Capítulo 9

Mínimos Cuadrados

Resumen

El tratamiento de datos experimentales tiene lugar en este capítulo. Se tratan dos tipos de problemas: los discretos y los continuos. En el caso discreto, se comienza con el planteamiento del problema de mínimos cuadrados. A continuación se trata el método de Householder que constituye uno de los más importantes métodos de resolución de sistemas de ecuaciones lineales. Se considera el ajuste de datos de algunos problemas que conducen a resolver sistemas de ecuaciones lineales en mínimos cuadrados. A continuación se trata el ajuste de datos en los que hay que determinar un parámetro. Posteriormente, se trata el problema de mínimos cuadrados continuos y se da aplicaciones a la aproximación de las series de Fourier.

9.1. Introducción

En muchas observaciones científicas se deben determinar los valores de ciertas constantes $a_1, \ldots, a_n$. Sin embargo, determinar o medir dichas constantes resulta muy difícil y por lo general imposible. En tales casos, el método indirecto siguiente es aplicado: en vez de observar los a_i resulta más fácil tomar una muestra de una cantidad medible "y" la cual depende de los a_i y de las mediciones experimentales que denotamos por x, esto es,

$$y = f\left(x, a_1, \ldots, a_n\right).$$

Con el propósito de determinar a_i, se realizan experimentos bajo m condiciones diferentes $x_1, \ldots, x_m$, obteniéndose m resultados diferentes: $y_k = f\left(x_k, a_1, \ldots, a_n\right) \quad k = 1, \ldots, m$. En general, al menos $m \geq n$ experimentos deben ejecutarse con el propósito de determinar $a_i \quad i = 1, \ldots, n$. Además, estos valores a_i deben satisfacer la relación precedente.

Si $m > n$, $y_k = f\left(x_k, a_1, \ldots, a_n\right) \quad k = 1, \ldots, m$ forman un sistema sobredeterminado para los parámetros desconocidos $a_1, \ldots, a_n$ que usualmente no tiene solución porque las cantidades observadas y_i están perturbadas por errores de medición. Consecuentemente, en vez de encontrar una solución exacta de dicho sistema, el problema se traduce en encontrar una mejor aproximación posible aplicando la conocida técnica de mínimos cuadrados. Este método fue publicado por primera vez por Legendre en 1805. Esta clase de problemas se los conoce como ajuste de datos.

La función f es conocida. Esta función se elige siguiendo varios criterios:

1. Se tiene un modelo matemático gobernado, por ejemplo, por ecuaciones diferenciales ordinarias y que tiene como solución la función f que depende de n parámetros $a_i \quad i = 1, \ldots, n$ que deben determinarse de la información experimental existente.$S = \{(x_i, y_i) \mid i = 1, \ldots, n\}$.
2. En base al conjunto de datos experimentales $S = \{(x_i, y_i) \mid i = 1, \ldots, n\}$ representados gráficamente y por alguna información suplementaria se intuye que siguen un comportamiento del tipo $y = f\left(x, a_1, \ldots, a_n\right)$.

3. Dado el conjunto de datos $S = \{(x_i, y_i) \mid i = 1, \ldots, n\}$ se fija directamente la función f así como los parámetros a determinar $a_1, \ldots, a_n$, esto es $y = f(x, a_1, \ldots, a_n)$.

En todos los casos suponemos que

$$y_k = f(x_k, a_1, \ldots, a_n) + r_k \quad k = 1, \ldots, m,$$

donde cada r_k es la perturbación del dato experimentas (x_k, y_k). Esta perturbación depende de los parámetros $a_1, \ldots, a_n$ que deben determinarse. Escribimos

$$r_k(a_1, \ldots, a_n) = y_k - f(x_k, a_1, \ldots, a_n) \quad k = 1, \ldots, m.$$

El método de mínimos cuadrados consiste en elegir la mejor opción de los parámetros $a_1, \ldots, a_n$ en el sentido que precisamos a continuación: sea $\Omega \subset \mathbb{R}^n$ abierto, se define la función E de Ω en $\mathbb{R}$ como sigue:

$$E(a_1, \ldots, a_n) = \sum_{k=1}^{m} r_k^2(a_1, \ldots, a_n) = \sum_{k=1}^{m} (y_k - f(x_k, a_1, \ldots, a_n))^2 \quad (a_1, \ldots, a_n) \in \Omega,$$

con lo que $E(a_1, \ldots, a_n) \geq 0 \quad \forall (a_1, \ldots, a_n) \in \Omega$; y se considera el problema siguiente:

$$\text{hallar } (\widehat{a}_1, \ldots, \widehat{a}_n) \in \Omega \text{ tal que } E(\widehat{a}_1, \ldots, \widehat{a}_n) = \min_{(a_1, \ldots, a_n) \in \Omega} E(a_1, \ldots, a_n).$$

Así, la mejor elección de los parámetros $a_1, \ldots, a_n$ es la solución del problema de minimización: $\min\limits_{(a_1, \ldots, a_n) \in \Omega} E(a_1, \ldots, a_n)$.

En la generalidad de los casos se supone que la función f es de clase C^1 en Ω, es decir que $\dfrac{\partial f}{\partial a_i}$ $i = 1, \ldots, n$, son continuas en Ω.

Para hallar un punto $(\widehat{a}_1, \ldots, \widehat{a}_n) \in \Omega$ en el que la función E alcanza su mínimo, aplicamos las condiciones necesarias de extremo, esto es:

$$\triangledown E(a_1, \ldots, a_n) = \overrightarrow{0} \Leftrightarrow \begin{cases} \dfrac{\partial E}{\partial a_1}(a_1, \ldots, a_n) = 0, \\ \qquad\vdots \\ \dfrac{\partial E}{\partial a_n}(a_1, \ldots, a_n) = 0, \end{cases}$$

con lo que obtenemos un sistema de ecuaciones lineal o no dependiendo exclusivamente de la función f. La solución de dicho sistema nos provee de un punto crítico $(\widehat{a}_1, \ldots, \widehat{a}_n)$ que puede ser en el que E alcanza su mínimo.

Puesto que para cada $i = 1, \ldots, m$, se tiene

$$\begin{aligned} \frac{\partial E}{\partial a_i}(a_1, \ldots, a_n) &= \sum_{k=1}^{m} 2(y_k - f(x_k, a_1, \ldots, a_n)) \left(-\frac{\partial f}{\partial a_i}(x_k, a_1, \ldots, a_n) \right) \\ &= -2 \sum_{k=1}^{m} \left[y_k \frac{\partial f}{\partial a_i}(x_k, a_1, \ldots, a_n) - \frac{\partial f}{\partial a_i}(x_k, a_1, \ldots, a_n) f(x_k, a_1, \ldots, a_n) \right], \end{aligned}$$

se sigue que el sistema de ecuaciones precedente está definido como sigue:

$$\begin{cases} \sum\limits_{k=1}^{m} \left[y_k \dfrac{\partial f}{\partial a_1}(x_k, a_1, \ldots, a_n) - \dfrac{\partial f}{\partial a_1}(x_k, a_1, \ldots, a_n) f(x_k, a_1, \ldots, a_n) \right] = 0 \\ \qquad\qquad\vdots \\ \sum\limits_{k=1}^{m} \left[y_k \dfrac{\partial f}{\partial a_n}(x_k, a_1, \ldots, a_n) - \dfrac{\partial f}{\partial a_n}(x_k, a_1, \ldots, a_n) f(x_k, a_1, \ldots, a_n) \right] = 0. \end{cases}$$

En el caso en que la función f es lineal respecto de las variables (parámetros a determinar) $a_1, \ldots, a_n$, el sistema de ecuaciones precedente es lineal. Dos ejemplos de funciones lineales nos proporcionan los polinomios de grado m :

$$f(x, a_0, \ldots, a_n) = a_0 + a_1 x + \cdots + a_n x^m \quad x \in \mathbb{R}, \quad (a_0, \ldots, a_n) \in \mathbb{R}^{n+1},$$

y los polinomios trigonométricos:

$$\begin{aligned}
f(x, a_1, \ldots, a_m) &= \sum_{k=1}^{m} a_k \operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L], \\
f(x, a_0, \ldots, a_m) &= \frac{a_0}{2} + \sum_{k=1}^{m} a_k \cos\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L], \\
f(x, a_0, a_1, \ldots, a_m, b_1, \ldots, b_m) &= \frac{a_0}{2} + \sum_{k=1}^{m} \left[a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right)\right],
\end{aligned}$$

donde $L > 0$, $a_0, \ldots, a_m$, $b_1, \ldots, b_m$ son los coeficientes de Fourier.

En el caso de los polinomios trigonométricos, el cálculo aproximado de los coeficientes de Fourier ya fue establecido en el capítulo de la aproximación numérica de las series de funciones. En este capítulo mostraremos que los coeficientes que figuran en los polinomios trigonométricos son efectivamente los coeficientes de Fourier.

En el caso de los polinomios de grado m focalizaremos nuestra atención a los polinomios de grado 1, 2 y 3; esto es,

$$\begin{aligned}
f(x, a, b) &= a + bx \quad x \in \mathbb{R}, \\
f(x, a, b, c) &= a + bx + cx^2 \quad x \in \mathbb{R}, \\
f(x, a, b, c, d) &= a + bx + cx^2 + dx^3 \quad x \in \mathbb{R}.
\end{aligned}$$

En el caso en que f no es lineal respecto de las variables $a_1, \ldots, a_n$, el sistema de ecuaciones a resolver es no lineal. En la generalidad de los casos se tiene f función de clase C^2 en Ω, y dicho sistema se resuelve en forma aproximada con el método de Newton, la linealización del sistema de ecuaciones se resuelve mediante el método de eliminación gaussiana. En pocos casos se conocen de métodos particulares para resolver dicho sistema cuando dependen de 1, 2, 3 parámetros. Más adelante trataremos algunos de estos ejemplos que tienen muchas aplicaciones.

Los problemas de mínimos cuadrados que acabamos de describir se extienden a funciones en dos o más variables, así a polinomios de grado $1, 2, \ldots$, en dos variables independientes $x,\ y$ como los que se describen a continuación:

$$\begin{aligned}
p(x, y) &= a + bx + cy \quad (x, y) \in \mathbb{R}^2, \\
p(x, y) &= a + bx + cy + dxy \quad (x, y) \in \mathbb{R}^2, \\
p(x, y) &= a + bx + cy + dxy + ex^2 + dy^2 \quad (x, y) \in \mathbb{R}^2,
\end{aligned}$$

que requieren, para el cálculo de las constantes que figuran en cada clase de polinomios, de un conjunto de datos $S = \left\{(x_k, y_k, z_k) \in \mathbb{R}^3 \mid k = 1, \ldots, n\right\}$. Otra clase de funciones son las conocidas funciones lineales y afines en n variables del tipo

$$z = f(x_1, \ldots, x_n) = \sum_{k=1}^{n} a_k x_n \quad (x_1, \ldots, x_n) \in \mathbb{R}^n$$

y

$$z = f(x_1, \ldots, x_n) = a_0 + \sum_{k=1}^{n} a_k x_n \quad (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

que requieren, para el cálculo de las constantes que figuran en cada una de esta clase de funciones, de un conjunto de datos

$$S = \left\{ \overrightarrow{x}_k = \left(x_1^{(k)}, \ldots, x_n^{(k)}, z_k \right) \in \mathbb{R}^{n+1} \mid k = 1, \ldots, m \right\}.$$

Comenzaremos con la resolución de los sistemas de ecuaciones lineales en mínimos cuadrados. Estos sistemas de ecuaciones provienen, en general, de problemas de ajuste de datos lineales. A continuación tratamos el ajuste de datos de datos polinomial. Luego tratamos el ajuste de datos de funciones afines del tipo

$$f\left(x_1, \ldots, x_n\right) = a_0 + \sum_{k=1}^{n} a_k x_n \qquad \left(x_1, \ldots, x_n\right) \in \mathbb{R}^n.$$

Concluimos con ejemplos de problemas de ajuste de datos de funciones no lineales.

9.2. Soluciones de sistemas de ecuaciones lineales en mínimos cuadrados.

Sean $m,\ n \in \mathbb{Z}^+$ con $m \geq n$, $A = (a_{ij}) \in M_{m\times n}[\mathbb{R}]$ no nula y de rango $\mathcal{R}(A) = n$, $\overrightarrow{b}^T = (b_1, ..., b_n) \in \mathbb{R}^m$. Consideramos el problema siguiente:

$$\text{hallar } \overrightarrow{x} \in \mathbb{R}^n \text{ solución del sistema de ecuaciones lineales} \quad A\overrightarrow{x} = \overrightarrow{b}.$$

Estos sistemas de ecuaciones se caracterizan por tener más ecuaciones que incógnitas. Estos sistemas, como hemos visto, surgen en la determinación de ciertos parámetros $x_1, ..., x_n$ que deben calcularse a partir de una información experimental y que corresponden a un modelo del tipo lineal.

Ponemos $A = [A_1, ..., A_n]$, donde A_j es la j-ésima columna de A y sea

$$W = L\left(A_1, ..., A_n\right) = \left\{ \sum_{j=1}^{n} \alpha_j A_j \mid \alpha_j \in \mathbb{R},\ \ j = 1, ..., n \right\},$$

el espacio constituido por todas las combinaciones lineales de $A_1, ..., A_n$. Por definición, la dimensión de W es el rango de A y que por hipótesis, $\mathcal{R}(A) = n$. Luego $\dim W = n$. Entonces, el sistema de ecuaciones $A\overrightarrow{x} = \overrightarrow{b}$ tiene solución si y solo si $\overrightarrow{b} \in W$. Esta situación se presenta en muy pocos casos. En la práctica, los sistemas de ecuaciones arriba propuesto, en general, no tienen solución.

Se define $\overrightarrow{r}(\overrightarrow{x}) = A\overrightarrow{x} - \overrightarrow{b}$, $\quad \overrightarrow{x} \in \mathbb{R}^n$. El vector $\overrightarrow{r}(\overrightarrow{x})$ se llama residuo. Proponemos un problema alterno denominado problema en mínimos cuadrados (P_a) que se indica a continuación:

$$\text{hallar, si existe, } \widehat{x} \in \mathbb{R}^n \text{ tal que} \quad \left\|\overrightarrow{r}(\widehat{x})\right\|^2 \leq \left\|\overrightarrow{r}(\overrightarrow{x})\right\|^2 \quad \forall \overrightarrow{x} \in \mathbb{R}^n,$$

o lo que es lo mismo

$$\left\|A\widehat{x} - \overrightarrow{b}\right\|^2 \leq \left\|A\overrightarrow{x} - \overrightarrow{b}\right\| \quad \forall \overrightarrow{x} \in \mathbb{R}^n,$$

que a su vez es equivalente al siguiente:

$$\text{hallar, si existe, } \widehat{x} \in \mathbb{R}^n \text{ tal que} \quad \left\|A\widehat{x} - \overrightarrow{b}\right\|^2 = \underset{\overrightarrow{x} \in \mathbb{R}^n}{\mathrm{Min}} \left\|A\overrightarrow{x} - \overrightarrow{b}\right\|^2,$$

donde $\|\cdot\|$ es la norma euclídea en $\mathbb{R}^m$.

Este problema lo enfrentamos de dos maneras. En la primera, probamos la existencia de $\widehat{x} \in \mathbb{R}^n$ mediante métodos del análisis matemático como minimización de un cierto funcional definido en $\mathbb{R}^n$. En la segunda, probamos la existencia de $\widehat{x} \in \mathbb{R}^n$ mediante métodos netamente del álgebra lineal utilizando la ortogonalidad.

