$

que es una norma en $C[a,b]$.

Con esto, el *problema de aproximación polinómica por mínimos cuadrados* se plantea de la siguiente manera:

(P. A.): *Dada una función $f \in C[a,b]$, encontrar un polinomio $p \in \Pi_n$, de grado n prefijado, de tal manera que*

$$\| f - p \|_{2,w} \leq \| f - q \|_{2,w}, \qquad \forall q \in \Pi_n \tag{2.1}$$

o, lo que es lo mismo, encontrar la mejor aproximación a f en Π_n.

El conocido Teorema de Weierstrass es el primero que nos dice que plantearse este problema **(P. A.)** tiene sentido:

Teorema 2.1 *(Teorema de aproximación de Weierstrass)*
Sea $f \in C[a,b]$ y cualquier $\epsilon > 0$. Existe un polinomio $p(x)$ de tal forma que

$$|f(x) - p(x)| < \epsilon, \quad a \leq x \leq b.$$

Este teorema nos dice que tiene sentido aproximar funciones continuas por polinomios, pero no nos resuelve el problema de aproximación **(P. A.)**.

2.1 Existencia y unicidad de la mejor aproximación

Vamos a ver si el problema **(P. A.)** este problema tiene solución.

Dado n fijo, Π_n es un espacio vectorial real de dimensión $n+1$. Sea

$$\mathcal{B} = \{p_0(x), p_1(x), \ldots, p_n(x)\} \tag{2.2}$$

una base cualquiera de Π_n ($n+1$ polinomios de Π_n linealmente independientes). Si existe, la mejor aproximación $p \in \Pi_n$ ha de representarse de forma unívoca como

$$p(x) = \sum_{i=0}^{n} c_i p_i(x). \tag{2.3}$$

Así que encontrar $p \in \Pi_n$ que verifique (2.1) es lo mismo que encontrar de forma única los $n+1$ coeficientes $\{c_i\}_{i=0}^n$ que hacen que $\| f - p \|_{2,w}$ sea mínima. Veamos cómo se escribe esta norma en función de los $\{c_i\}$:

$$\begin{aligned}
\| f - p \|_{2,w}^2 &= \langle f-p, f-p\rangle_w = \langle f, f\rangle_w - 2\langle f, p\rangle_w + \langle p, p\rangle_w \\
&= \int_a^b (f(x))^2 w(x)\, dx - 2\int_a^b f(x) \left(\sum_{i=0}^n c_i p_i(x)\right) w(x)\, dx + \int_a^b \left(\sum_{i=0}^n c_i p_i(x)\right)^2 w(x)\, dx \\
&=: G(c_0, c_1, \ldots, c_n)
\end{aligned}$$

Vemos que la norma es una función G real que depende de $n+1$ variables $\{c_i\}$. Para que esta G alcance un mínimo en un punto $(c_0, c_1, \ldots, c_n)$ tiene que cumplirse

(a) $\dfrac{\partial G}{\partial c_j}(c_0, c_1, \ldots, c_n) = 0, \ \forall j = 0, 1, \ldots, n.$

(b) El Hessiano

$$H = \left(\frac{\partial^2 G}{\partial c_k \partial c_j}(c_0, c_1, \ldots, c_n) \right)_{k,j=0}^{n}$$

ha de verificar que $\det H > 0$ y que todos sus menores principales también tienen que ser > 0.

Calculemos estas parciales para $j = 0, 1, \ldots, n$:

$$\frac{\partial G}{\partial c_j} = -2 \int_a^b f(x) p_j(x) w(x)\, dx + 2 \int_a^b \left(\sum_{i=0}^{n} c_i p_i(x) \right) p_j(x) w(x)\, dx = 0, \tag{2.4}$$

de donde obtenemos fácilmente que se ha de verificar

$$\sum_{i=0}^{n} c_i \langle p_i, p_j \rangle_w = \langle f, p_j \rangle_w, \qquad \forall 0 \le j \le n. \tag{2.5}$$

De forma clara vemos que esto es un sistema lineal en las $\{c_i\}$, que se expresa en forma matricial como

$$AC = B \tag{2.6}$$

$$A = \begin{pmatrix} \langle p_0, p_0 \rangle_w & \langle p_1, p_0 \rangle_w & \cdots & \langle p_n, p_0 \rangle_w \\ \langle p_0, p_1 \rangle_w & \langle p_1, p_1 \rangle_w & \cdots & \langle p_n, p_1 \rangle_w \\ \vdots & \vdots & \ddots & \vdots \\ \langle p_0, p_n \rangle_w & \langle p_1, p_n \rangle_w & \cdots & \langle p_n, p_n \rangle_w \end{pmatrix}, \quad C = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad B = \begin{pmatrix} \langle f, p_0 \rangle_w \\ \langle f, p_1 \rangle_w \\ \vdots \\ \langle f, p_n \rangle_w \end{pmatrix}.$$

Por tanto, la existencia de los $\{c_i\}$ pasa porque esta matriz de coeficientes A sea regular ($\det A \neq 0$). Además, estudiando las parciales de segundo orden, derivando respecto c_k, $k = 0, 1, \ldots, n$ en (2.4) se tiene

$$\frac{\partial^2 G}{\partial c_k \partial c_j} = 2 \int_a^b p_k(x) p_j(x) w(x)\, dx = 2 \langle p_k, p_j \rangle_w,$$

por lo que

$$H = 2A.$$

Entonces, existirá un mínimo de G para un punto $(c_0, c_1, \ldots, c_n)$ sii

(i) $\det A > 0$.

(ii) Todos los menores principales de A son > 0.