1. **Minimización de un funcional cuadrático definido en $\mathbb{R}^n$.**

Sea J de $\mathbb{R}^n$ en $\mathbb{R}$ el funcional definido por $J(\overrightarrow{x}) = \left\| A\overrightarrow{x} - \overrightarrow{b} \right\|^2, \quad \overrightarrow{x} \in \mathbb{R}^n$. Se sabe que J es un funcional convexo. Hallemos la derivada de Gâteaux de J, esto es, la derivada direccional de J en $\overrightarrow{x} \in \mathbb{R}^n$ según la dirección $\overrightarrow{y}$, denotada $D_{\overrightarrow{y}} J(\overrightarrow{x})$ y definida como

$$D_{\overrightarrow{y}} J(\overrightarrow{x}) = \lim_{t \to 0} \frac{J(\overrightarrow{x} + t\overrightarrow{y}) - J(\overrightarrow{x})}{t}.$$

De la definición del producto escalar en $\mathbb{R}^n$, se tiene

$$J(\overrightarrow{x}) = \left\| A\overrightarrow{x} - \overrightarrow{b} \right\|^2 = \left(A\overrightarrow{x} - \overrightarrow{b}\right)^T \left(A\overrightarrow{x} - \overrightarrow{b}\right) \qquad \overrightarrow{x} \in \mathbb{R}^n.$$

Sean $\overrightarrow{x}$, $\overrightarrow{y} \in \mathbb{R}^n$ fijos, $t \neq 0$. Entonces,

$$\begin{aligned} J(\overrightarrow{x} + t\overrightarrow{y}) - J(\overrightarrow{x}) &= \left(A(\overrightarrow{x} + t\overrightarrow{y}) - \overrightarrow{b}\right)^T \left(A(\overrightarrow{x} + t\overrightarrow{y}) - \overrightarrow{b}\right) - \left(A\overrightarrow{x} - \overrightarrow{b}\right)^T \left(A\overrightarrow{x} - \overrightarrow{b}\right) \\ &= 2t\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T A\overrightarrow{y} + t^2 (A\overrightarrow{y})^T A\overrightarrow{y}. \end{aligned}$$

Se ha utilizado el hecho que $A\overrightarrow{x} - \overrightarrow{b}$, $A\overrightarrow{x} \in \mathbb{R}^m$, y, $\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T A\overrightarrow{y} = (A\overrightarrow{y})^T \left(A\overrightarrow{x} - \overrightarrow{b}\right)$, pués el producto escalar en $\mathbb{R}^m$ es conmutativo. Luego, para $t \neq 0$ se tiene

$$\frac{J(\overrightarrow{x} + t\overrightarrow{y}) - J(\overrightarrow{x})}{t} = 2\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T A\overrightarrow{y} + t(A\overrightarrow{y})^T A\overrightarrow{y},$$

de donde

$$\begin{aligned} D_{\overrightarrow{y}} J(\overrightarrow{x}) &= \lim_{t \to 0} \frac{J(\overrightarrow{x} + t\overrightarrow{y}) - J(\overrightarrow{x})}{t} = \lim_{t \to 0} \left(2\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T A\overrightarrow{y}\right) + t(A\overrightarrow{y})^T A\overrightarrow{y} \\ &= 2\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T A\overrightarrow{y}. \end{aligned}$$

Así, la derivada de Gâteaux de J en $\overrightarrow{x}$ según la dirección $\overrightarrow{y}$ está definida como

$$D_{\overrightarrow{y}} J(\overrightarrow{x}) = 2\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T A\overrightarrow{y}.$$

Por otro lado, un resultado muy conocido del Análisis Matemático sobre las derivadas direccionales es que si $D_{\overrightarrow{y}} J(\overrightarrow{x})$ es continuo en $\overrightarrow{x} \in \mathbb{R}^n$ en toda dirección $\overrightarrow{y} \in \mathbb{R}^n$, entonces

$$D_{\overrightarrow{y}} J(\overrightarrow{x}) = (\nabla J(\overrightarrow{x}))^T \overrightarrow{y},$$

donde $\nabla J(\overrightarrow{x})$ denota el gradiente de J en $\overrightarrow{x}$ definido por $(\nabla J(\overrightarrow{x}))^T = \left(\frac{\partial J}{\partial x_1}(\overrightarrow{x}), ..., \frac{\partial J}{\partial x_n}(\overrightarrow{x})\right)$. Luego,

$$(\nabla J(\overrightarrow{x}))^T \overrightarrow{y} = 2\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T A\overrightarrow{y} \qquad \forall \overrightarrow{y} \in \mathbb{R}^n,$$

con lo cual

$$\nabla J(\overrightarrow{x}) = \left(A\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T A\right)^T = 2A^T\left(A\overrightarrow{x} - \overrightarrow{b}\right)^T.$$

Las condiciones necesarias de extremo implican

$$\nabla J(\overrightarrow{x}) = 0 \Leftrightarrow A^T\left(A\overrightarrow{x} - \overrightarrow{b}\right) = 0 \Leftrightarrow A^T A\overrightarrow{x} = A^T\overrightarrow{b}.$$

El sistema de ecuaciones lineales $A^T A\overrightarrow{x} = A^T\overrightarrow{b}$, se llama sistema de ecuaciones normales. Note que la matriz $A^T A$ es una matriz simétrica y en consecuencia es una matriz normal.

Por hipótesis $R(A) = n$. Luego $R\left(A^T\right) = n$ y $R\left(A^T A\right) = n$. Como $A^T A \in M_{n\times n}[\mathbb{R}]$ y $R\left(A^T A\right) = n$, se sigue que $A^T A$ es invertible y en consecuencia

$$A^T A\overrightarrow{x} = A^T\overrightarrow{b} \Leftrightarrow \overrightarrow{x} = \left(A^T A\right)^{-1} A^T\overrightarrow{b}.$$

Ponemos

$$\widehat{x} = \left(A^T A\right)^{-1} A^T\overrightarrow{b}.$$

Probemos que $\left\|A\widehat{x} - \overrightarrow{b}\right\|^2 \leq \left\|A\overrightarrow{x} - \overrightarrow{b}\right\|^2 \quad \forall\overrightarrow{x} \in \mathbb{R}^n$. Sea $\overrightarrow{x} \in \mathbb{R}^n$. Entonces

$$\begin{aligned}\|\overrightarrow{r}(\overrightarrow{x})\|^2 &= (\overrightarrow{r}(\overrightarrow{x}))^T\overrightarrow{r}(\overrightarrow{x}) = [A(\overrightarrow{x}-\widehat{x}) + \overrightarrow{r}(\overrightarrow{x})]^T[A(\overrightarrow{x}-\widehat{x}) + \overrightarrow{r}(\overrightarrow{x})] \\ &= [A(\overrightarrow{x}-\widehat{x})]^T A(\overrightarrow{x}-\widehat{x}) + (\overrightarrow{r}(\widehat{x}))^T\overrightarrow{r}(\widehat{x}) = \|A(\overrightarrow{x}-\widehat{x})\|^2 + \|\overrightarrow{r}(\widehat{x})\|^2,\end{aligned}$$

donde el producto $(A(\overrightarrow{x}-\widehat{x}))^T\overrightarrow{r}(\widehat{x}) = 0$, pues $A^T\left(A\widehat{x} - \overrightarrow{b}\right) = 0$ y en consecuencia

$$[A(\overrightarrow{x}-\widehat{x})]^T\overrightarrow{r}(\widehat{x}) = [A(\overrightarrow{x}-\widehat{x})]^T\left(A\widehat{x} - \overrightarrow{b}\right) = (\overrightarrow{x}-\widehat{x})^T A^T\left(A\widehat{x} - \overrightarrow{b}\right) = 0.$$

Por lo tanto,

$$\|\overrightarrow{r}(\overrightarrow{x})\|^2 = \|A(\overrightarrow{x}-\widehat{x})\|^2 + \|\overrightarrow{r}(\widehat{x})\|^2$$

y como $\|A(\overrightarrow{x}-\widehat{x})\|^2 \geq 0$, se sigue que $\|\overrightarrow{r}(\widehat{x})\|^2 \leq \|\overrightarrow{r}(\overrightarrow{x})\|^2 \quad \forall\overrightarrow{x} \in \mathbb{R}^n$, es decir

$$\left\|A\widehat{x} - \overrightarrow{b}\right\|^2 = \operatorname*{Min}_{\overrightarrow{x}\in\mathbb{R}^n}\left\|A\overrightarrow{x} - \overrightarrow{b}\right\|^2.$$

A la solución $\widehat{x}$ del problema (P_a) lo denominaremos solución en mínimos cuadrados.

2. **Proyección ortogonal.**

Ponemos $A = [A_1, ..., A_n]$, con A_j la j-ésima columna de la matriz A. Sea $\overrightarrow{x}^T = (x_1, ..., x_n) \in \mathbb{R}^n$. Entonces

$$A\overrightarrow{x} = \sum_{j=1}^{n} x_j A_j \in W.$$

El ortogonal de W, por definición, es el conjunto notado $W^\perp$ y definido como sigue:

$$\begin{aligned}W^\perp &= \{\overrightarrow{y} \in \mathbb{R}^m \mid \langle\overrightarrow{y}, A\overrightarrow{x}\rangle = 0 \quad \forall\overrightarrow{x} \in \mathbb{R}^n\} = \left\{\overrightarrow{y} \in \mathbb{R}^m \mid (A\overrightarrow{x})^T\overrightarrow{y} = 0 \quad \forall\overrightarrow{x} \in \mathbb{R}^n\right\} \\ &= \left\{\overrightarrow{y} \in \mathbb{R}^m \mid \overrightarrow{x}^T A^T\overrightarrow{y} = 0 \quad \forall\overrightarrow{x} \in \mathbb{R}^m\right\}.\end{aligned}$$

Luego

$$\overrightarrow{y} \in W^\perp \Leftrightarrow \overrightarrow{y} \in \ker\left(A^T\right) = \left\{\overrightarrow{y} \in \mathbb{R}^m \mid A^T\overrightarrow{y} = 0\right\}.$$

Se tiene la siguiente suma directa $\mathbb{R}^m = W \oplus W^\perp$, y de esta, para cada $\overrightarrow{b} \in \mathbb{R}^m$, existe un único $\widehat{x} \in \mathbb{R}^n$ y $\widehat{y} \in W^\perp$ tales que $\begin{cases} A\widehat{x} \perp \widehat{y}, \\ \overrightarrow{b} = A\widehat{x} + \widehat{y}, \end{cases}$ de donde $\widehat{y} = \overrightarrow{b} - A\widehat{x}$. En la figura siguiente se ilustran el subespacio W y su ortogonal $W^\perp$, el vector $A\widehat{x} \in W$ y el vector ortogonal $\overrightarrow{y} \in W^\perp$.

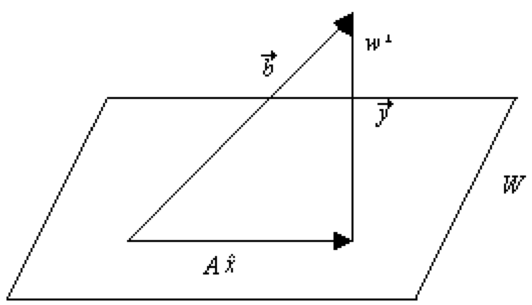

Figura 82

Sea $\overrightarrow{x} \in \mathbb{R}^n$. Entonces $A\overrightarrow{x} \in W$ y en consecuencia

$$(A\overrightarrow{x})^T \widehat{y} = 0 \Leftrightarrow (A\overrightarrow{x})^T \left(\overrightarrow{b} - A\widehat{x}\right) = 0 \Leftrightarrow \overrightarrow{x}^T A^T \left(\overrightarrow{b} - A\widehat{x}\right) = 0 \Leftrightarrow \overrightarrow{x}^T \left(A^T\overrightarrow{b} - A^T A\widehat{x}\right) = 0,$$

de donde

$$A^T A\widehat{x} = A^T \overrightarrow{b}.$$

Resulta que $\|\widehat{y}\|^2 \leq \|\overrightarrow{y}\|^2 \quad \forall \overrightarrow{y} \in W^{\perp}$, o bien

$$\left\|\overrightarrow{b} - A\widehat{x}\right\|^2 = \underset{\overrightarrow{x} \in \mathbb{R}^n}{\text{Min}} \left\|\overrightarrow{b} - A\overrightarrow{x}\right\|^2.$$

El resultado que acabamos de obtener se le conoce como proyección de un vector $\overrightarrow{b} \in \mathbb{R}^m$ con $\overrightarrow{b} \notin W$, sobre el subespacio cerrado W de $\mathbb{R}^m$. El vector $\widehat{x} \in \mathbb{R}^n$ se le conoce como solución en mínimos cuadrados.

Observaciones

1. El sistema de ecuaciones $A\overrightarrow{x} = \overrightarrow{b}$, con $m, n \in \mathbb{Z}^+$, $m \geq n$, $A = (a_{ij}) \in M_{m\times n}[\mathbb{R}]$ no nula y de rango $\mathcal{R}(A) = n$, $\overrightarrow{b}^T = (b_1, ..., b_n) \in \mathbb{R}^m$, puede ser resuelto directamente aplicando el método de factorización QR de Householder, que se expone más adelante.

2. Supóngase que $A \in M_{n\times n}[\mathbb{R}]$ matriz invertible y consideremos el sistema de ecuaciones lineales $A\overrightarrow{x} = \overrightarrow{b}$, donde $\overrightarrow{b} \in \mathbb{R}^n$ dado, cuya solución en mñinimos cuadrados es $\widehat{x} = \left(A^T A\right)^{-1} A^T \overrightarrow{b}$. Como A es invertible, se sabe que A^T es también invertible y en consecuencia existe $\left(A^T\right)^{-1}$ tal que $\left(A^T\right)^{-1} A^T = I$. Luego

$$\widehat{x} = \left(A^T A\right)^{-1} A^T \overrightarrow{b} = A^{-1} \left(A^T\right)^{-1} A^T \overrightarrow{b} = A^{-1}\overrightarrow{b},$$

que muestra que la solución en mínimos cuadrados coincide con la solución del sistema de ecuaciones $A\overrightarrow{x} = \overrightarrow{b}$.

3. En la práctica, la solución en mínimos cuadrados se calcula como sigue: del sistema de ecuaciones normal $A^T A\overrightarrow{x} = A^T\overrightarrow{b}$, se definen $B = A^T A$, $\overrightarrow{c} = A^T\overrightarrow{b}$, con lo que dicho sistema se escribe como $B\overrightarrow{x} = \overrightarrow{c}$, sistema de ecuaciones lineales que puede ser resuelto mediante el método de eliminación gaussiana, de Choleski o de factorización LU, dependiendo de las características de la matriz B.

4. Sean $m, n \in \mathbb{Z}^+$ con $m \leq n$, $V = \mathbb{R}^n$, $\{\overrightarrow{y}_1, \ldots, \overrightarrow{y}_m\} \subset \mathbb{R}^n$ tal que$\{\overrightarrow{y}_1, \ldots, \overrightarrow{y}_m\}$ linealmente independiente,

$$W = \left\{\sum_{i=1}^{m} x_i \overrightarrow{y}_i \mid x_i \in \mathbb{R},\ i = 1, \ldots, m\right\}.$$

Se tiene que W es un subespacio de $\mathbb{R}^n$ de dimensión m. Del resultado arriba establecido de la proyección de un vector $\overrightarrow{b} \in \mathbb{R}^n$ con $\overrightarrow{b} \notin W$, sobre el subespacio cerrado W, existe $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_m) \in \mathbb{R}^m$ tal que

$$\left\|\overrightarrow{b} - \sum_{i=1}^{m} \widehat{x}_i \overrightarrow{y}_i\right\|^2 = \underset{\overrightarrow{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m}{\text{Min}} \left\|\overrightarrow{b} - \sum_{i=1}^{m} x_i \overrightarrow{y}_i\right\|^2.$$

Además, se prueba que

$$\left\langle \overrightarrow{b} - \sum_{i=1}^{m} \widehat{x}_i \overrightarrow{y}_i, \overrightarrow{w} \right\rangle = 0 \quad \forall \overrightarrow{w} \in W.$$

En la figura siguiente se ilustra la solución en mínimos cuadrados.

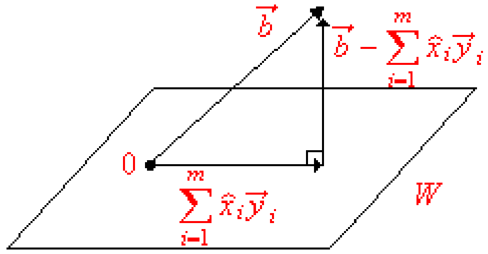

Figura 83

9.3. Método de Householder y mínimos cuadrados

Sea $A \in M_{n\times n}[\mathbb{R}]$. Supongamos que las columnas de A son ortogonales, entonces el rango de A es n, además A es invertible. En efecto, escribamos la matriz A en la forma $A = [A_1, \ldots, A_n]$, donde A_j es la j-ésima columna de A, luego $A^T = \begin{bmatrix} A_1^T \\ \vdots \\ A_n^T \end{bmatrix}$ y

$$A^T A = \begin{bmatrix} A_1^T \\ \vdots \\ A_n^T \end{bmatrix} [A_1, \ldots, A_n] = \begin{bmatrix} A_1^T A_1 & \ldots & A_1^T A_n \\ & \vdots & \\ A_n^T A_1 & \ldots & A_n^T A_n \end{bmatrix}$$

Como las columnas de A son ortogonales se sigue que $A_i^T A_j = 0$ si $i \neq j$, y $A_i^T A_i = \|A_i\|^2 = 1$.

Luego $A^T A = I$ y de esta relación resulta que $A^T = A^{-1}$.

Consideremos el sistema de ecuaciones lineales $A\overrightarrow{x} = \overrightarrow{b}$ con $\overrightarrow{b} \in \mathbb{R}^n$ dado. Entonces

$$\overrightarrow{x} = A^{-1}\overrightarrow{b} = A^T\overrightarrow{b}.$$

El siguiente resultado constituye la base del algoritmo de Householder para la resolución de sistemas de ecuaciones lineales y factorización de una matriz A en la forma QR.

Teorema 1 ***(de Householder)***
Sea $\overrightarrow{v} \in \mathbb{R}^n$ con $\overrightarrow{v} \neq 0$. Existe una matriz ortogonal H y $\alpha \in \mathbb{R}$ tales que

$$H\overrightarrow{v} = \alpha\overrightarrow{e}_1,$$

de donde $\overrightarrow{e}_1^T = (1, 0, ..., 0)$ es el primer vector de la base canónica de $\mathbb{R}^n$.

Demostración. Sea $\overrightarrow{u} \in \mathbb{R}^n$ tal que $\|\overrightarrow{u}\| = 1$. La matriz de Householder

$$H = I - 2\overrightarrow{u}\,\overrightarrow{u}^T$$

es simétrica y ortogonal.

Sea $\overrightarrow{v} \in \mathbb{R}^n$ con $\overrightarrow{v} \neq 0$. Mostremos que existe $\overrightarrow{u} \in \mathbb{R}^n$ tal que $\|\overrightarrow{u}\| = 1$ y $H\overrightarrow{v} = \alpha\overrightarrow{e}_1$. En efecto, sea $\alpha \in \mathbb{R}$ tal que $\|\overrightarrow{v}\| = |\alpha|$. Entonces

$$H\overrightarrow{v} = \left(I - 2\overrightarrow{u}\,\overrightarrow{u}^T\right)\overrightarrow{v} = \overrightarrow{v} - 2\overrightarrow{u}\,\overrightarrow{u}^T\overrightarrow{v},$$

y como $H\overrightarrow{v} = \alpha\overrightarrow{e}_1$, se sigue que

$$\begin{aligned} \overrightarrow{v} - 2\overrightarrow{u}\,\overrightarrow{u}^T\overrightarrow{v} &= \alpha\overrightarrow{e}_1 \\ 2\overrightarrow{u}\,\overrightarrow{u}^T\overrightarrow{v} &= \overrightarrow{v} - \alpha\overrightarrow{e}_1. \end{aligned}$$

Sea $p = 2\overrightarrow{u}^T\overrightarrow{v}$. Se tiene

$$p\overrightarrow{u} = \overrightarrow{v} - \alpha\overrightarrow{e}_1$$

de donde

$$\|\overrightarrow{v} - \alpha\overrightarrow{e}_1\| = \|p\overrightarrow{u}\| = |p|\,\|\overrightarrow{u}\| = |p|\,.$$

Para evitar que $p = 0$, elegimos α del modo siguiente:

$$\overrightarrow{v} - \alpha\overrightarrow{e}_1 = \begin{bmatrix} v_1 - \alpha \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

entonces

$$\alpha = \begin{cases} -\operatorname{sign}(v_1)\,\|\overrightarrow{v}\|\,, & si \quad v_1 \neq 0, \\ -\|\overrightarrow{v}\|\,, & si \quad v_1 = 0. \end{cases}$$

Supongamos primeramente que $v_1 \neq 0$. Se tiene $\alpha = -\operatorname{sign}(v_1)\,\|\overrightarrow{v}\|$, luego

$$\|\overrightarrow{v} - \alpha\overrightarrow{e}_1\|^2 = \|\overrightarrow{v} + \operatorname{sign}(v_1)\,\|\overrightarrow{v}\|\,\overrightarrow{e}_1\|^2 = (v_1 + \operatorname{sign}(v_1)\,\|\overrightarrow{v}\|)^2 + \sum_{k=2}^{n} v_k^2.$$

Si $v_1 > 0$, $\operatorname{sign}(v_1) = 1$, y

$$v_1 + \operatorname{sign}(v_1)\,\|\overrightarrow{v}\| = v_1 + \|\overrightarrow{v}\|\,.$$

Si $v_1 < 0$, $\operatorname{sign}(v_1) = -1$, y

$$v_1 + \operatorname{sign}(v_1)\,\|\overrightarrow{v}\| = v_1 - \|\overrightarrow{v}\| = -(-v_1 + \|\overrightarrow{v}\|) = -(|v_1| + \|\overrightarrow{v}\|)\,.$$

Por lo tanto

$$\begin{aligned} \|\overrightarrow{v} - \alpha\overrightarrow{e}_1\| &= (|v_1| + \|\overrightarrow{v}\|)^2 + \sum_{k=2}^{n} v_k^2 = v_1^2 + 2|v_1|\,\|\overrightarrow{v}\| + \|\overrightarrow{v}\|^2 + \sum_{k=2}^{n} v_k^2 \\ &= 2|v_1|\,\|\overrightarrow{v}\| + \|\overrightarrow{v}\|^2 + \sum_{k=1}^{n} v_k^2 = 2|v_1|\,\|\overrightarrow{v}\| + 2\,\|\overrightarrow{v}\|^2\,. \end{aligned}$$

Si $v_1 = 0$, $\quad \alpha = -\|\overrightarrow{v}\|$. Entonces

$$\|\overrightarrow{v} - \alpha\overrightarrow{e}_1\|^2 = \|\overrightarrow{v} + \|\overrightarrow{v}\|\,\overrightarrow{e}_1\|^2 = \|\overrightarrow{v}\|^2 + \sum_{k=2}^{n} v_k^2 = \|\overrightarrow{v}\|^2 + \sum_{k=1}^{n} v_k^2 = 2\,\|\overrightarrow{v}\|^2\,.$$

Consecuentemente, de la igualdad $p\overrightarrow{u} = \overrightarrow{v} - \alpha\overrightarrow{e}_1$, se sigue que

$$\begin{aligned} \overrightarrow{u} &= \frac{\overrightarrow{v} - \alpha\overrightarrow{e}_1}{p} = \frac{\overrightarrow{v} - \alpha\overrightarrow{e}_1}{\|\overrightarrow{v} - \alpha\overrightarrow{e}_1\|} = \frac{\overrightarrow{v} - \alpha\overrightarrow{e}_1}{\left(2\,\|\overrightarrow{v}\|^2 + 2|v_1|\,\|\overrightarrow{v}\|\right)^{\frac{1}{2}}} \quad si \quad v_1 \neq 0, \\ \overrightarrow{u} &= \frac{\overrightarrow{v} - \alpha\overrightarrow{e}_1}{\sqrt{2}\,\|\overrightarrow{v}\|} \quad si \quad v_1 = 0. \end{aligned}$$

La matriz H queda definida como

$$H = I - 2\overrightarrow{u}\,\overrightarrow{u}^T = I - \frac{1}{\|\overrightarrow{v}\|^2 + |v_1|\,\|\overrightarrow{v}\|}\,(\overrightarrow{v} - \alpha\overrightarrow{e}_1)\,(\overrightarrow{v} - \alpha\overrightarrow{e}_1)^T \text{ si } v_1 \neq 0,$$

y

$$H = I - \frac{1}{\|\overrightarrow{v}\|^2}\left(\overrightarrow{v} - \alpha\overrightarrow{e}_1\right)\left(\overrightarrow{v} - \alpha\overrightarrow{e}_1\right)^T, \text{ si } v_1 = 0.$$

Si ponemos

$$\begin{array}{rcl} r & = & \left\{ \begin{array}{c} \|\overrightarrow{v}\|^2 + |v_1|\,\|\overrightarrow{v}\|, \quad si \quad v_1 \neq 0, \\ 2\,\|\overrightarrow{v}\|^2, \quad si \quad v_1 = 0, \end{array} \right. \\ \overrightarrow{w} & = & \overrightarrow{v} - \alpha\,\overrightarrow{e}_1, \end{array}$$

entonces

$$H = I - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T.$$

■

Observación

i) Se puede escribir $H = I - 2\dfrac{(s\overrightarrow{u})\,(s\overrightarrow{u})^t}{s^2}$; el cálculo explícito de $\overrightarrow{u}$ no es necesario.

Ponemos $r = \dfrac{s^2}{2}$ y $\overrightarrow{w} = s\overrightarrow{u}$ entonces $H = I - \dfrac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T$ donde $\overrightarrow{w}$ y r se determinan por el sistema:

$$\left\{ \begin{array}{l} |\alpha| = \|\overrightarrow{v}\|, \\ \overrightarrow{w} = \overrightarrow{v} - \alpha\,\overrightarrow{e}_1, \\ r = \alpha\,(\alpha - v_1). \end{array} \right.$$

ii) El signo de α se tomará el opuesto al de v_1

iii) Para calcular $H\overrightarrow{v}$ podemos proceder de la siguiente manera:

$$H\overrightarrow{v} = \left(I - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T\right)\overrightarrow{v} = \overrightarrow{v} - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T\overrightarrow{v} = \overrightarrow{v} - \frac{1}{r}\left(\overrightarrow{w}^T\overrightarrow{v}\right)\overrightarrow{w},$$

cuyos cálculos sucesivos son:

$$\begin{array}{rcl} q & = & \overrightarrow{w}^T\overrightarrow{v}, \\ p & = & \dfrac{q}{r}, \\ H\overrightarrow{v} & = & \overrightarrow{v} - p\overrightarrow{w}. \end{array}$$

Algoritmo de Householder

Sea $A \in M_{n\times n}[\mathbb{R}]$ invertible, $\overrightarrow{b} \in \mathbb{R}^n$ dado y consideramos el sistema de ecuaciones $A\overrightarrow{x} = \overrightarrow{b}$.

Aplicaremos el lema 2, m veces para triangularizar la matriz A. Ponemos $A^{(1)} = A$, $\overrightarrow{b}^{(1)} = \overrightarrow{b}$. Generaremos $m-1$ ecuaciones equivalentes a la ecuación original propuesta: $A^{(k)}\overrightarrow{x} = \overrightarrow{b}^{(k)}$ $k = 2, \ldots, n$ donde $A^{(k)}$ y $\overrightarrow{b}^{(k)}$ son de la forma:

$$A^{(k)} = \left[\begin{array}{cccc|ccc} a_{11}^{(2)} & \cdots & \cdots & a_{1k-1}^{(2)} & a_{1k}^{(2)} & \cdots & a_{1n}^{(2)} \\ & a_{22}^{(3)} & \cdots & a_{2k-1}^{(3)} & \vdots & & \vdots \\ & & \ddots & \vdots & \vdots & & \vdots \\ 0 & & & a_{k-1,k-1}^{(k)} & a_{k-1,k}^{(k)} & \cdots & a_{k-1,n}^{(k)} \\ \hline & & & & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ & 0 & & & \vdots & & \vdots \\ & & & & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} \end{array}\right], \quad \overrightarrow{b}^{(k)} = \left[\begin{array}{c} b_1^{(2)} \\ \vdots \\ b_{k-1}^{(k)} \\ \hline b_k^{(k)} \\ \vdots \\ b_n^{(k)} \end{array}\right]$$

que de manera abreviada se escribe:

$$A^{(k)} = \left[\begin{array}{c|c} A_{11}^{(k)} & A_{12}^{(k)} \\ \hline 0 & A_{22}^{(k)} \end{array}\right], \quad \overrightarrow{b}^{(k)} = \left[\begin{array}{c} \overrightarrow{c}^{(k)} \\ \hline \overrightarrow{d}^{(k)} \end{array}\right]$$

Para pasar de $A^{(k)}$ a $A^{(k+1)}$ y de $\overrightarrow{b}^{(k)}$ a $\overrightarrow{b}^{(k+1)}$, buscamos una matriz ortogonal elemental (matrix de Householder) $H^{(k)}$ tal que $H^{(k)}A^{(k)}$ es una matriz cuyas primeras k columnas forman una matriz triangular de la forma de $A_{11}^{(k)}$.