El álgebra lineal nos dice que si una matriz es simétrica y definida positiva ($x^T A x \ge 0$ y $x^T A x = 0$ sii $x = 0$), entonces verifica (i) y (ii). Como el producto interior $\langle \cdot \rangle_w$ es conmutativo, es claro que A es simétrica. Veamos si es definida positiva.

Sea cualquier $x = (x_0, x_1, \ldots, x_n)^T$. Haciendo una cuenta elemental tenemos

$$x A^T x = \sum_{k=0}^{n} \sum_{j=0}^{n} x_k x_j \langle p_k, p_j \rangle_w = \langle \sum_{k=0}^{n} x_k p_k, \sum_{j=0}^{n} x_j p_j \rangle_w = \| \sum_{k=0}^{n} x_k p_k \|_{2,w}^2 \ge 0.$$

Por otra parte

$$x A^T x = 0 \iff \| \sum_{k=0}^{n} x_k p_k \|_{2,w}^2 = 0 \iff \sum_{k=0}^{n} x_k p_k = 0.$$

Como $\mathcal{B}$ es una base de Π_n, esto implica directamente que $x_k = 0$, $\forall k = 0, 1, \ldots, n$, es decir $x = 0$. El recíproco es trivial. Por tanto, A es definida positiva.

En conclusión, lo anterior nos dice que la solución del sistema lineal (2.6) existe y es única y además, en dicho punto, G alcanza el mínimo. Es decir, el problema **(P. A.)** tiene solución única $p \in \Pi_n$.

Para calcular esta mejor aproximación por mínimos cuadrados, lo que tenemos que hacer entonces es calcular una base $\mathcal{B}$ de Π_n, calcular los elementos de la matriz A y el vector B y resolver el sistema lineal (2.6).

Abordar este cálculo directamente con cualquier base de Π_n es casi imposible en la práctica, pues el coste de calcular todos los productos interiores $\langle p_j, p_k\rangle_w, \langle f, p_j\rangle_w$ es muy costoso. La idea que surge naturalmente es elegir la base $\mathcal{B}$ de tal forma que este cálculo sea sencillo. De hecho, se van a construir un tipo de bases de Π_n de forma que los productos $\langle p_j, p_k\rangle_w$ son conocidos y no hará falta calcularlos: *las bases de polinomios ortogonales.*

2.2 Polinomios ortogonales

Se dice que dos funciones f y g de $C[a, b]$ son *ortogonales respecto a la función peso* w sii

$$\langle f, g\rangle_w = 0.$$

A partir de esto, se dice que una base $\{p_0, p_1, \dots, p_n\}$ de Π_n es ortogonal sii $\langle p_i, p_j\rangle_w = 0,\ \forall i \neq j$. Se dice que es *base ortonormal de* Π_n sii

$$\langle p_i, p_j\rangle_w = \delta_{ij} = \begin{cases} 0, & \text{si } i \neq j \\ 1, & \text{si } i = j \end{cases} \tag{2.7}$$

Usando el proceso de ortonormalización de Gram-Schmidt[1] sobre el conjunto de polinomios Π definidos en $[a, b]$ obtenemos:

Teorema 2.2 *Existe una sucesión de polinomios* $\{p_n(x)\,/\,n \geq 0\}$ *con grado*$(p_n) = n$*, para todo* n *y*

$$\langle p_n, p_m\rangle_w = 0 \text{ para todo } n \neq m, \quad n, m \geq 0,$$

Es más, podemos construir esta sucesión con las siguientes propiedades adicionales:

(1) $\langle p_n, p_n\rangle_w = 1,\ \forall n$

(2) El coeficiente director de p_n *es positivo.*

Con estas propiedades adicionales, la sucesión de polinomios $\{p_n(x)\,/\,n \geq 0\}$ *es única.*

Esta sucesión de polinomios se llama *sucesión de polinomios ortogonales respecto a* $w(x)$ *en* $[a, b]$.

Volviendo al problema de aproximación por mínimos cuadrados, para minimizar la norma $\| f - p \|_{2,w}$ entre todos los polinomios $p \in \Pi_n$, elegimos como la base $\mathcal{B}$ en (2.2) una base ortonormal $\{p_0, p_1, \dots, p_n\}$ de Π_n verificando las propiedades del Teorema 2.2 anterior. Así el polinomio de aproximación buscado se expresa como vimos como (2.3), donde los coeficientes $\{c_i\}_{i=0}^n$ son la solución del sistema lineal (2.6).

La gran ventaja de este tipo de base ortonormal es que en este caso dicho sistema (2.6) resulta

$$A = I, \qquad C = B,$$

de donde la solución única es

$$c_i = \langle f, p_i\rangle_w, \qquad i = 0, 1, \dots, n. \tag{2.8}$$

Por tanto, la mejor aproximación a f por mínimos cuadrados en Π_n viene dada por

$$p(x) = \sum_{i=0}^{n} \langle f, p_i\rangle_w p_i(x). \tag{2.9}$$

[1]Ver Anexo

Además, volviendo a la construcción de $G(c_0, c_1, \ldots, c_n)$ el valor de la distancia de f a dicha mejor aproximación (o, lo que es lo mismo, la distancia de f a Π_n), llamada *desviación mínima de la mejor aproximación*, resulta

$$\begin{aligned} \| f - p \|_{2,w}^2 &= \| f \|_{2,w}^2 - 2\langle f, \sum_{i=0}^{n} \langle f, p_i \rangle_w p_i \rangle_w + \langle \sum_{i=0}^{n} \langle f, p_i \rangle_w p_i, \sum_{j=0}^{n} \langle f, p_j \rangle_w p_j \rangle_w \\ &= \| f \|_{2,w}^2 - \sum_{i=0}^{n} (\langle f, p_i \rangle_w)^2 = \| f \|_{2,w}^2 - \| p \|_{2,w}^2, \end{aligned} \tag{2.10}$$

o, de otro modo

$$\| f - p \|_{2,w} = \left[\| f \|_{2,w}^2 - \| p \|_{2,w}^2 \right]^{1/2}. \tag{2.11}$$