Tomamos un vector $\overrightarrow{u}^{(k)}$ tal que $u_1^{(k)} = u_2^{(k)} = \cdots = u_{k-1}^{(k)} = 0$ y notamos con $\widetilde{u}^{(k)}$ el vector $(u_k, \ldots, u_n) \in \mathbb{R}^{n-k+1}$ y

$$\widetilde{H}^{(k)} = I_{n-k+1} - 2\widetilde{u}^{(k)}\widetilde{u}^{(k)T}.$$

La matriz $H\,(k) = I - 2u^{(k)}u^{(k)T}$ es tal que $H^{(k)}A^{(k)}$ deja fijas las $k-1$ filas y columnas de $A^{(k)}$, además $H^{(k)}\overrightarrow{b}^{(k)}$ no modifica las $k-1$ primeros elementos de $\overrightarrow{b}^{(k)}$. Luego $H^{(k)}A^{(k)}$ se escribe:

$$\begin{aligned} H^{(k)}A^{(k)} &= \left[\begin{array}{c|c} I_{k-1} & 0 \\ \hline 0 & \widetilde{H}^{(k)} \end{array}\right] \left[\begin{array}{c|c} A_{11}^{(k)} & A_{12}^{(k)} \\ \hline 0 & A_{22}^{(k)} \end{array}\right] = \left[\begin{array}{c|c} A_{11}^{(k)} & A_{12}^{(k)} \\ \hline 0 & \widetilde{H}^{(k)}A_{22}^{(k)} \end{array}\right] \\ H^{(k)}\overrightarrow{b}^{(k)} &= \left[\begin{array}{c|c} I_{k-1} & 0 \\ \hline 0 & \widetilde{H}^{(k)} \end{array}\right] \left[\begin{array}{c} \overrightarrow{c}^{(k)} \\ \hline \overrightarrow{d}^{(k)} \end{array}\right] = \left[\begin{array}{c} \overrightarrow{c}^{(k)} \\ \hline \widetilde{H}^{(k)}\overrightarrow{d}^{(k)} \end{array}\right]. \end{aligned}$$

Como $\det\left(A^{(k)}\right) = \left[\prod_{i=1}^{k-1} a_{ii}^{(i+1)}\right] \det\left(A_{22}^{(k)}\right) \neq 0$, los elementos de la primera columna de $A_{22}^{(k)}$ no son nulos, consecuentemente se puede aplicar el lema2 con $\overrightarrow{v} = \left(a_{kk}^{(k)}, \ldots, a_{nk}^{(k)}\right) \in \mathbb{R}^{n-k+1}$ para anular todos salvo el primero; cuyo algoritmo de cálculo es:

$$\begin{aligned} a_{kk}^{(k-1)} &= -sign\left(a_{kk}^{(k)}\right)\left(\sum_{i=k}^{n}\left[a_{ik}^{(k)}\right]^2\right)^{\frac{1}{2}}, \\ r^{(k)} &= a_{kk}^{(k+1)}\left(a_{kk}^{(k+1)} - a_{kk}^{(k)}\right), \\ w_k^{(k)} &= a_{kk}^{(k)} - a_{kk}^{(k+1)}, \\ w_i^{(k)} &= a_{ik}^{(k)} \quad i = k+1, \ldots, n \end{aligned}$$

que permiten determinar $H^{(k)}$ cuyo cálculo de $\widetilde{H}^{(k)}A_{22}^{(k)}$ es el siguiente:

$$\begin{aligned} q_j^{(k)} &= \sum_{i=k}^{n} w_i^{(k)}a_{ij}^{(k)}, \\ p_j^{(k)} &= \frac{q_i^{(k)}}{r^{(k)}} \quad j = k+1, \ldots, n. \\ a_{ij}^{(k+1)} &= a_{ij}^{(k)} - p_j^{(k)}w_i, \quad i = k+1, \ldots, n \end{aligned}$$

Para el cálculo de $\widetilde{H}^{(k)}\overrightarrow{b}^{(k)}$ se emplea el siguiente algoritmo;

$$\begin{aligned} q^{(k)} &= \sum_{i=k}^{n} w_i^{(k)}b_i^{(k)}, \\ p^{(k)} &= \frac{q^{(k)}}{r^{(k)}}, \\ b_i^{(k+1)} &= b_i^{(k)} - p^{(k)}w_i^{(k)} \quad i = k, \ldots, n. \end{aligned}$$

Finalmente, ponemos $H^{(k)}A^{(k)} = A^{(k+1)}$ y $H^{(k)}\overrightarrow{b}^{(k)} = \overrightarrow{b}^{(k+1)}$.

La matriz $A^{(n)}$ es una matriz triangular superior que permite resolver fáclmente el sistema: $A^{(n)}\overrightarrow{x} = \overrightarrow{b}^{(n)}$.

Observación

1. El método de Housholder permite calcular $\det(A)$. Pués de la factorización siguiente:

$$A^{(n)} = H^{(n-1)}H^{(n-2)}\dots H^{(1)}A^{(1)},$$

y como $A^{(1)} = A$ y $\det\left(H^{(k)}\right) = -1 \quad \forall k = 1, \cdots, n$, se sigue que

$$\det(A) = (-1)^{n-1}\det\left(A^{(n)}\right) = (-1)^{n-1}\prod_{i=1}^{n} a_{ii}^{(n)}.$$

2. Para mejorar la estabilidad del método, podemos proceder como sigue: en la k-ésima etapa, en lugar de transformar la k-ésima columna de $A^{(k)}$, se elige entre las últimas columnas de $A^{(k)}$ aquel elemento que hace: $a_{ij}^{(k)} = \sum\limits_{i=k}^{n}\left(a_{ij}^{(k)}\right)^2$ máximo. Se permuta aquella con la k-ésima columna, si en $k = L$ se alcanza tal máximo, entonces

$$a_{kk}^{(k+1)} = -sign\left(a_{kk}^{(k)}\right)\sqrt{a_L^{(k)}}.$$

Ejemplos

1. Sea $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Construyamos la matriz de Householder y factoremos la matriz en la forma $A = QR$ con Q una matriz ortogonal y R una matriz triangular superior.

   Sea $\overrightarrow{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, se tiene $\alpha = -1$,

$$\overrightarrow{w} = \overrightarrow{v} - \alpha\overrightarrow{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

   Además $r = \alpha(\alpha - v_1) = -1(-1-2) = 2$. Luego

$$\begin{aligned} H &= I - \frac{1}{r}\overrightarrow{w}\overrightarrow{w}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 2 \\ 0 \end{bmatrix}(2,0) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \\ A^{(1)} &= HA = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} = R. \end{aligned}$$

   Obtenemos $A = H^T R = QR \quad$ con $Q = H^T = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$.

2. Sea $A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Construyamos la matriz de Householder y factoremos la matriz en la forma $A = QR$ con Q una matriz ortogonal y R una matriz triangular superior.

   Sea $\overrightarrow{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, se tiene $v_1 = 1$, $\alpha = -1$,

$$\overrightarrow{w} = \overrightarrow{v} - \alpha\overrightarrow{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}.$$

Además $r=\alpha(\alpha-v_1)=-1(-1-1)=2$. Luego

$$\begin{aligned} H^{(1)} &= I-\frac{1}{r}\overrightarrow{w}\overrightarrow{w}^T=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}-\frac{1}{2}\begin{bmatrix}2\\0\\0\end{bmatrix}(2,0,0)=\begin{bmatrix}-1&0&0\\0&1&0\\0&0&1\end{bmatrix}\\ A^{(1)} &= H^{(1)}A=\begin{bmatrix}-1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}=\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}. \end{aligned}$$

Segunda etapa

Definimos $\overrightarrow{v}=\begin{bmatrix}1\\0\end{bmatrix}$. Procediendo tal como en la primera etapa obtenemos $\widetilde{H}^{(2)}=\begin{bmatrix}-1&0\\0&1\end{bmatrix}$,

$$\begin{aligned} H^{(2)} &= \left[\begin{array}{c|c}I&0\\\hline 0&\widetilde{H}^{(2)}\end{array}\right]=\begin{bmatrix}1&0&0\\0&-1&0\\0&0&1\end{bmatrix},\\ A^{(2)} &= H^{(2)}A^{(1)}=\begin{bmatrix}1&0&0\\0&-1&0\\0&0&1\end{bmatrix}\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}=\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}=R. \end{aligned}$$

De la definición de las matrices precedentes obtenemos $R=H^{(2)}A^{(1)}=H^{(2)}H^{(1)}A\Rightarrow A=QR$ con $Q=\left(H^{(2)}H^{(1)}\right)^T$, y

$$H^{(2)}H^{(1)}=\begin{bmatrix}1&0&0\\0&-1&0\\0&0&1\end{bmatrix}\begin{bmatrix}-1&0&0\\0&1&0\\0&0&1\end{bmatrix}=\begin{bmatrix}-1&0&0\\0&-1&0\\0&0&1\end{bmatrix}.$$

3. Sea $A=\begin{bmatrix}0&0&0\\0&0&0\\1&0&0\end{bmatrix}$. Construyamos la matriz de Householder y factoremos la matriz en la forma $A=QR$ con Q una matriz ortogonal y R una matriz triangular superior.

Sea $\overrightarrow{v}=\begin{bmatrix}0\\0\\1\end{bmatrix}$, se tiene $\|v\|=1$, $v_1=0$, $\alpha=-1$,

$$\overrightarrow{w}=\overrightarrow{v}-\alpha\overrightarrow{e}_1=\begin{bmatrix}0\\0\\1\end{bmatrix}+\begin{bmatrix}1\\0\\0\end{bmatrix}=\begin{bmatrix}1\\0\\1\end{bmatrix}.$$

Además $r=\alpha(\alpha-v_1)=1$. Luego

$$\begin{aligned} H^{(1)} &= I-\frac{1}{r}\overrightarrow{w}\overrightarrow{w}^T=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}-\begin{bmatrix}1\\0\\1\end{bmatrix}(1,0,1)=\begin{bmatrix}0&0&-1\\0&1&0\\-1&0&0\end{bmatrix}\\ A^{(2)} &= H^{(1)}A=\begin{bmatrix}0&0&-1\\0&1&0\\-1&0&0\end{bmatrix}\begin{bmatrix}0&0&0\\0&0&0\\1&0&0\end{bmatrix}=\begin{bmatrix}-1&0&0\\0&0&0\\0&0&0\end{bmatrix}. \end{aligned}$$

Segunda etapa

Definimos $\overrightarrow{v}=\begin{bmatrix}1\\0\end{bmatrix}$. Procediendo tal como en la primera etapa obtenemos

$$\begin{aligned} H^{(2)} &= \begin{bmatrix}1&0&0\\0&-1&0\\0&0&1\end{bmatrix},\\ A^{(3)} &= H^{(2)}A^{(2)}=\begin{bmatrix}1&0&0\\0&-1&0\\0&0&1\end{bmatrix}\begin{bmatrix}-1&0&0\\0&0&0\\0&0&0\end{bmatrix}=\begin{bmatrix}-1&0&0\\0&0&0\\0&0&0\end{bmatrix}=R. \end{aligned}$$

Consecuentemente $R = H^{(2)}A^{(2)} = H^{(2)}H^{(1)}A \Rightarrow A = QR$ con $Q = \left(H^{(2)}H^{(1)}\right)^T$, y

$$H^{(2)}H^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

4. Sea $A = \begin{bmatrix} 2 & -1 & 4 \\ 2 & 0 & -1 \\ 1 & 3 & 1 \end{bmatrix}$. Construyamos una matriz triangular superior aplicando el algoritmo que acabamos de describir.

i) Sea $\overrightarrow{v} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ obtenido como la primera columna de la matriz A, entonces $\|\overrightarrow{v}\| = \sqrt{9} = 3$. Determinemos α, $\overrightarrow{w}$ y r. Tenemos

$$\begin{aligned} \alpha &= |\alpha|\, sign\,(\alpha) = \|\overrightarrow{v}\|\, sign\,(\alpha) = -\|\overrightarrow{v}\| = -3, \qquad \text{pués } sign\,(\alpha) = -sign\,(v_1) = -1, \\ \overrightarrow{w} &= \overrightarrow{v} - \alpha\,\overrightarrow{e}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} - (-3) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}, \\ r &= \alpha\,(\alpha - v_1) = -3\,(-3-2) = 15, \end{aligned}$$

y la matriz de Householder está definida como sigue:

$$H^{(1)} = I - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{15} \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix} (5, 2, 1) = \begin{bmatrix} -\frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ -\frac{2}{3} & \frac{11}{15} & -\frac{2}{15} \\ -\frac{1}{3} & -\frac{2}{15} & \frac{14}{15} \end{bmatrix},$$

entonces

$$A^{(2)} = H^{(1)}A^{(1)} = \begin{bmatrix} -\frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ -\frac{2}{3} & \frac{11}{15} & -\frac{2}{15} \\ -\frac{1}{3} & -\frac{2}{15} & \frac{14}{15} \end{bmatrix} \begin{bmatrix} 2 & -1 & 4 \\ 2 & 0 & -1 \\ 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} -3 & -\frac{1}{3} & -\frac{7}{3} \\ 0 & \frac{4}{15} & -\frac{53}{15} \\ 0 & \frac{47}{15} & -\frac{4}{15} \end{bmatrix}.$$

ii) Continuado con el algoritmo, elegimos el vector $\overrightarrow{v} = \begin{bmatrix} \frac{4}{15} \\ \frac{47}{15} \end{bmatrix}$ obtenido de la segunda columna de la matriz $A^{(2)}$, luego $\|\overrightarrow{v}\| = \frac{1}{15}\sqrt{2225} = \frac{1}{3}\sqrt{89} \simeq 3{,}1446603777$. Determinemos α, $\overrightarrow{w}$ y r. Tenemos

$$\begin{aligned} \alpha &= -\frac{1}{3}\sqrt{89} = -3{,}1446603777, \\ \overrightarrow{w} &= \begin{bmatrix} \frac{4}{15} \\ \frac{47}{15} \end{bmatrix} + 3{,}1446603777 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3{,}4113270440 \\ 3{,}1333333333 \end{bmatrix}, \\ r &= \alpha\,(\alpha - v_1) = -3{,}1446603777 \left(-3{,}1446603777 - \frac{4}{15}\right) = 10{,}7274649895, \\ \frac{1}{r} &= 0{,}09322186682. \end{aligned}$$

En esta etapa la matriz de Householder está definida como sigue:

$$\begin{aligned}
H^{(2)} &= I - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&\quad -0{,}09322186682 \begin{bmatrix} 0 \\ 3{,}4113270440 \\ 3{,}1333333333 \end{bmatrix} (0, 3{,}4113270440, 3{,}1333333333) \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0{,}0847998304 & -0{,}9963980072 \\ 0 & -0{,}9963980072 & 0{,}0847998304 \end{bmatrix}.
\end{aligned}$$

luego, calculamos la matriz $A^{(3)} = H^{(2)}A^{(2)}$, tenemos

$$\begin{aligned}
A^{(3)} &= H^{(2)}A^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0{,}0847998304 & -0{,}9963980072 \\ 0 & -0{,}9963980072 & 0{,}0847998304 \end{bmatrix} \begin{bmatrix} -3 & -\frac{1}{3} & -\frac{7}{3} \\ 0 & \frac{4}{15} & -\frac{53}{15} \\ 0 & \frac{47}{15} & -\frac{4}{15} \end{bmatrix} \\
&= \begin{bmatrix} -3 & -\frac{1}{3} & -\frac{7}{3} \\ 0 & -3{,}1446603774 & 0{,}56533322027 \\ 0 & 0 & 3{,}4979930040 \end{bmatrix},
\end{aligned}$$

que es la matriz triangular superior buscada. Note que

$$A^{(3)} = H^{(2)}A^{(2)} = H^{(2)}H^{(1)}A^{(1)} = H^{(2)}H^{(1)}A = QA$$

con $Q = H^{(2)}H^{(1)}$ matriz ortogonal. Ponemos $R = A^{(3)}$ y se tiene $A = Q^T R$.