Observaciones:

1. Para simplificar los cálculos a menudo no se exige que la base sea ortonormal, sino que sólo se pide que sea ortogonal, es decir $\langle p_i, p_j \rangle_w = 0$ si $i \neq j$, pero $\langle p_i, p_i \rangle_w \neq 1$. En estos casos, la matriz de coeficientes del sistema de la mejor aproximación (2.6) resulta

$$A = \text{diag}(\langle p_0, p_0 \rangle_w, \langle p_1, p_1 \rangle_w, \ldots, \langle p_n, p_n \rangle_w),$$

   por lo que la solución es

$$c_i = \frac{\langle f, p_i \rangle_w}{\langle p_i, p_i \rangle_w}, \qquad i = 0, 1, \ldots, n. \tag{2.12}$$

2. En el Teorema 2.2 la base que aparece es infinita, pues en el desarrollo no se necesita que n esté acotado. Es más, se puede ver fácilmente de la construcción de la sucesión que una base ortonormal $\{p_0, \ldots, p_n, p_{n+1}\}$ de Π_{n+1} es una base ortonormal de Π_n ($\{p_0, \ldots, p_n\}$) más un polinomio p_{n+1} unitario adicional que es ortogonal a todos los n anteriores. A partir de aquí, es inmediato ver que la mejor aproximación a f en Π_{n+1} va a ser

$$p^{(n+1)}(x) = p^{(n)}(x) + \langle f, p_{n+1} \rangle_w p_{n+1}$$

   (donde $p_{(n)}$ es la mejor aproximación a f en Π_n

2.3 Polinomios ortogonales más usados

Dependiendo de la función peso $w(x)$ y el intervalo $[a, b]$ en que estemos aproximando se van a usar diferentes familias de polinomios ortogonales.

Los más habituales en las aplicaciones son los siguientes, que usualmente se expresan en una forma no unitaria, es decir, son ortogonales pero $\| p_i \|_{2,w} \neq 1$:

2.3.1 Polinomios ortogonales de Legendre

Es una sucesión de polinomios ortogonales respecto a la función peso

$$w(x) \equiv 1, \qquad \text{sobre el intervalo } [-1, 1].$$

Se definen como

$$P_n(x) = \frac{(-1)^n}{2^n n!} \cdot \frac{d^n}{dx^n} \left[(1 - x^2)^n \right], \quad n \geq 1, \qquad P_0(x) \equiv 1. \tag{2.13}$$

Son ortogonales en $[-1, 1]$, el grado de P_n es n y $P_n(1) = 1$, $\forall n$. Su norma resulta

$$\langle P_n, P_n \rangle_w = \frac{2}{2n + 1},$$

por lo que la sucesión ortonormal que se usa es

$$p_n(x) = \sqrt{\frac{2n + 1}{2}} P_n(x).$$

Además, es habitual usarlos también en el caso $[a, b]$ con la misma función peso haciendo un cambio de variables a $[-1, 1]$.

Estos polinomios son un caso particular de los llamados *polinomios ortogonales de Jacobi*, ortogonales respecto a la función peso

$$w(x) = (1-x)^{\alpha}(1+x)^{\beta} \quad \text{en el intervalo } [-1, 1],$$

cuya fórmula es

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!}(1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^n}{dx^n}\left[(1-x)^{\alpha+n}(1+x)^{\beta+n}\right].$$

2.3.2 Polinomios ortogonales de Laguerre

Esta sucesión es de polinomios ortogonales respecto la función peso

$$w(x) = e^{-x}, \qquad \text{en el intervalo } [a, b] = [0, \infty).$$

En este caso vienen dados como

$$L_n(x) = \frac{1}{n!e^{-x}} \cdot \frac{d^n}{dx^n}(x^n e^{-x}), \quad n \geq 0.$$

Se tiene que

$$\| L_n \|_{2,w} = 1, \quad \forall n.$$

2.3.3 Polinomios ortogonales de Chebychev

Esta es la clase más importante de polinomios ortogonales en la teoría de la aproximación numérica por muchas razones. La primera es que es la que mejor resultados aporta en la práctica. La segunda es que a partir de ellos se introduce de forma natural la aproximación trigonométrica, que estará íntimamente relacionada con los desarrollos en serie de Fourier. Finalmente, son un conjunto de polinomios muy usados también en otros campos del análisis numérico, como, por ejemplo, la integración numérica.