5. Sean $A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 1 & 2 & 4 & 1 \\ 2 & 4 & 0 & 2 \\ 3 & 1 & 2 & 14 \end{bmatrix}$, $\overrightarrow{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$. Apliquemos el método de Householder para resolver el sistema de ecuaciones $A\overrightarrow{x} = \overrightarrow{b}$. Para el efecto, primeramente construimos una matriz triangular superior $A^{(4)} = H^{(3)}H^{(2)}H^{(1)}A$, a continuación obtenemos el vector $\overrightarrow{b}^{(4)} = H^{(3)}H^{(2)}H^{(1)}\overrightarrow{b}$. Entonces $A\overrightarrow{x} = \overrightarrow{b} \iff A^{(4)}\overrightarrow{x} = \overrightarrow{b}^{(4)}$, este último sistema de ecuaciones lineales es triangular superior.

i) De acuerdo al algoritmo antes descrito, elegimos $\overrightarrow{v} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \end{bmatrix}$ que es la primera columna de la matriz A, luego $\|\overrightarrow{v}\| = \sqrt{15} = 3{,}8729833462$. A continuación determinamos $\alpha = -3{,}8729833462$, y

$$r = \alpha\,(\alpha - v_1) = -\sqrt{15}\left(-\sqrt{15} - 1\right) = 18{,}8729833462, \qquad \frac{1}{r} = 0{,}05298579259,$$

y defiinos el vector $\overrightarrow{w}$ como sigue:

$$\overrightarrow{w} = \overrightarrow{v} - \alpha\,\overrightarrow{e}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \end{bmatrix} + \sqrt{15}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1+\sqrt{15} \\ 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4{,}8729833462 \\ 1 \\ 2 \\ 3 \end{bmatrix}.$$

La matriz de Householder $H^{(1)}$ está definida como

$$\begin{aligned} H^{(1)} &= I - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - 0{,}05298579259 \begin{bmatrix} 4{,}8729833462 \\ 1 \\ 2 \\ 3 \end{bmatrix} (4{,}8729833462, 1, 2, 3) \\ &= \begin{bmatrix} -0{,}2581988897 & -2{,}5819889747 & -0{,}5163977795 & -0{,}7745966692 \\ -2{,}5819889747 & 0{,}9470142064 & -0{,}1059715872 & -0{,}1589573807 \\ -0{,}5163977795 & -0{,}1059715872 & -0{,}7880568256 & -0{,}3179147615 \\ -0{,}7745966692 & -0{,}1589573807 & -0{,}3179147615 & 0{,}52312785769 \end{bmatrix}. \end{aligned}$$

y en consecuencia

$$A^{(2)} = H^{(1)}A^{(1)} = \begin{bmatrix} -3{,}8729833462 & -3{,}6147844565 & -3{,}0983866769 & -12{,}9099444873 \\ 0{,}0 & 1{,}052987936 & 2{,}9537442846 & -2{,}2649289679 \\ 0{,}0 & 2{,}1059715872 & -2{,}0925114308 & -4{,}5298579359 \\ 0{,}0 & -1{,}8410426192 & -1{,}1387671462 & 4{,}2052130962 \end{bmatrix}.$$

ii) Tomando en consideración la segunda columna de la matriz $A^{(2)}$, elegimos el vector $\overrightarrow{v} = \begin{bmatrix} 0 \\ 1{,}052987936 \\ 2{,}1059715872 \\ -1{,}8410426192 \end{bmatrix}$ y calculamos su norma, tenemos $\|\overrightarrow{v}\| = 2{,}9888682362 \simeq 3$ con lo que $\alpha = -2{,}9888682362$. Calculemos r y r^{-1} :

$$\begin{aligned} r &= \alpha\,(\alpha - v_2) = -2{,}9888682362\,(-2{,}9888682362 - 1{,}052987936) = 12{,}08056912496, \\ \frac{1}{r} &= 0{,}082777557055. \end{aligned}$$

Determinamos el vector $\overrightarrow{w}$:

$$\overrightarrow{w} = \overrightarrow{v} - \alpha\,\overrightarrow{e}_2 = \begin{bmatrix} 0 \\ 1{,}052987936 \\ 2{,}1059715872 \\ -1{,}8410426192 \end{bmatrix} + 2{,}9888682362 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0{,}0 \\ 4{,}04185402978 \\ 2{,}1059715872 \\ -1{,}8410426192 \end{bmatrix},$$

y con este pasamos al cálculo de la matriz de Householder $H^{(2)}$:

$$H^{(2)} = I - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T = \begin{bmatrix} 1{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\ 0{,}0 & -0{,}3523025139 & -0{,}7046050279 & 0{,}6159664708 \\ 0{,}0 & -0{,}7046050279 & 0{,}632871905277 & 0{,}32094377398 \\ 0{,}0 & 0{,}6159664708 & 0{,}3209437773988 & 0{,}71943060871 \end{bmatrix}$$

entonces

$$A^{(3)} = H^{(2)}A^{(2)} = \begin{bmatrix} -3{,}8729833462 & -3{,}6147844565 & -3{,}0983866769 & -12{,}9099444873 \\ 0{,}0 & -2{,}9888682362 & -0{,}2676598420 & 6{,}5799711169 \\ 0{,}0 & 0 & -3{,}7709949958 & 0{,}07869747780 \\ 0{,}0 & 0 & 0{,}32956498588 & 0{,}176409012875 \end{bmatrix}$$

iii) Continuando con el método de Householder, de la tercera columna de la matriz $A^{(3)}$, elegimos el vector $\overrightarrow{v} = \begin{bmatrix} 0{,}0 \\ 0{,}0 \\ -3{,}7709949958 \\ 0{,}32956498588 \end{bmatrix}$ y calculamos su norma, obtenemos $\|\overrightarrow{v}\| = 3{,}78528178726$ y con este valor tenemos $\alpha = 3{,}78528178726$. Calculemos r y r^{-1} :

$$\begin{aligned} r &= \alpha\,(\alpha - v_3) = 3{,}78528178726\,(3{,}78528178726 + 3{,}7709949958) = 28{,}6026368867 \\ \frac{1}{r} &= 0{,}04596181153. \end{aligned}$$

Calculamos el vector correspondiente $\overrightarrow{w}$:

$$\overrightarrow{w} = v - \alpha \overrightarrow{e}_3 = \begin{bmatrix} 0{,}0 \\ 0{,}0 \\ -3{,}7709949958 \\ 0{,}32956498588 \end{bmatrix} - 3{,}78528178726 \begin{bmatrix} 0{,}0 \\ 0{,}0 \\ 1{,}0 \\ 0{,}0 \end{bmatrix} = \begin{bmatrix} 0{,}0 \\ 0{,}0 \\ -7{,}5562767831 \\ 0{,}3285649859 \end{bmatrix}$$

y la matriz de Householder queda definida como

$$H^{(3)} = I - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T = \begin{bmatrix} 1{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\ 0{,}0 & 1{,}0 & 0{,}0 & 0{,}0 \\ 0{,}0 & 0{,}0 & -0{,}99622569938 & 0{,}086800667518 \\ 0{,}0 & 0{,}0 & 0{,}086800667518 & 0{,}99622569938 \end{bmatrix}.$$

Con esta matriz, calculamos la matriz $A^{(4)}$ siguiente:

$$A^{(4)} = H^{(3)}A^{(3)} = \begin{bmatrix} -3{,}8729833462 & -3{,}6147844565 & -3{,}0983866769 & -12{,}9099444873 \\ 0{,}0 & -2{,}9888682362 & -0{,}2676598420 & 6{,}5799711169 \\ 0{,}0 & 0 & 3{,}7852817872 & -0{,}06308802978 \\ 0{,}0 & 0 & 0 & 0{,}1825741858 \end{bmatrix}.$$

Puesto que $\overrightarrow{b}^{(4)} = H^{(3)}H^{(2)}H^{(1)}\overrightarrow{b}$, se sigue que

$$\overrightarrow{b}^{(2)} = H^{(1)}\overrightarrow{b} = \begin{bmatrix} 0{,}516397779494 \\ -0{,}09924150898 \\ -0{,}19848301796 \\ -1{,}29772452694 \end{bmatrix}, \qquad \overrightarrow{b}^{(3)} = H^{(2)}\overrightarrow{b}^{(2)} = \begin{bmatrix} 0{,}516397779494 \\ -0{,}6245396314 \\ -0{,}4721848668 \\ -1{,}0584544077 \end{bmatrix},$$

$$\overrightarrow{b}^{(4)} = H^{(3)}\overrightarrow{b}^{(3)} = \begin{bmatrix} 0{,}516397779494 \\ -0{,}6245396314 \\ 0{,}37852817872 \\ -1{,}0954451150 \end{bmatrix}.$$

Resolvamos el sistema de ecuaciones $A\overrightarrow{x} = \overrightarrow{b}$. Tenemos $A^{(4)}\overrightarrow{x} = \overrightarrow{b}^{(4)}$:

$$\begin{bmatrix} -3{,}8729833462 & -3{,}6147844565 & -3{,}0983866769 & -12{,}9099444873 \\ 0{,}0 & -2{,}9888682362 & -0{,}2676598420 & 6{,}5799711169 \\ 0{,}0 & 0 & 3{,}7852817872 & -0{,}06308802978 \\ 0{,}0 & 0 & 0 & 0{,}1825741858 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$
$$= \begin{bmatrix} 0{,}516397779494 \\ -0{,}6245396314 \\ 0{,}37852817872 \\ -1{,}0954451150 \end{bmatrix}$$

y la solución de este sistema triangular superior es $\overrightarrow{x} = \begin{bmatrix} 32 \\ -13 \\ 0 \\ -6 \end{bmatrix}$.

9.3.1. Número de operaciones elementales

Para pasar de $A^{(k)}$ a $A^{(k+1)}$ se requiere las siguientes operaciones elementales:

$n-k+1$	elevaciones al cuadrado (multiplicaciones),
$n-k$	adiciones,
1	raíz cuadrada para calcular $a_{kk}^{(k+1)}$,
1	sustracción,
1	multiplicación para calcular $r^{(k)}$,
1	sustracción para calcular $w_i^{(k)}, \quad i=k,\ldots,n$,
$n-k+1$	multiplicaciones para calcular $q_j^{(k)}$,
$n-k$	adiciones,
1	división para calcular $q_i^{(k)}$,
$n-k+1$	multiplicaciones para calcular $a_{ij}^{(k+1)}, \quad i=k,\ldots,n$,
$n-k+1$	sustracciones

Para pasar de $\overrightarrow{b}^{(k)}$ a $\overrightarrow{b}^{(k+1)}$, se requieren de:

$n-k+1$	multiplicaciones para calcular $q^{(k)}$,
$n-k$	adiciones,
1	división para calcular $p^{(k)}$,
$n-k+1$	multiplicaciones para calcular $b_i^{(k+1)}, \quad i=k,\ldots,n$,
$n-k+1$	sustracciones,

Para pasar del sistema $A^{(k)}\overrightarrow{x} = \overrightarrow{b}^{(k)}$ al sistema $A^{(k+1)}\overrightarrow{x} = \overrightarrow{b}^{(k+1)}$ se requieren de:

$2n^2-4nk+5n+2k^2-5k+4$	multiplicaciones,
$2n^2-4nk+4n+2k^2-4k+3$	adiciones o sustracciones,
$n-k+1$	divisiones,
1	raíz cuadrada,

Por lo tanto, para la triangularización de A se necesitan

$$\begin{array}{rl}
(n-1)\left[\dfrac{2n^2}{3}+\dfrac{13n}{6}+4\right] & \text{multiplicaciones,}\\
(n-1)\left[\dfrac{2n^2}{3}+\dfrac{5n}{3}+3\right] & \text{adiciones o sustracciones,}\\
(n-1)\,\dfrac{n+2}{2} & \text{divisiones,}\\
n-1 & \text{raíces cuadradas,}
\end{array}$$

Para la resolución del sistema triangular $A^{(n)}\overrightarrow{x} = \overrightarrow{b}^{(n)}$ se necesitan de n^2 operaciones elementales.

El número total de operaciones en el método de Householder es:

$$T_H = \frac{4n^3}{3} + 4n^2 + \frac{14n}{3} - 9.$$

Metodo de Householder

El método de Householder puede aplicarse para resolver problemas de aproximación con el método de mínimos cuadrados.

Sea $A \in M_{m\times n}[\mathbb{R}]$ con $m \geq n$ y $\overrightarrow{b} \in \mathbb{R}^m$. Consideramos el sistema de ecuaciones $A\overrightarrow{x} = \overrightarrow{b}$ y el problema en mínimos cuadrados:

$$\text{hallar } \widehat{x} \in \mathbb{R}^n \text{ tal que } \left\| A\widehat{x} - \overrightarrow{b} \right\| = \min_{\overrightarrow{x}\in\mathbb{R}^n} \left\| A\overrightarrow{x} - \overrightarrow{b} \right\|.$$

Apliquemos el método de ortogonalización de House holder para resolver el sistema de ecuaciones $A\overrightarrow{x} = \overrightarrow{b}$.

Ponemos $A^{(0)} = A$ y $\overrightarrow{b}^{(0)} = \overrightarrow{b}$. Mediante el método de Householder construimos matrices ortogonales $Q_i \in M_{m\times m}[\mathbb{R}]$, matrices $A^{(i)}$ tales que $A^{(i)} = Q_i A^{(i-1)}$ y $\overrightarrow{b}^{(i)} = Q_i \overrightarrow{b}^{(i-1)}$ $\quad i = 1, \ldots, n$.

Sea $Q = Q_{n-1}Q_{n-2}\ldots Q_1$ entonces $QA^{(n)} = \begin{bmatrix} R \\ 0 \end{bmatrix}$, donde $R = \begin{bmatrix} r_{11} & \ldots & r_{1n} \\ & \ddots & \vdots \\ 0 & & r_{nn} \end{bmatrix}$, $0 \in M_{m\times m}[\mathbb{R}]$.

Ponemos $\overrightarrow{h} = \overrightarrow{b}^{(n)} = Q\overrightarrow{b}$ y $\overrightarrow{h} = \begin{bmatrix} \overrightarrow{h}_1 \\ \overrightarrow{h}_2 \end{bmatrix}$ con $\overrightarrow{h}_1 \in \mathbb{R}^n$ y $\overrightarrow{h}_2 \in \mathbb{R}^{m-n}$.

Puesto que la matriz Q es ortogonal, se tiene $\|Q\overrightarrow{u}\| = \|\overrightarrow{u}\| \quad \forall \overrightarrow{u} \in \mathbb{R}^m$ y en consecuencia

$$\left\|A\overrightarrow{x} - \overrightarrow{b}\right\| = \left\|Q\left(A\overrightarrow{x} - \overrightarrow{b}\right)\right\| = \left\|QA\overrightarrow{x} - Q\overrightarrow{b}\right\| = \left\|A^{(n)}\overrightarrow{x} - \overrightarrow{h}\right\|,$$

y como

$$A^{(n)}\overrightarrow{x} - \overrightarrow{h} = \begin{bmatrix} R \\ 0 \end{bmatrix}\overrightarrow{x} - \begin{bmatrix} \overrightarrow{h}_1 \\ \overrightarrow{h}_2 \end{bmatrix} = \begin{bmatrix} R\overrightarrow{x} - \overrightarrow{h}_1 \\ -\overrightarrow{h}_2 \end{bmatrix},$$

con lo que

$$\left\|A^{(n)}\overrightarrow{x} - \overrightarrow{h}\right\| = \left\|\begin{bmatrix} R\overrightarrow{x} - \overrightarrow{h}_1 \\ -\overrightarrow{h}_2 \end{bmatrix}\right\| = \left(\left\|R\overrightarrow{x} - \overrightarrow{h}_1\right\|^2 + \left\|\overrightarrow{h}_2\right\|^2\right)^{\frac{1}{2}}$$

y de esta igualdad, se sigue que $\left\|A^{(n)}\overrightarrow{x} - \overrightarrow{h}\right\|$ tendrá norma mínima si se elige $\overrightarrow{x}$ como solución del sistema de ecuaciones lineales $R\overrightarrow{x} = \overrightarrow{h}_1$, de donde $\overrightarrow{x} = R^{-1}\overrightarrow{h}_1$.

Note que la matriz R tiene inversa si y solo si las columnas de A son linealmente independientes, esto es, $\dim R(A) = n$.

Teorema 2 *Sea $A \in M_{m\times n}[\mathbb{R}]$ tal que $\dim \mathcal{R}(A) = n$ con $m \geq n$. Entonces A puede factorarse en un producto de la forma $A = Q\widetilde{R}$, donde $Q \in M_{m\times m}[\mathbb{R}]$ es una matriz ortogonal y $\widetilde{R} = \begin{bmatrix} R \\ 0 \end{bmatrix}$ con $R \in M_{n\times n}[\mathbb{R}]$ una matriz triangular superio invertible.*
Si $m = n$, $\ A = QR$.

Demostración. Basta aplicar el método de Householder. ∎

Observación

Sea $A \in M_{n\times n}[\mathbb{R}]$ con $\dim \mathcal{R}(A) = n$, $\overrightarrow{b} \in \mathbb{R}^n$ y consideramos el sistema de ecuaciones lineales $A\overrightarrow{x} = \overrightarrow{b}$. Ponemos $A = QR$. La solución del sistema de ecuaciones en mínimos cuadrados viene dada como

$$\begin{aligned} \overrightarrow{x} &= \left(A^T A\right)^{-1} A^T\overrightarrow{b} = \left[(QR)^T QR\right]^{-1} (QR)^T \overrightarrow{b} \\ &= \left[R^T Q^T QR\right]^{-1} R^T Q^T \overrightarrow{b} = \left(R^T R\right)^{-1} R^T Q^T \overrightarrow{b} \\ &= R^{-1}\left(R^T\right)^{-1} R^T Q^T \overrightarrow{b} = R^{-1} Q^T \overrightarrow{b}. \end{aligned}$$

Así, $\overrightarrow{x} = R^{-1}Q^T\overrightarrow{b} \Leftrightarrow R\overrightarrow{x} = Q^T\overrightarrow{b}$, además dicha solución $\overrightarrow{x}$ es única.

Ejemplo

Consideremos el sistema de ecuaciones lineales $\begin{cases} x - 2y + z = 1 \\ 2x + 5y + z = 2 \\ 4y - z = 3 \\ 2x + 3y + z = 4 \end{cases}$. Calculemos la solución en mínimos cuadrados aplicando el método de Householder.

Ponemos $A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 5 & 1 \\ 0 & 5 & -1 \\ 2 & 3 & 1 \end{bmatrix}$, $\overrightarrow{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$. Determinaremos una matriz ortogonal Q y una matriz triangular superior invertible R tales que $QA = \begin{bmatrix} R \\ 0 \end{bmatrix}$ y $\overrightarrow{h} = Q\overrightarrow{b} = \begin{bmatrix} \overrightarrow{h}_1 \\ \overrightarrow{h}_2 \end{bmatrix}$ con $\overrightarrow{h}_1 \in \mathbb{R}^n$, $\overrightarrow{h}_2 \in \mathbb{R}^{m-n}$.

Ponemos $A^{(1)} = A$ y $\overrightarrow{b}^{(1)} = \overrightarrow{b}$. Apliquemos el método de Householder.

1. $A^{(2)} = Q^{(1)}A^{(1)}$ con $Q^{(1)} = I - \dfrac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T$ y $\alpha = -\left\|\overrightarrow{v}\right\| sign\,(v_1)\,,\ r = \alpha\,(\alpha - v_1)\,,\ \overrightarrow{w} = \overrightarrow{v} - \alpha\,\overrightarrow{e}_1$.

   Sea $\overrightarrow{v} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 2 \end{bmatrix}$ vector obtenido de la primera columna de la matriz A. Calculamos su norma: $\left\|\overrightarrow{v}\right\| = 3$, y en consecuencia $\alpha = -3$. Puesto que $v_1 = 1$, se sigue que $r = -3\,(-3-1) = 12$. Con esta información pasamos a calcular el vector $\overrightarrow{w}$. Tenemos

$$\overrightarrow{w} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 8 \end{bmatrix},$$

   con lo que la matriz de Householder está definida como

$$Q^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \frac{1}{12}\begin{bmatrix} 4 \\ 2 \\ 0 \\ 2 \end{bmatrix}(4,2,0,2) = \frac{1}{3}\begin{bmatrix} -1 & -2 & 0 & -2 \\ -2 & 2 & 0 & -1 \\ 0 & 0 & 3 & 0 \\ -2 & -1 & 0 & 2 \end{bmatrix},$$

   y con esta calculamos la matriz $A^{(2)}$ siguiente:

$$A^{(2)} = Q^{(2)}A^{(1)} = \frac{1}{3}\begin{bmatrix} -1 & -2 & 0 & -2 \\ -2 & 2 & 0 & -1 \\ 0 & 0 & 3 & 0 \\ -2 & -1 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & -2 & 1 \\ 2 & 5 & 1 \\ 0 & 5 & -1 \\ 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} -3 & -\frac{14}{3} & -\frac{5}{3} \\ 0 & \frac{11}{3} & -\frac{1}{3} \\ 0 & 4 & -1 \\ 0 & \frac{5}{3} & -\frac{1}{3} \end{bmatrix}.$$

2. De la segunda columna de la matriz $A^{(2)}$, elegimos elvector $\overrightarrow{v} = \begin{bmatrix} 0 \\ \frac{11}{3} \\ 4 \\ \frac{5}{3} \end{bmatrix}$ y calculamos su norma, tenemos $\left\|\overrightarrow{v}\right\| = \dfrac{\sqrt{290}}{3}$, con lo que $\alpha = -\dfrac{\sqrt{290}}{3}$, pués $v_2 = \dfrac{11}{3}$. Luego

$$\frac{1}{r} = \frac{1}{\alpha\,(\alpha - v_2)} = \frac{1}{-\frac{\sqrt{290}}{3}\left(-\frac{\sqrt{290}}{3} - \frac{11}{3}\right)} = \frac{9}{\sqrt{290}\,(11+\sqrt{290})} \simeq 0{,}018855.$$

   Se define $\overrightarrow{w} = \begin{bmatrix} 0. \\ 9{,}34313 \\ 4. \\ 1{,}6667 \end{bmatrix}$. La matriz de Housholder está definida como $Q^{(2)} = I - \dfrac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T$, esto es,

$$
\begin{aligned}
Q^{(2)} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - 0{,}018855 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 81{,}2941 & 37{,}37252 & 15{,}57191 \\ 0 & 37{,}37252 & 16{,}0 & 6{,}666667 \\ 0 & 15{,}57191 & 6{,}666667 & 2{,}777778 \end{bmatrix} \\
&= \begin{bmatrix} 1. & 0. & 0. & 0. \\ 0. & -0{,}64593 & -0{,}70466 & -0{,}29361 \\ 0. & -0{,}70466 & 0{,}69832 & -0{,}12570 \\ 0. & -0{,}29361 & -0{,}12573 & 0{,}94762 \end{bmatrix}
\end{aligned}
$$

Se define $A^{(3)} = Q^{(2)} A^{(2)}$. Resulta

$$
\begin{aligned}
A^3 &= \begin{bmatrix} 1. & 0. & 0. & 0. \\ 0. & -0{,}64593 & -0{,}70466 & -0{,}29361 \\ 0. & -0{,}70466 & 0{,}69832 & -0{,}12570 \\ 0. & -0{,}29361 & -0{,}12573 & 0{,}94762 \end{bmatrix} \begin{bmatrix} -3 & -\frac{13}{3} & -\frac{5}{3} \\ 0. & \frac{11}{3} & -\frac{1}{3} \\ 0. & 4. & -1 \\ 0. & \frac{5}{3} & -\frac{1}{3} \end{bmatrix} \\
&= \begin{bmatrix} -3 & -4{,}66667 & -1{,}66667 \\ 0. & -5{,}67639 & 1{,}01784 \\ 0. & 0. & -0{,}42153 \\ 0. & 0. & -0{,}09231 \end{bmatrix}.
\end{aligned}
$$

Por lo tanto la matriz R está definida como

$$
R = \begin{bmatrix} -3 & -4{,}6667 & -1{,}66667 \\ 0. & -5{,}67639 & 1{,}01784 \\ 0. & 0. & -0{,}42153 \end{bmatrix}
$$

y su inversa

$$
R^{-1} = \begin{bmatrix} -0{,}33333 & 0{,}2740 & 1{,}9796 \\ 0. & -0{,}1761 & -0{,}4254 \\ 0. & 0. & -2{,}3723 \end{bmatrix}.
$$

Se define $\overrightarrow{h} = Q\overrightarrow{b} = Q^{(2)} Q^{(1)} \overrightarrow{b}$. Se tiene

$$
\begin{aligned}
\overrightarrow{h} &= \frac{1}{3} \begin{bmatrix} 1. & 0. & 0. & 0. \\ 0. & -0{,}64593 & -0{,}70466 & -0{,}29361 \\ 0. & -0{,}70466 & 0{,}69832 & -0{,}12570 \\ 0. & -0{,}29361 & -1{,}12570 & 0{,}94762 \end{bmatrix} \begin{bmatrix} -1 & -2 & 0 & -2 \\ -2 & 2 & 0 & -1 \\ 0 & 0 & 3 & 0 \\ -2 & -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \\
&= \begin{bmatrix} -4{,}3333 \\ -2{,}0748 \\ 2{,}3971 \\ 1{,}0821 \end{bmatrix} = \begin{bmatrix} \overrightarrow{h}_1 \\ \overrightarrow{h}_2 \end{bmatrix},
\end{aligned}
$$

de donde $\overrightarrow{h}_1 = \begin{bmatrix} -4{,}3333 \\ -2{,}0748 \\ 2{,}3971 \end{bmatrix}$

3. La solución en mínimos cuadrado está definida como

$$
\overrightarrow{x} = R^{-1} \overrightarrow{h}_1 = \begin{bmatrix} -0{,}3333 & 0{,}2740 & 1{,}9796 \\ 0. & -0{,}1761 & -0{,}4254 \\ 0. & 0. & -2{,}3723 \end{bmatrix} \begin{bmatrix} -4{,}3333 \\ -2{,}0748 \\ 2{,}3971 \end{bmatrix} = \begin{bmatrix} 5{,}6213 \\ -0{,}6544 \\ -5{,}6867 \end{bmatrix}.
$$

Note que en realidad no se requiere del cálculo de la matriz R^{-1}. Se resuelve directamente el sistema de ecuaciones lineales triangular superior $R\overrightarrow{x} = \overrightarrow{h}_1$.

9.4. Ajuste de datos polinomial

Para simplificar la escritura y que a su vez no pierda de generalidad, hemos seleccionado como problema de ajuste polinomial con un polinomio de tercer grado. El procedimiento que a continuación se describe, se aplica directamente al ajuste polinomial con polinomios de grados uno, dos, etc.

Supongamos que se dispone de un conjunto de n pares de datos experimentales $S = \{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, ..., n\}$. Se desea encontrar un polinomio P de grado 3: $P(x) = a + bx + cx^2 + dx^3 \quad x \in \mathbb{R}$, de modo que P se ajuste de la mejor manera al conjunto de datos S. El polinomio P queda perfectamente bien definido si se conocen todos sus coeficientes a, b, c, d. Estos coeficientes son calculados mediante el denominado método de mínimos cuadrados discreto que describimos a continuación.

Denotemos con r_i el residuo en cada medición, esto es,

$$y_i = P(x_i) + r_i = a + bx_i + cx_i^2 + dx_i^3 + r_i \quad i = 1, ..., n.$$

En forma matricial, el conjunto de ecuaciones precedente, se escribe como el siguiente sistema de ecuaciones lineales:

$$\begin{bmatrix} y_1 \\ \\ \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} + \begin{bmatrix} r_1 \\ \vdots \\ \vdots \\ r_n \end{bmatrix}.$$

El residuo en cada medición depende de los coeficientes a, b, c, d del polinomio P.

Definimos los vectores $\overrightarrow{x}$, $\overrightarrow{y}$ y el residuo $\overrightarrow{r}(\overrightarrow{x})$ con como sigue:

$$\overrightarrow{x} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \quad \overrightarrow{y} = \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}, \quad \overrightarrow{r}(\overrightarrow{x}) = \begin{bmatrix} r_1(\overrightarrow{x}) \\ \vdots \\ \vdots \\ r_n(\overrightarrow{x}) \end{bmatrix}.$$

Definimos la matriz A siguiente: $A = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}.$

El sistema de ecuaciones lineales arriba propuesto se transforma en el siguiente: $\overrightarrow{y} = A\overrightarrow{x} + \overrightarrow{r}(\overrightarrow{x})$, de donde el residuo $\overrightarrow{r}(\overrightarrow{x}) = \overrightarrow{y} - A\overrightarrow{x}$ con $\overrightarrow{x} \in \mathbb{R}^4$. El problema de hallar el "mejor polinomio" que se ajusta al conjunto de datos S se expresa como sigue:

$$\text{hallar } \widehat{x}^T = (\widehat{a}, \widehat{b}, \widehat{c}, \widehat{d}) \in \mathbb{R}^4 \text{ tal que } \|\overrightarrow{r}(\widehat{x})\|^2 = \underset{\overrightarrow{x} \in \mathbb{R}^4}{\text{Min}} \|\overrightarrow{r}(\widehat{x})\|^2,$$

o de modo equivalente

$$\|\overrightarrow{y} - A\widehat{x}\|^2 = \underset{\overrightarrow{x} \in \mathbb{R}^4}{\text{Min}} \|\overrightarrow{y} - A\widehat{x}\|^2.$$

Este problema, como ya hemos señalado, se conoce como método de mínimos cuadrados y se ha demostrado que conduce a resolver el sistema de ecuaciones $A^T A\overrightarrow{x} = A^T\overrightarrow{y}$, donde A^T denota la matriz transpuesta de A.

Veamos la generalización de este problema. Sea $C = \left\{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots, n\right\}$ un conjunto de datos experimentales y p un polinomio de grado $\leq m$. Ponemos

$$p(x) = a_0 + a_1 x_1 + \ldots + a_m x^m \quad x \in \mathbb{R},$$

donde $a_0, \ldots, a_m \in \mathbb{R}$ se determinan mediante el método de mínimos cuadrados. Supongamos

$$\begin{cases} y_1 = a_0 + a_1 x_1 + \ldots + a_m x_1^m + r_1 \\ \vdots \\ y_n = a_0 + a_1 x_n + \ldots + a_m x_n^m + r_n, \end{cases}$$

con $r_1, \ldots, r_n$ los errores cometidos en la medición. En forma matricial, el sistema de ecuaciones precedente, se expresa como $\overrightarrow{y} = A\overrightarrow{x} + \overrightarrow{r}$, donde $\overrightarrow{y} = (y_1, \ldots, y_n)^T$, $\overrightarrow{r} = (r_1, \ldots, r_n)^T$, $\overrightarrow{x} = (a_0, \ldots, a_m)^T$,

$$A = (a_{ij}) = \begin{bmatrix} 1 & x_1 & \ldots & x_1^m \\ & \vdots & & \\ 1 & x_n & \ldots & x_n^m \end{bmatrix} = (A_0, \ldots, A_m),$$

con $A_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \ldots, A_0 = \begin{bmatrix} x_1^m \\ \vdots \\ x_n^m \end{bmatrix}$ las columnas de A.

Note que si $\overrightarrow{x} = (a_0, \ldots, a_m)^T \in \mathbb{R}^{m+1}$, se tiene $A\overrightarrow{x} = a_0 A_0 + \ldots + a_m A_m$. Sea $W = \left\{A\overrightarrow{x} \mid \overrightarrow{x} \in \mathbb{R}^{m+1}\right\}$. Resulta que W es un subespacio de $\mathbb{R}^n$ con $\dim W = m < n$. Se define $\overrightarrow{r}(\overrightarrow{x}) = \overrightarrow{y} - A\overrightarrow{x} \quad \overrightarrow{x} \in \mathbb{R}^{m+1}$. El problema de mínimos cuadrados consiste en determinar $\hat{x} \in \mathbb{R}^{m+1}$ tal que

$$\|\overrightarrow{r}(\hat{x})\|^2 \leq \|\overrightarrow{r}(\overrightarrow{x})\|^2 \quad \forall \overrightarrow{x} \in \mathbb{R}^{m+1} \Longleftrightarrow \|\overrightarrow{y} - A\hat{x}\|^2 = \underset{\overrightarrow{x} \in \mathbb{R}^{m+1}}{\text{Min}} \|\overrightarrow{y} - A\overrightarrow{x}\|^2.$$

Note que, en general, $\overrightarrow{y} \notin W$. Por el teorema precedente, existe $\hat{y} \in W$ tal que

$$\|\overrightarrow{y} - \hat{y}\| \leq \|\overrightarrow{y} - \overrightarrow{z}\| \quad \forall \overrightarrow{z} \in W,$$

$$\langle \overrightarrow{y} - \hat{y},\ \overrightarrow{w} \rangle = 0 \quad \forall \overrightarrow{w} \in W.$$

Como $\hat{y} \in W$ si y solo si $\overrightarrow{x} \in \mathbb{R}^{m+1}$ tal que $\hat{y} = A\hat{x}$, $\overrightarrow{z} \in W$ si y solo si existe $\overrightarrow{x} \in \mathbb{R}^{m+1}$ tal que $\overrightarrow{z} = A\overrightarrow{x}$.