Se define la *sucesión de polinomios de Chebychev (de primera especie)* a la sucesión de polinomios ortogonales respecto a la función peso

$$w(x) = \frac{1}{\sqrt{1-x^2}} \quad \text{en el intervalo } [-1, 1],$$

dados por

$$T_n(x) = \cos(n \arccos x), \qquad n \geq 0. \tag{2.14}$$

Aplicando propiedades básicas trigonométricas es fácil ver que resultan

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \\ &\vdots \end{aligned}$$

Estudiémoslo con un poco más de detalle. Veamos que son ortogonales entre sí con respecto a la función peso dada:

$$\langle T_j, T_k \rangle_w = \int_{-1}^{1} T_j(x)T_k(x)\frac{dx}{\sqrt{1-x^2}}.$$

Haciendo el cambio de variable $\theta = \arccos x$, $d\theta = -1/\sqrt{1-x^2}$ se tiene:

$$\langle T_j, T_k \rangle_w = -\int_{\pi}^{0} \cos(j\theta)\cos(k\theta)\, d\theta.$$

Aplicando la igualdad trigonométrica

$$\cos a \cos b = \frac{1}{2}(\cos(a+b) + \cos(a-b)),$$

se llega fácilmente a que

$$\langle T_j, T_k \rangle_w = \begin{cases} 0, & \text{si } j \neq k \\ \pi, & \text{si } j = k = 0 \\ \dfrac{\pi}{2}, & \text{si } j = k > 0 \end{cases}$$

Como no son ortonormales, la mejor aproximación a $f \in C[-1,1]$ en Π_n respecto a la función peso $w(x) = 1/\sqrt{1-x^2}$ va a ser

$$p_n(x) = \sum_{k=0}^{n} c_k T_k(x)$$

donde

$$c_0 = \frac{\langle f, T_0 \rangle_w}{\langle T_0, T_0 \rangle_w} = \frac{1}{\pi} \int_{-1}^{1} f(x) \frac{dx}{\sqrt{1-x^2}},$$

$$c_k = \frac{\langle f, T_k \rangle_w}{\langle T_k, T_k \rangle_w} = \frac{2}{\pi} \int_{-1}^{1} f(x) T_k(x) \frac{dx}{\sqrt{1-x^2}}, \qquad k = 1, 2, \ldots, n.$$

Observación: También existen los *polinomios ortogonales de Chebyshev de segunda especie*, que son ortogonales respecto la función peso $w(x) = \sqrt{1-x^2}$ en $[-1,1]$ y que vienen dados por

$$U_n(x) = \frac{\sin((n+1)\arccos x)}{\sqrt{1-x^2}}.$$

No se suelen estudiar pues no tienen mucha aplicación práctica.

2.4 Más sobre polinomios ortogonales

La sucesión de polinomios ortogonales dada por el Teorema 2.2 no sólo tiene aplicación en la aproximación numérica. Como veremos también jugarán un importante papel en la integración numérica, aparte de otros campos del análisis numérico.

Veamos algunas propiedades más de estos polinomios.

Si $\{\varphi_n(x) \,/\, n \geq 0\}$ es una familia de polinomios ortogonales en $[a,b]$ con función peso $w(x)$ y con grado$(\varphi_n) = n$, $n \geq 0$, se tienen las siguientes propiedades:

1. Cualquier polinomio p de grado m se puede expresar de forma única como

$$p(x) = \sum_{n=0}^{m} d_n \varphi_n(x), \quad x \in [a,b],$$

   donde

$$d_n = \frac{\langle p, \varphi_n \rangle_w}{\langle \varphi_n, \varphi_n \rangle_w}, \quad 0 \leq n \leq m.$$

2. Si p es un polinomio de grado $\leq m-1$ entonces

$$\langle p, \varphi_m \rangle_w = 0,$$

   o sea, φ_m es ortogonal a p.

3. El polinomio φ_n tiene exactamente n raíces reales distintas (y por tanto simples) en el intervalo abierto (a,b).

4. Satisfacen una ley de recurrencia a tres términos. Esto es, si

$$\varphi_n(x) = A_n x^n + B_n x^{n-1} + \cdots, \quad a_n = \frac{A_{n+1}}{A_n}, \quad \gamma_n = \langle \varphi_n, \varphi_n \rangle_w > 0,$$

se tiene

$$\varphi_{n+1}(x) = (a_n x + b_n)\varphi_n(x) - c_n \varphi_{n-1}(x)$$

con

$$b_n = a_n \left(\frac{B_{n+1}}{A_{n+1}} - \frac{B_n}{A_n} \right), \qquad c_n = \frac{A_{n+1} A_{n-1}}{A_n^2} \cdot \frac{\gamma_n}{\gamma_{n-1}}.$$

(No veremos la demostración por ahora).

Las leyes a recurrencia a tres términos son muy útiles a la hora de computar estos polinomios. En los casos de polinomios ortogonales más habituales tenemos:

1. **Los polinomios de Legendre:** Su fórmula de recurrencia a tres términos resulta

$$P_{n+1}(x) = \frac{2n+1}{n+1} x P_n(x) - \frac{n}{n+1} P_{n-1}(x), \quad n \geq 1, \qquad P_0(x) = 1, \ P_1(x) = x.$$

2. **Los polinomios de Laguerre:** En este caso, la relación de recurrencia a tres términos es:

$$L_{n+1}(x) = \frac{1}{n+1}(2n+1-x) L_n(x) - \frac{n}{n+1} L_{n-1}(x), \quad n \geq 1, \qquad L_0(x) = 1, \ L_1(x) = 1 - x.$$

3. **Los polinomios de Chebyshev de primera especie:** Para estos, la relación de recurrencia a tres términos es:

$$T_{n+1}(x) = 2x T_n(x) - T_{n-1}(x), \quad n \geq 1, \qquad T_0(x) = 1, \ T_1(x) = x.$$

3 Convergencia de la mejor aproximación

Ya hemos visto que el problema **(P. A.)** tiene solución única. Modificando un poco la notación para que quede patente la dependencia de n, esto quiere decir que para cada n existe un único $p_n \in \Pi_n$ de tal forma que

$$\| f - p_n \|_{2,w} \leq \| f - q_n \|_{2,w}, \quad \forall q_n \in \Pi_n,$$

o, lo que es lo mismo, p_n es lo más "cerca" (en norma 2) que se puede estar de f en Π_n. Pero como nosotros lo que queremos es aproximar f, la pregunta inmediata es ¿p_n se aproxima a f realmente?, ¿está lo suficientemente cerca? ¿$\lim_{n\to\infty} p_n = f$?