Así,

$$\|\overrightarrow{y} - A\hat{x}\|^2 \leq \|\overrightarrow{y} - A\overrightarrow{x}\|^2 \quad \forall \overrightarrow{x} \in \mathbb{R}^{m+1},$$

$$\langle \overrightarrow{y} - A\hat{x},\ \overrightarrow{w} \rangle = 0 \quad \forall \overrightarrow{w} \in W.$$

Para $\overrightarrow{w} = A\overrightarrow{x} \quad \overrightarrow{x} \in \mathbb{R}^{m+1}$, se obtiene

$$0 = \langle \overrightarrow{y} - A\hat{x},\ Ax \rangle = (A\overrightarrow{x})^T (\overrightarrow{y} - A\hat{x}) = \overrightarrow{x}^T A^T \overrightarrow{y}^T - \overrightarrow{x}^T A^T A\hat{x} = \overrightarrow{x}^T \left(A^T \overrightarrow{y} - A^T A\hat{x}\right).$$

Luego

$$\overrightarrow{x}^T \left(A^T \overrightarrow{y} - A^T A\hat{x}\right) = 0 \quad \forall \overrightarrow{x} \in \mathbb{R}^{m+1} \Longleftrightarrow A^T A\hat{x} = A^T \overrightarrow{y},$$

que se conoce como ecuación normal. O sea $\hat{x}$ es solución del sistema de ecuaciones normal.

9.4.1. Ajuste de datos con polinomios de grado 1.

Consideramos el conjunto de datos $S = \left\{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots, n\right\}$ y supongamos que dicho conjunto de puntos tiene una tendencia como la de un polinomio de grado 1, esto es, $p(x) = a + bx \quad x \in \mathbb{R}$, donde $a, b \in \mathbb{R}$ se determinan como soluciones en mínimos cuadrados. Tenemos

$$y_i = a + bx_i + r_i \quad i = 1, \ldots, n.$$

Ponemos $\overrightarrow{x}^T = (a, b) \in \mathbb{R}^2$, $\overrightarrow{y}^T = (y_1, \ldots, y_n)$, $A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, el residuo definido como $\overrightarrow{r}(\overrightarrow{x})^T = (r_1(\overrightarrow{x}), \ldots, r_n(\overrightarrow{x}))$. El sistema de ecuaciones precedente se escribe como $\overrightarrow{y} = A\overrightarrow{x} + \overrightarrow{r}(\overrightarrow{x}) \quad \overrightarrow{x} \in \mathbb{R}^2$. Luego, el problema en mínimos cuadrados está definido como

$$E(a, b) = \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \|\overrightarrow{r}(\overrightarrow{x})\|^2 = \|\overrightarrow{y} - A\overrightarrow{x}\|^2 \quad \overrightarrow{x} \in \mathbb{R}^2.$$

La solución en mínimos cuadrados se expresa como

$$\widehat{x} = \left(A^T A\right)^{-1} A^T \overrightarrow{y}.$$

Puesto que

$$\begin{aligned}
A^T A &= \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 \end{bmatrix}, \\
A^T \overrightarrow{y} &= \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i y_i \end{bmatrix}.
\end{aligned}$$

Es este caso resulta fácil el cálculo de $\left(A^T A\right)^{-1}$, tenemos

$$\left(A^T A\right)^{-1} = \frac{1}{n \sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum\limits_{i=1}^{n} x_i^2 & -\sum\limits_{i=1}^{n} x_i \\ -\sum\limits_{i=1}^{n} x_i & n \end{bmatrix},$$

luego

$$\begin{aligned}
\widehat{x} &= \begin{bmatrix} \widehat{a} \\ \widehat{b} \end{bmatrix} = \left(A^T A\right)^{-1} A^T \overrightarrow{y} = \frac{1}{n \sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum\limits_{i=1}^{n} x_i^2 & -\sum\limits_{i=1}^{n} x_i \\ -\sum\limits_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i y_i \end{bmatrix} \\
&= \frac{1}{n \sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \left(\sum\limits_{i=1}^{n} x_i^2\right)\left(\sum\limits_{i=1}^{n} y_i\right) - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} x_i y_i\right) \\ -\left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right) + n \sum\limits_{i=1}^{n} x_i y_i \end{bmatrix}
\end{aligned}$$

y de esta igualdad obtenemos los coeficientes $\widehat{a}$, $\widehat{b}$:

$$\begin{aligned}
\widehat{a} &= \frac{\left(\sum\limits_{i=1}^{n} x_i^2\right)\left(\sum\limits_{i=1}^{n} y_i\right) - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} x_i y_i\right)}{n\left(\sum\limits_{i=1}^{n} x_i^2\right) - \left(\sum\limits_{i=1}^{n} x_i\right)^2}, \\
\widehat{b} &= \frac{n \sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{n\left(\sum\limits_{i=1}^{n} x_i^2\right) - \left(\sum\limits_{i=1}^{n} x_i\right)^2}.
\end{aligned}$$

Si aplicamos las condiciones necesarias de extremo a la función $E\left(a,b\right)$, tenemos

$$\left\{ \begin{array}{l} \dfrac{\partial E}{\partial a}\left(a,b\right) = 0 \\ \dfrac{\partial E}{\partial b}\left(a,b\right) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \sum\limits_{i=1}^{n} \left(y_i - a - b x_i\right) = 0 \\ \sum\limits_{i=1}^{n} \left(y_i - a - b x_i\right) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} na + b \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i \\ a \sum\limits_{i=1}^{n} x_i + b \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i, \end{array} \right.$$

cuya solución es

$$
\begin{aligned}
a &= \frac{\left(\sum\limits_{i=1}^{n} x_i^2\right)\left(\sum\limits_{i=1}^{n} y_i\right) - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} x_i y_i\right)}{n\left(\sum\limits_{i=1}^{n} x_i^2\right) - \left(\sum\limits_{i=1}^{n} x_i\right)^2}, \\
b &= \frac{n\sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2},
\end{aligned}
$$

que coincide exactamente con $\widehat{a}$, $\widehat{b}$.

Note que hemos supuesto que $\left(\sum\limits_{i=1}^{n} x_i\right)^2 \neq n\sum\limits_{i=1}^{n} x_i^2$.

Para calcular las constantes $\widehat{a}$, $\widehat{b}$ mediante el método de mínimos cuadrados se requiere de la siguiente información: el número de puntos $n \geq 2$ y el conjunto de datos $S = \{(x_i, y_i) \mid i = 1, \ldots, n\}$; y, se deben calcular las sigientes sumas: $S_1 = \sum\limits_{i=1}^{n} x_i$, $S_2 = \sum\limits_{i=1}^{n} x_i^2$, $R_1 = \sum\limits_{i=1}^{n} y_i$, $R_2 = \sum\limits_{i=1}^{n} x_i y_i$. Con estos resultados se pasa al cálculo de $\widehat{a}$ y $\widehat{b}$. Se propone el siguiente algoritmo de cálculo de $\widehat{a}$ y $\widehat{b}$.

<u>Algoritmo</u>

Datos de entrada: $n \in \mathbb{Z}^+$, $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^2$.

Datos de salida: Mensaje1, Mensaje 2, a, b.

1. Si $n < 2$, continuar en 14)
2. $S_1 = 0$.
3. $S_2 = 0$.
4. $R_1 = 0$.
5. $R_2 = 0$.
6. Para $i = 1, \ldots, n$

   $S_1 = S_1 + x_i$

   $S_2 = S_2 + x_i^2$

   $R_1 = R_1 + y_i$

   $R_2 = R_2 + x_i y_i$

   Fin de bucle i.
7. $z = nS_2 - S_1^2$
8. Si $z = 0$, continuar en 13)
9. $a = \dfrac{S_2 R_1 - S_1 R_2}{z}$,
10. $b = \dfrac{nR_2 - S_1 R_1}{z}$
11. Imprimir a, b, continuar en 14)
12. Mensaje 1: $n \geq 2$ continuar en 14)

13. Mensaje 2: el sistema no tiene solución o tiene infinitas soluciones.

14. Fin

En la gráfica siguiente se ilustra un conjunto de puntos $S = \{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots, n\}$ que siguen una tendencia de una recta, y, la gráfica de la recta solución en mínimos cuadrados, de ecuación $y = a + bx$ $x \in \mathbb{R}$.

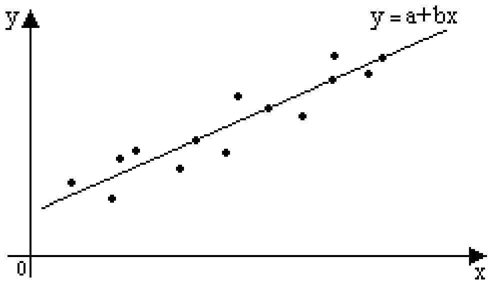

Figura 84

Ejemplo

Apliquemos el algoritmo al siguiente conjunto de datos:

$$S = \{(1,\ 3{,}6), (1{,}5,\ 4{,}35), (2{,}1,\ 5{,}25), (2{,}9,\ 6{,}45), (3{,}2,\ 6{,}9)\}.$$

Buscamos una función p de la forma $p(x) = a + bx \quad x \in \mathbb{R}$, con $a,\ b$ constantes calculadas con el método de mínimos cuadrados

$$\begin{aligned}
S_1 &= \sum_{i=1}^{5} x_i = 1 + 1{,}5 + 2{,}1 + 2{,}9 + 3{,}2 = 10{,}7, \\
S_2 &= \sum_{i=1}^{n} x_i^2 = 1 + 2{,}25 + 4{,}41 + 8{,}41 + 10{,}24 = 26{,}31, \\
R_1 &= \sum_{i=1}^{n} y_i = 3{,}6 + 4{,}35 + 5{,}25 + 6{,}45 + 6{,}9 = 26{,}55, \\
R_2 &= \sum_{i=1}^{n} x_i y_i = 3{,}6 + 6{,}525 + 11{,}025 + 18{,}705 + 22{,}08 = 61{,}935. \\
z &= nS_2 - S_1^2 = 5 \times 26{,}31 - (10{,}7)^2 = 17{,}06.
\end{aligned}$$

La solución es

$$\begin{aligned}
a &= \frac{S_2 \times R_1 - S_1 R_2}{z} = \frac{26{,}55 \times 26{,}31 - 10{,}7 \times 61{,}935}{17{,}06} = 2{,}1 \\
b &= \frac{nR_2 - S_1 \times R_1}{z} = \frac{5 \times 61{,}935 - 10{,}07 \times 26{,}55}{17{,}06} = 1{,}5.
\end{aligned}$$

El polinomio de grado 1 buscado es $p(x) = 2{,}1 + 1{,}5x \quad x \in \mathbb{R}$. Note que $p(x)$ es la ecuación cartesiana de la recta a la que se le denomina recta de mejor ajuste en mínimos cuadrados.

9.4.2. Ajuste polinomial con polinomios de grado 2.

Dado el conjunto de datos $S = \{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots, n\}$ que tiene una tendencia de una función cuadrática del tipo $p(x) = a + bx + cx^2 \quad x \in \mathbb{R}$, donde $a,\ b,\ c$ son constantes reales que se determinan como soluciones del método de mínimos cuadrados. tenemos

$$y_i = a + bx_i + cx_i^2 + r_i \quad i = 1, \ldots, n.$$

Siguiendo el mismo procedimiento descrito anteriormente, definimos los vectores de $\mathbb{R}^n$ siguientes: $\overrightarrow{y}^T = (y_1, \ldots, y_n)$, $\overrightarrow{r}(\overrightarrow{x})^T = (r_1(\overrightarrow{x}), \ldots, r_n(\overrightarrow{x}))$ con $\overrightarrow{x}^T = (a, b, c) \in \mathbb{R}^3$; y, se define la matriz $A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$. Entonces

$$\overrightarrow{y} = A\overrightarrow{x} + \overrightarrow{r}(\overrightarrow{x}) \qquad \overrightarrow{x} \in \mathbb{R}^3.$$

Se define la función E de $\mathbb{R}^3$ en $[0, \infty[$ como sigue

$$E(a,b,c) = \|\overrightarrow{r}(a,b,c)\|^2 = \|\overrightarrow{y} - A\overrightarrow{x}\|^2 = \sum_{i=1}^{n} \left(y_i - a - bx_i - cx_i^2\right)^2 \qquad (a,b,c) \in \mathbb{R}^3.$$

El sistema de ecuaciones normal está definido como $A^T A\overrightarrow{x} = A^T\overrightarrow{y}$. Calculemos $A^T A$ y $A^T\overrightarrow{y}$. Tenemos

$$\begin{aligned} A^T A &= \begin{bmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_n \\ x_1^2 & \ldots & x_n^2 \end{bmatrix} \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} = \begin{bmatrix} n & \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i^3 \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^4 \end{bmatrix}, \\ A^T\overrightarrow{y} &= \begin{bmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_n \\ x_1^2 & \ldots & x_n^2 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i y_i \\ \sum\limits_{i=1}^{n} x_i^2 y_i \end{bmatrix}. \end{aligned}$$

La solución $\widehat{x}^T = \left(\widehat{a}, \widehat{b}, \widehat{c}\right)$ se obtiene del sistema de ecuaciones lineales

$$\begin{bmatrix} n & \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i^3 \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i y_i \\ \sum\limits_{i=1}^{n} x_i^2 y_i \end{bmatrix}.$$

Si aplicamos las condiciones necesarias de extremo a la función $E(a,b,c)$, tenemos

$$\begin{cases} \dfrac{\partial E}{\partial a}(a,b,c) = 0 \\ \dfrac{\partial E}{\partial b}(a,b,c) = 0 \\ \dfrac{\partial E}{\partial c}(a,b,c) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum\limits_{i=1}^{n} \left(y_i - a - bx_i - cx_i^2\right) = 0 \\ \sum\limits_{i=1}^{n} \left(y_i - a - bx_i - cx_i^2\right) x_i = 0 \\ \sum\limits_{i=1}^{n} \left(y_i - a - bx_i - cx_i^2\right) x_i^2 = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} na + b\sum\limits_{i=1}^{n} x_i + c\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} y_i, \\ a\sum\limits_{i=1}^{n} x_i + b\sum\limits_{i=1}^{n} x_i^2 + c\sum\limits_{i=1}^{n} x_i^3 = \sum\limits_{i=1}^{n} x_i y_i, \\ a\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i^3 + c\sum\limits_{i=1}^{n} x_i^4 = \sum\limits_{i=1}^{n} x_i^2 y_i, \end{cases}$$

que es exactamente el sistema de ecuaciones normal.

Primeramente, para resolver el sistema de ecuaciones, hemos de calcular cada uno de los sumatorios. Ponemos

$$\begin{aligned} S_1 &= \sum_{i=1}^{n} x_i,\ S_2 = \sum_{i=1}^{n} x_i^2,\ S_3 = \sum_{i=1}^{n} x_i^3,\ S_4 = \sum_{i=1}^{n} x_i^4, \\ R_1 &= \sum_{i=1}^{n} y_i,\ R_2 = \sum_{i=1}^{n} x_i y_i,\ R_3 = \sum_{i=1}^{n} x_i^2 y_i, \end{aligned}$$

con lo que el sistema de ecuaciones lineales se escribe como

$$\begin{bmatrix} n & S_1 & S_2 \\ S_1 & S_2 & S_3 \\ S_2 & S_3 & S_4 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix}.$$

Este sistema de ecuaciones se resuelve con el método de eliminación gaussiana.