La respuesta a esta pregunta es que depende de qué entendemos por este límite, es decir, en qué norma lo estemos calculando.

3.1 Convergencia en norma 2

El primer resultado es que sí que hay convergencia en norma 2 (con peso $w(x)$):

Teorema 3.1 *Si $[a, b]$ es finito, $f \in C[a, b]$, se tiene*

$$\lim_{n\to\infty} \| f - p_n \|_{2,w} = 0.$$

Este teorema aporta una interesante propiedad de los coeficientes de la mejor aproximación.

Teorema 3.2 *Si $f \in C[a, b]$, y $\{\varphi_0, \varphi_1, \ldots, \varphi_n, \ldots\}$ es la base ortonormal dada en el Teorema 2.2 respecto a la función peso $w(x)$, se tiene*

1. *Desigualdad de Bessel:*

$$\| p_n \|_{2,w}^2 = \sum_{j=0}^{n} \langle f, \varphi_j \rangle_w^2 \leq \| f \|_{2,w}^2 \, . \tag{3.1}$$

2. *Igualdad de Parseval:*

$$\| f \|_{2,w} = \left(\sum_{j=0}^{\infty} \langle f, \varphi_j \rangle_w^2 \right)^{1/2} . \tag{3.2}$$

Observación: De este teorema se deduce que la serie de términos positivos

$$\sum_{j=0}^{\infty} \langle f, \varphi_j \rangle_w^2$$

es convergente, lo que implica que sus términos deben tender a 0 a medida que j aumenta:

$$\lim_{j \to \infty} \langle f, \varphi_j \rangle_w = 0,$$

es decir, que los coeficientes c_j de la mejor aproximación por mínimos cuadrados convergen a 0.

3.2 Convergencia uniforme

A la hora de tomar en la práctica el valor de la mejor aproximación en mínimos cuadrados $p_n(x)$ como aproximante de $f(x)$, se intenta demostrar una convergencia más fuerte, con la norma infinito

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

Para ello ya hay que exigir buenas condiciones de derivabilidad a la función f. Además, ya es más difícil encontrar resultados generales que se verifiquen para cualquier sucesión de polinomios ortogonales. Veremos sólo los casos de los polinomios de Legendre y de los de Chebychev.

Teorema 3.3 *Considérese la base de polinomios ortogonales de Legendre $\{P_i(x)\}$, dados en (2.13), en $[-1,1]$ respecto a la función peso $w(x) \equiv 1$, y la mejor aproximación por mínimos cuadrados de grado n:*

$$p_n(x) = \sum_{i=0}^{n} c_i P_i(x), \qquad c_i = \frac{\langle f, P_i \rangle_w}{\langle P_i, P_i \rangle_w}.$$

Si $f \in C^2[-1,1]$ y $\epsilon > 0$, existe $N \geq 1$ tal que $\forall n \geq N$:

$$\| f - p_n \|_\infty \leq \frac{\epsilon}{\sqrt{n}}.$$

Por tanto,

$$\lim_{n \to \infty} \| f - p_n \|_\infty = 0$$

y además converge con una velocidad mayor que la de $1/\sqrt{n}$.

Los polinomios de Chebychev mejoran esta velocidad de convergencia.

Teorema 3.4 *Considérese la base de polinomios ortogonales de Chebychev $\{T_i(x)\}$, dados en (2.14), en $[-1,1]$ respecto a la función peso $w(x) = 1/\sqrt{1-x^2}$, y la mejor aproximación por mínimos cuadrados de grado n:*

$$p_n^*(x) = \sum_{i=0}^{n} c_i^* T_i(x), \qquad c_i^* = \frac{\langle f, T_i \rangle_w}{\langle T_i, T_i \rangle_w}.$$

Si $f \in C^2[-1,1]$ con $r \geq 1$, se tiene que

$$\lim_{n \to \infty} \| f - p_n^* \|_\infty = 0.$$

Concluyendo, p_n^* converge uniformemente a f siempre que f sea al menos $C^2[-1,1]$. Hay un resultado más complejo de demostrar que rebaja la restricción a que f sea al menos $C^1[-1,1]$ y que nos dice, además, que se aumenta la velocidad de convergencia cuanto más suave sea f:

Teorema 3.5 *Considérese la base de polinomios ortogonales de Chebychev $\{T_i(x)\}$ y la mejor aproximación por mínimos cuadrados p_n^* de grado n dada en el teorema anterior. Si $f \in C^r[-1,1]$ con $r \geq 1$, se tiene que*

$$\| f - p_n^* \|_\infty \leq \frac{B \ln n}{n^r}, \qquad n \geq 2$$

donde B es una constante que depende de f y de r.

Observaciones:

- Considerando la definición de las series funcionales reales, la convergencia uniforme lo que nos da es la convergencia uniforme de las series

  $$\sum_{k=0}^{\infty} c_k P_k = \lim_{n\to\infty} p_n = f$$

  en el caso de Legendre y

  $$\sum_{k=0}^{\infty} c_k^* T_k = \lim_{n\to\infty} p_n^* = f$$

  en el caso de Chebychev.

- Si se quiere rebajar la condición de derivabilidad de la función f ya se pasa a conseguir resultados sólo sobre la convergencia puntual de la sucesión de aproximantes por mínimos cuadrados. Esta cuestión es mejor dejarla para cursos más especializados que el que nos ocupa.

4 Aproximación trigonométrica

Como ya vimos en el anterior tema sobre la interpolación, las funciones periódicas aparecen en muchas aplicaciones. Vimos también que podíamos trabajar sin pérdida de generalidad sólo con las funciones periódicas de periodo 2π:

$$f(\theta + 2\pi) = f(\theta), \quad -\infty < \theta < \infty,$$

como es el caso de las funciones trigonométricas. Denotemos

$$C_{2\pi} := \{\text{funciones reales periódicas de periodo } 2\pi\}.$$

Para aproximar este tipo de funciones en lugar de utilizar polinomios ordinarios, pues nunca podrán imitar el comportamiento periódico, usaremos la misma clase de funciones que usamos en la interpolación: la clase de *polinomios trigonométricos de grado n*:

$$p_n(\theta) = a_0 + \sum_{j=1}^{n} a_j \cos(j\theta) + b_j \sin(j\theta). \tag{4.1}$$

Para facilitar el desarrollo posterior, denotemos también

$$\mathcal{T}_n = \{p_n(\theta) \text{ polinomios trigonométricos de grado } \leq n\},$$

$$\mathcal{T} = \{\text{polinomios trigonométricos}\}$$

En el espacio de funciones $C_{2\pi}$ consideremos el producto interior con función peso $w(\theta) = 1$ dado por

$$\langle f, g \rangle = \int_0^{2\pi} f(\theta) g(\theta)\, d\theta.$$

Entonces cabe plantearse el problema de encontrar la *mejor aproximación a una función* $f \in C_{2\pi}$ *dentro del conjunto de polinomios trigonométricos de grado n,* $\mathcal{T}_n$*, respecto a la norma euclídea asociada al producto interior*:

$$\| f \|_2^2 = \langle f, f \rangle = \int_0^{2\pi} f(\theta)^2 \, d\theta.$$

En este caso se realiza un desarrollo análogo al deducido en el caso de los polinomios ordinarios con la ventaja de que ya partimos de una base especial de $\mathcal{T}$:

$$\mathcal{B} = \{1, \cos\theta, \sin\theta, \cos 2\theta, \sin 2\theta, \dots, \cos n\theta, \sin n\theta, \dots\}. \tag{4.2}$$

Esta base es especial porque es *ortogonal* pues (en el Anexo incluimos las igualdades trigonométricas que se usan):

$$\int_0^{2\pi} \cos k\theta \cos j\theta \, d\theta = \begin{cases} 0, & \text{si } k \neq j \\ 2\pi, & \text{si } k = j = 0 \\ \pi, & \text{si } k = j > 0 \end{cases}$$

$$\int_0^{2\pi} \sin k\theta \sin j\theta \, d\theta = \begin{cases} 0, & \text{si } k \neq j \\ \pi, & \text{si } k = j > 0 \end{cases} \tag{4.3}$$

$$\int_0^{2\pi} \sin k\theta \cos j\theta \, d\theta = 0, \qquad \forall k, j \geq 0.$$

Y, por tanto, la mejor aproximación trigonométrica por mínimos cuadrados va a ser

$$p_n(\theta) = a_0 + \sum_{j=1}^{n} a_j \cos(j\theta) + b_j \sin(j\theta). \tag{4.4}$$

donde

$$\begin{aligned} a_0 &= \frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) \, d\theta, \\ a_j &= \frac{\langle f, \cos j\theta \rangle}{\langle \cos j\theta, \cos j\theta \rangle} = \frac{1}{\pi} \int_0^{2\pi} f(\theta) \cos j\theta \, d\theta, \quad j = 1, \dots, n \\ b_j &= \frac{\langle f, \sin j\theta \rangle}{\langle \sin j\theta, \sin j\theta \rangle} = \frac{1}{\pi} \int_0^{2\pi} f(\theta) \sin j\theta \, d\theta, \quad j = 1, \dots, n \end{aligned} \tag{4.5}$$

Es decir, los coeficientes de la mejor aproximación son exactamente *los coeficientes del desarrollo en serie de Fourier*[2] *de la función* f. O, lo que es lo mismo, la mejor aproximación es la *suma parcial* $n-$*ésima de la serie de Fourier de* f.

Esta mejor aproximación va a converger puntualmente a $f(\theta)$ a medida que n aumenta:

Teorema 4.1 *Si* f *es una función continua en* $[0, 2\pi]$ *y diferenciable en* θ*, entonces,*

$$\lim_{n \to \infty} p_n(\theta) = f(\theta).$$

Observación: Realmente, en el caso general de la aproximación polinómica por mínimos cuadrados a una función $f \in C[a, b]$, donde se obtiene la mejor aproximación (2.9) con los coeficientes dados por (2.8) en la literatura se suele llamar a dichos c_k *coeficientes de Fourier de* f *en* $[a, b]$ *respecto a la base ortonormal dada.*

Así el estudio del desarrollo en serie de Fourier de una función pasa por construir un sistema ortonormal de funciones y calcular la mejor aproximación a f respecto a dicho sistema ortonormal.

[2]Ver Anexo

5 Aproximación discreta por mínimos cuadrados

Hasta ahora hemos aproximado funciones dadas de forma analítica y la hemos aproximado por polinomios también dados analíticamente. Pero lo más habitual en la práctica es que sólo se conozca un conjunto de valores de una función (nube de puntos) sin saber exactamente cuál es su expresión analítica. Así que lo que se va a buscar es un polinomio que sustituya a dicha función desconocida mediante la minimización del error cuadrático, pero sólo en los puntos conocidos. Esto tiene mucha aplicación en Estadística, en los llamados *métodos de regresión lineal*, que son un caso particular de la aproximación discreta por mínimos cuadrados que vamos a ver.

Igual que en el caso general, el problema de aproximación discreta por mínimos cuadrados de una función f consistirá en encontrar el polinomio p_n, de un grado prefijado n, que esté más "cerca" de f, o, lo que es lo mismo, que minimice la "distancia cuadrática" al conjunto de puntos de f conocidos. La diferencia es que la "distancia" en norma 2 que hemos visto hasta ahora necesita tener información de la función f en todo un intervalo, y ahora sólo tenemos la información en un conjunto de puntos discreto.

Para usar sólo la información discreta que tenemos, vamos a definir una nueva distancia cuadrática discreta.

Dado un conjunto de puntos $\{x_i\}_{i=1}^m \subset [a,b]$ y una función peso $w(x)$ en $[a,b]$ (de la que conocemos los valores $w(x_i) > 0$, $\forall i$), definimos el *pseudo producto interior discreto*:

$$\langle f,g\rangle_w = \sum_{i=1}^m f(x_i)g(x_i)w(x_i), \qquad f,g \in C[a,b].$$

Lo llamamos pseudo producto interior, pues *no verifica* la condición de los productos interiores de que $\langle f,f\rangle_w = 0$ sii $f = 0$.

A partir de este pseudo producto definimos la *seminorma en $C[a,b]$* como

$$\| f \|_{2,w} = \langle f,f\rangle_w^{1/2} = \left(\sum_{i=1}^m f(x_i)^2 w(x_i)\right)^{1/2}.$$

Este funcional verifica las condiciones de que $\| \alpha f \|_{2,w} = |\alpha| \; \| f \|_{2,w}$, $\forall \alpha \in I\!R$ y la desigualdad triangular, pero no la condición de que $\| f \|_{2,w} = 0$ sii $f = 0$, por lo que no es una norma propiamente dicha. Pero $\| f-g \|_{2,w}$ sí que mide la distancia cuadrática entre f y g sobre los puntos x_i, que es lo que nos interesa.

Así que el problema de aproximación polinómica discreta por mínimos cuadrados ahora se plantea de la forma siguiente:

Dada una función $f \in C[a,b]$ y un $n > m$ prefijado, encontrar un polinomio $p \in \Pi_n$ de tal forma que

$$\| f-p \|_{2,w} \leq \| f-q \|_{2,w}, \quad \forall q \in \Pi_n \tag{5.1}$$

De manera análoga al caso continuo, se considera una base de Π_n, $\{p_0, p_1, \ldots, p_n\}$, y se busca el polinomio p de la forma

$$p(x) = \sum_{j=0}^n c_j p_j(x)$$

que verifique (5.1), es decir, que minimice $G(c_0, c_1, \ldots, c_n) := \| f-p \|_{2,w}$. Siguiendo el mismo razonamiento que en el caso continuo, se llega a que esto es equivalente a que la matriz de coeficientes del sistema lineal

$$AC = B \tag{5.2}$$

$$A = \begin{pmatrix} \langle p_0,p_0\rangle_w & \langle p_1,p_0\rangle_w & \cdots & \langle p_n,p_0\rangle_w \\ \langle p_0,p_1\rangle_w & \langle p_1,p_1\rangle_w & \cdots & \langle p_n,p_1\rangle_w \\ \vdots & \vdots & \ddots & \vdots \\ \langle p_0,p_n\rangle_w & \langle p_1,p_n\rangle_w & \cdots & \langle p_n,p_n\rangle_w \end{pmatrix}, \quad C = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad B = \begin{pmatrix} \langle f,p_0\rangle_w \\ \langle f,p_1\rangle_w \\ \vdots \\ \langle f,p_n\rangle_w \end{pmatrix},$$

sea definida positiva (es simétrica siempre), con la única diferencia de que ahora $\langle\cdot\rangle_w$ representa al pseudo producto interior. Es decir, tenemos que ver que $c^TAc \geq 0, \forall c$ y que $c^TAc = 0$ sii $c = 0$. Para ello, elegimos cualquier vector $c = (c_0, c_1, \ldots, c_n)$ y desarrollamos el producto:

$$\begin{aligned} c^TAc &= \sum_{k=0}^{n}\sum_{j=0}^{n} c_kc_j\langle p_k, p_j\rangle_w = \sum_{k=0}^{n}\sum_{j=0}^{n} c_kc_j \sum_{i=1}^{m} p_k(x_i)p_j(x_i)w(x_i) \\ &= \sum_{i=1}^{m}\left\{\sum_{k=0}^{n}\sum_{j=0}^{n} c_kc_jp_k(x_i)p_j(x_i)\right\} w(x_i) = \sum_{i=1}^{m}\left(\sum_{k=0}^{n} c_kp_k(x_i)\right)\left(\sum_{j=0}^{n} c_jp_j(x_i)\right) w(x_i) \\ &= \sum_{i=1}^{m} q(x_i)^2w(x_i) \geq 0 \end{aligned}$$

donde $q(x)$ es el polinomio de Π_n dado por $q(x) = \sum_{j=0}^{n} c_jp_j(x)$. Además, del mismo desarrollo anterior, $c^TAc = 0$ sii

$$q(x_i)^2w(x_i) = 0, \quad \forall i = 1, \ldots, m,$$

o, lo que es equivalente desde que $w(x_i) > 0$, $\forall i$ que

$$q(x_i) = 0, \forall i = 1, \ldots, m.$$

Por tanto, q es un polinomio de grado $\leq n$ con $m > n$ raíces distintas entre sí. Esto es imposible salvo que sea el polinomio nulo $q \equiv 0$, o, lo que es lo mismo, que $c = 0$.

En conclusión, la matriz A es simétrica y definida positiva, por lo que el sistema (5.2) tendrá solución única y además, es fácil de computar numéricamente, obteniendo de forma unívoca la mejor aproximación discreta a f en Π_n por mínimos cuadrados.

En este caso, se resuelve numéricamente el sistema (5.2) sin tener que echar mano de los polinomios ortogonales, pues el cálculo de los coeficientes de A y de b es mucho más simple.

6 Anexo

- **Proceso de ortonormalización de Gram–Schmidt:** Dada una base $\{v_1, v_2, \ldots, v_n\}$ cualquiera de un espacio vectorial V de dimensión n (con un producto interior definido $\langle\cdot,\cdot\rangle$), se puede construir una base $\{u_1, u_2, \ldots, u_n\}$ de V ortonormal, esto es,

  $$\langle u_i, u_j\rangle = \delta_{ij}.$$

  Para ello, se define primero el vector unitario

  $$u_1 = \frac{v_1}{\| v_1 \|_2}.$$

  Para calcular el segundo vector de la nueva base se define previamente

  $$v_2' = v_2 - \langle v_2, u_1\rangle u_1,$$

  que es no nulo (si lo fuera v_1 y v_2 serían linealmente dependientes) y ortogonal a u_1 pues

  $$\langle v_2', u_1\rangle = \langle v_2, u_1\rangle - \langle v_2, u_1\rangle \| u_1 \|_2^2 = 0.$$

  A partir de éste definimos el segundo vector de la base ortonormal como

  $$u_2 = \frac{v_2'}{\| v_2' \|_2}.$$

  Para calcular el tercero, consideramos el vector

  $$v_3' = v_3 - \langle v_3, u_1\rangle u_1 - \langle v_3, u_2\rangle u_2,$$

que vuelve a ser no nulo por la independencia de los v_i y tal que

$$\langle v_3', u_1\rangle = \langle v_3, u_1\rangle - \langle v_3, u_1\rangle \parallel u_1 \parallel_2^2 - \langle v_3, u_2\rangle\langle u_2, u_1\rangle = 0,$$
$$\langle v_3', u_2\rangle = \langle v_3, u_2\rangle - \langle v_3, u_1\rangle\langle u_1, u_2\rangle - \langle v_3, u_2\rangle \parallel u_2 \parallel_2^2 = 0,$$

por lo que tomamos como tercer vector de la base ortonormal a

$$u_3 = \frac{v_3'}{\parallel v_3' \parallel_2}.$$

Por inducción se define el vector

$$v_r' = v_r - \sum_{i=1}^{r-1} \langle v_r, u_i\rangle u_i,$$

que es no nulo pues v_r no depende de $\{u_1, u_2, \dots, u_{r-1}\}$ y ortogonal a dichos $u_1, u_2, \dots, u_{r-1}$ ya que para $j = 0, 1, \dots, r-1$:

$$\langle v_r', u_j\rangle = \langle v_r, u_j\rangle - \sum_{i=1}^{r-1} \langle v_r, u_i\rangle\langle u_i, u_j\rangle = \langle v_r, u_j\rangle - \langle v_r, u_j\rangle = 0.$$

Y finalmente se toma como $r-$ésimo vector de la base ortonormal a

$$u_r = \frac{v_r'}{\parallel v_r' \parallel_2}.$$

- **Igualdades trigonométricas necesarias para demostrar (4.3)**:

$$\cos\alpha\cos\beta = \frac{1}{2}(\cos(\alpha-\beta) + \cos(\alpha+\beta)),$$
$$\sin\alpha\sin\beta = \frac{1}{2}(\cos(\alpha-\beta) - \cos(\alpha+\beta)),$$
$$\sin\alpha\cos\beta = \frac{1}{2}(\sin(\alpha-\beta) + \sin(\alpha+\beta)).$$

- **Desarrollo en serie de Fourier:** La idea original propuesta por Fourier (1807) fue que una función "arbitraria" se puede expresar como combinación lineal de senos y cosenos. Esta idea es muy importante cuando estudiamos sobre todo funciones periódicas (movimiento ondulatorio, por ejemplo).

  La clase de funciones que usa son funciones $f \in L^2(I)$ donde $I \subseteq \mathbb{R}$ es un intervalo, es decir, funciones complejas medibles en I y con cuadrado $|f|^2$ integrable (mediante la integral de Lebesgue). En dicho espacio se define el producto interior

$$\langle f, g\rangle = \int_I f(x)\overline{g(x)}\, dx,$$

  y la norma 2 asociada $\parallel f \parallel_2 = (\langle f, f\rangle)^{1/2}$.

  Se dan dos variantes: real y compleja.

  1. Desarrollo en serie de Fourier real: $I = [0, 2\pi]$:

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos nt\, dt, \qquad b_n = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin nt\, dt.$$

2. Desarrollo en serie de Fourier complejo: $I = [0, 2\pi]$, f toma valores complejos:

$$f(x) \sim \sum_{n=0}^{\infty} c_n e^{inx},$$

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int}\, dt.$$

Obsérvese que se expresa con el símbolo $\sim$ en lugar del igual pues el estudio de la convergencia de la serie es muy complicado.

La teoría de Fourier demuestra que si f es continua en $[0, 2\pi]$ y periódica de período 2π y $\{s_n\}$ es la sucesión de sumas parciales de la serie de Fourier de f, es decir,

$$s_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kx + b_k \sin kx,$$

entonces

$$\lim_{n\to\infty} \int_0^{2\pi} |s_n(x) - f(x)|^2\, dx = 0,$$

o, lo que es lo mismo, hay convergencia en norma 2 $\|\cdot\|_2$.