Proponemos como ejercicio la elaboración de un algoritmo para el cálculo de $\widehat{x}^T = \left(\widehat{a}, \widehat{b}, \widehat{c}\right)$.

Ejemplo

Consideremos el conjunto de datos S siguiente: $S = \{(0,3), (1,2), (2,3), (3,6), (4,11)\}$. Determinemos un polinomio $p(x) = a_1 + a_2x + a_3x^2 \quad x \in \mathbb{R}$ en mínimos cuadrados.

De los resultados arriba establecidos, se tienen $\overrightarrow{x}^T$, $\overrightarrow{y}^T$ y A definidos como sigue: $\overrightarrow{x} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$,

$\overrightarrow{y}^T = (3,2,3,6,11)$, $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}$. Luego

$$\begin{aligned} A^TA &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 4 & 9 & 16 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} = \begin{bmatrix} 5 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix}, \\ A^T\overrightarrow{y} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 4 & 9 & 16 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ 3 \\ 6 \\ 11 \end{bmatrix} = \begin{bmatrix} 25 \\ 70 \\ 244 \end{bmatrix}. \end{aligned}$$

El sistema normal de ecuaciones lineales es

$$\begin{bmatrix} 5 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 25 \\ 70 \\ 244 \end{bmatrix}.$$

Apliquemos el método de eliminación gaussiana con pivoting total. Para el efecto, ponemos

$$\widetilde{A} = \left[\begin{array}{ccc|c} 354. & 100. & 30. & 244. \\ 100. & 30. & 10. & 70. \\ 30. & 10. & 5. & 25. \end{array}\right]$$

entonces

$$\begin{aligned} \widetilde{A}^{(1)} &= \left[\begin{array}{ccc|c} 354. & 100. & 30. & 244. \\ 0. & 1{,}75141243 & 1{,}525423729 & 1{,}073446329 \\ 0. & 1{,}525423729 & 2{,}457627119 & 4{,}322033899 \end{array}\right], \\ \widetilde{A}^{(2)} &= \left[\begin{array}{ccc|c} 354. & 100. & 30. & 244. \\ 0. & 1{,}75141243 & 1{,}525423729 & 1{,}073446329 \\ 0. & 0. & 1{,}129032259 & 3{,}387096774 \end{array}\right], \end{aligned}$$

y de este sistema de ecuaciones triangular superior, se obtiene

$$\widehat{a}_1 = 2{,}999999997 \simeq 3{,}0, \quad \widehat{a}_2 = -1{,}999999996 \simeq -2{,}0, \quad \widehat{a}_3 = 1.$$

El polinomio buscado es $p(x) = 3 - 2x + x^2 \quad x \in \mathbb{R}$.

9.5. Ajuste de datos con funciones afines de n variables

Supongamos que se dispone de un conjunto de datos

$$S = \left\{ \left(x_1^{(k)}, \ldots, x_n^{(k)}, z_k \right) \in \mathbb{R}^{n+1} \mid k = 1, \ldots, m \right\},$$

y se desea hallar los coeficientes de la función f definida como

$$z = f(x_1, \ldots, x_n) = a_0 + a_1 x + \cdots + a_n x_n \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Se supone que $m \geq n + 1$.

Si $n = 1$, la función f se escribe como $f(x) = a + bx \quad x \in \mathbb{R}$ y coincide con la de un polinomio de grado 1 que ya hemos arriba tratado. Del punto de vista geométrico, la gráfica de f representa una recta del plano, por este motivo llamamos recta de mejor ajuste.

Si $n = 2$, se tiene $z = f(x, y) = a_0 + a_1 x + a_2 y \qquad (x, y) \in \mathbb{R}^2$, que del punto de vista geométrico, la gráfica de f se identifica con un plano.

Si $n \geq 3$, a la función f lo identificamos con la ecuación cartesiana de un hiperplano.

Tal como en el caso del ajuste de datos polinomial, se tiene el siguiente sistema de ecuaciones:

$$z_k = a_0 + a_1 x_1^{(k)} + \cdots + a_n x_n^{(k)} + r_k \quad k = 1, \ldots, m,$$

donde r_k denota la perturbación del dato experimental $\left(x_1^{(k)}, \ldots, x_n^{(k)}, z_k \right)$ y que depende de $a_0, a_1, \ldots, a_n$.

Se define $\overrightarrow{x}^T = (a_0, a_1, \ldots, a_n) \in \mathbb{R}^{n+1}$, $\overrightarrow{z}^T = (z_1, \ldots z_m)$, $\overrightarrow{r}(\overrightarrow{x})^T = (r_1(\overrightarrow{x}), \ldots, r_m(\overrightarrow{x}))$ y
$A = \begin{bmatrix} 1 & x_1^{(1)} & \ldots & x_n^{(1)} \\ \vdots & \vdots & & \vdots \\ 1 & x_n^{(1)} & \ldots & x_n^{(m)} \end{bmatrix}$.

El sistema de ecuaciones precedente se escribe en forma matricial como $\overrightarrow{z} = A\overrightarrow{x} + \overrightarrow{r}(\overrightarrow{x})$ o bien $r(\overrightarrow{x}) = \overrightarrow{z} - A\overrightarrow{x} \quad \overrightarrow{x} \in \mathbb{R}^{n+1}$.

Se define la función E de $\mathbb{R}^{n+1}$ en $[0, \infty[$ como sigue:

$$E(a_0, a_1, \ldots, a_n) = \|\overrightarrow{r}(\overrightarrow{x})\|^2 = \|\overrightarrow{z} - A\overrightarrow{x}\|^2 = \sum_{k=1}^{m} \left(z_k - a_0 - a_1 x_1^{(k)} - \cdots - a_n x_n^{(k)} \right)^2$$

y consideramos el problema de minimización siguiente:

$$\text{hallar } \widehat{x}^T = (\widehat{a}_0, \ldots, \widehat{a}_n) \in \mathbb{R}^{n+1} \text{ tal que } E(\widehat{a}_0, \ldots, \widehat{a}_n) = \min_{(a_0, \ldots, a_n) \in \mathbb{R}^{n+1}} E(a_0, \ldots, a_n).$$

Las condiciones necesarias de extremo implican

$$\begin{cases} \dfrac{\partial E}{\partial a_0}(a_0, \ldots, a_n) = 0 \\ \qquad \vdots \\ \dfrac{\partial E}{\partial a_n}(a_0, \ldots, a_n) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum\limits_{k=1}^{m} \left(z_k - a_0 - a_1 x_1^{(k)} - \cdots - a_n x_n^{(k)} \right) = 0, \\ \qquad \vdots \\ \sum\limits_{k=1}^{m} \left(z_k - a_0 - a_1 x_1^{(k)} - \cdots - a_n x_n^{(k)} \right) x_n^{(k)} = 0, \end{cases}$$

que se expresa en forma matricial como

$$\left(A^T A \right) \overrightarrow{x} = A^T \overrightarrow{z} \quad x \in \mathbb{R}^{n+1},$$

que es el sistema normal de ecuaciones lineales.

Este sistema se obtuvo como solución en mínimos cuadrados del sistema de ecuaciones lineales $A\overrightarrow{x} = \overrightarrow{z}$ $\overrightarrow{x} \in \mathbb{R}^{n+1}$.

Ejemplos

1. Consideremos el conjunto de datos S siguiente:

$$S = \{(1, 0{,}5, 9), (2, 2{,}5, 16), (3, 3, 20), (4, 3{,}5, 24), (5, 5, 32)\}.$$

Determinemos una función f del tipo $f(x, y, z) = ax + by + c \quad (x, y, z) \in \mathbb{R}^3$ y $a,\ b,\ c$ constantes determinadas con el método de mínimos cuadrados.

Definimos la matriz A y los vectores $\overrightarrow{x}$, $\overrightarrow{y}$ como sigue:

$$A = \begin{bmatrix} 1 & 0{,}5 & 1 \\ 2 & 2{,}5 & 1 \\ 3 & 3 & 1 \\ 4 & 3{,}5 & 1 \\ 5 & 6 & 1 \end{bmatrix}, \quad \overrightarrow{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3, \quad \overrightarrow{y} = \begin{bmatrix} 9 \\ 16 \\ 20 \\ 24 \\ 32 \end{bmatrix}$$

El problema en mínimos cuadrados es el siguiente:

$$\text{hallar } \widehat{x} \in \mathbb{R}^3 \text{ solución de } \|A\widehat{x} - \widehat{y}\|^2 = \min_{\overrightarrow{x} \in \mathbb{R}^3} \|A\overrightarrow{x} - \overrightarrow{y}\|^2.$$

La solución en mínimos cuadrados satisface la ecuación normal $A^T A\overrightarrow{x} = A^T\overrightarrow{y}$. Note que el rango de la matriz es 3, luego $A^T A$ tiene también rango 3 y en consecuencia $A^T A$ es invertible. Se tiene

$$A^T A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0{,}5 & 2{,}5 & 3 & 3{,}5 & 6 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0{,}5 & 1 \\ 2 & 2{,}5 & 1 \\ 3 & 3 & 1 \\ 4 & 3{,}5 & 1 \\ 5 & 6 & 1 \end{bmatrix} = \begin{bmatrix} 55 & 58{,}5 & 15 \\ 58{,}5 & 63{,}75 & 15{,}5 \\ 15 & 15{,}5 & 5 \end{bmatrix}$$

$$A^T\overrightarrow{y} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0{,}5 & 2{,}5 & 3 & 3{,}5 & 6 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 16 \\ 20 \\ 24 \\ 32 \end{bmatrix} = \begin{bmatrix} 357 \\ 380{,}5 \\ 101 \end{bmatrix}.$$

El sistema normal de ecuaciones lineales se escribe:

$$\begin{bmatrix} 55 & 58{,}5 & 15 \\ 58{,}5 & 63{,}75 & 15{,}5 \\ 15 & 15{,}5 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 357 \\ 380{,}5 \\ 101 \end{bmatrix}.$$

Para hallar la solución de este sistema de ecuacones lineales aplicamos el método de eliminación gaussiana con pivoting parcial. Tenemos la matriz ampliada en la que se intercambiaron la primera y segunda filas

$$\widetilde{A} = \left[\begin{array}{ccc|c} 58{,}5 & 63{,}75 & 15{,}5 & 380{,}5 \\ 55 & 58{,}5 & 15 & 357 \\ 15 & 15{,}5 & 5 & 101 \end{array}\right],$$

$$\widetilde{A}^{(1)} = \left[\begin{array}{ccc|c} 58{,}5 & 63{,}75 & 15{,}5 & 380{,}5 \\ 0. & -1{,}43589744 & 0{,}42735043 & -0{,}7350427 \\ 0. & -0{,}84615385 & 1{,}025641026 & 3{,}43589744 \end{array}\right],$$

$$\widetilde{A}^{(2)} = \left[\begin{array}{ccc|c} 58{,}5 & 63{,}75 & 15{,}5 & 380{,}5 \\ 0. & -1{,}43589744 & 0{,}42735043 & -0{,}7350427 \\ 0. & 0. & 0{,}7738095222 & 3{,}869047603 \end{array}\right].$$

La solución del sistema de ecuaciones triangular superior nos da los resultados siguientes:

$$\widehat{c} = 4{,}99999999 \simeq 5{,}0, \quad \widehat{b} = 1{,}9999999976 \simeq 2{,}0 \quad , \widehat{a} \simeq 3{,}0$$

La función buscada está definda como

$$f(x, y) = 3x + 2y + 5 \quad (x, y) \in \mathbb{R}^2.$$

2. Consideremos el conjunto de datos S siguiente:

$$S = \{(5, 1, 2{,}5)\,, \ \ (15, 3, 3{,}5)\,, \ \ (20, 4, 4)\,, \ \ (40, 8, 6)\}\,.$$

Determinemos, siempre que sea posible, una función del tipo $z = f(x, y) = a+bx+cy \quad (x, y) \in \mathbb{R}^2$ y $a,\ b,\ c$ constantes que se determinen con el método de mínimos cuadrados.

Tal como en el ejemplo anterior, definimos los vectores $\overrightarrow{x}$, $\overrightarrow{z}$ y la matriz A como sigue:

$$\overrightarrow{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3, \quad \overrightarrow{z} = \begin{bmatrix} 2{,}5 \\ 3{,}5 \\ 4{,}0 \\ 6{,}0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 5 & 1 \\ 1 & 15 & 3 \\ 1 & 29 & 4 \\ 1 & 40 & 8 \end{bmatrix}.$$

Entonces

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 15 & 20 & 40 \\ 1 & 3 & 4 & 8 \end{bmatrix} \begin{bmatrix} 1 & 5 & 1 \\ 1 & 15 & 3 \\ 1 & 29 & 4 \\ 1 & 40 & 8 \end{bmatrix}. = \begin{bmatrix} 4 & 80 & 16 \\ 80 & 2250 & 450 \\ 16 & 450 & 90 \end{bmatrix},$$

$$A^T \overrightarrow{z} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 15 & 20 & 40 \\ 1 & 3 & 4 & 8 \end{bmatrix} \begin{bmatrix} 2{,}5 \\ 3{,}5 \\ 4{,}0 \\ 6{,}0 \end{bmatrix} = \begin{bmatrix} 16 \\ 385 \\ 77 \end{bmatrix}.$$

El sistema normal de ecuaciones lineales $A^T A \overrightarrow{x} = A^T \overrightarrow{z}$ se expresa como

$$\begin{bmatrix} 4 & 80 & 16 \\ 80 & 2250 & 450 \\ 16 & 450 & 90 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 16 \\ 385 \\ 77 \end{bmatrix}.$$

Debemos notar que la matriz A tiene rango 2, por lo tanto la matriz $A^T A$ no es invertible. En este caso el sistema normal de ecuaciones lineales debe ser resuelto con el método QR de Householder.

3. Considerar el conjunto de datos dados en la tabla siguiente

i	x_i	y_i	z_i	w_i
1	0	0,25	4.	14,5
2	1	0,5	6.	21.
3	2	1.	10.	33.
4	3	1,5	20.	63.
5	4	2.	25.	78.

Sea f la función real definida por

$$w = f(x, y, z) = a + bx + cy + dz, \quad (x, y, z) \in \mathbb{R}^3,$$

donde $a,\ b,\ c,\ d$ son constantes reales que deben determinarse utilizando el método de mínimos cuadrados.

i) Halle el sistema normal de ecuaciones.

ii) Aplique el método de Householder al sistema de ecuaciones normales para hallar una solución (si existe) del sistema de ecuaciones normales

Solución

Se busca una función real f definida como

$$f(x, y, z) = a + bx + cy + dz, \quad (x, y, z) \in \mathbb{R}^3,$$

donde a, b, c, d son constantes reales que deben determinarse utilizando el método de mínimos cuadrados y usando la información suministrada en la tabla, esto es

$$w_i = f(x_i, y_i, z_i) = a + bx_i + cy_i + dz_i + r_i, \quad i = 1, \ldots, n \tag{1}$$

con $r_i \in \mathbb{R}$ el error cometido en cada observación.

Ponemos $\overrightarrow{x}^T = (a, b, c, d)$, $\overrightarrow{w}^T = (w_1, w_2, w_3, w_4, w_5)$, $\overrightarrow{r}^T = (r_1, r_2, r_3, r_4, r_5)$ y $A = (a_{ij}) \in M_{4\times 5}[\mathbb{R}]$ dada por

$$A = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \\ 1 & x_5 & y_5 & z_5 \end{bmatrix}.$$

El valor $\overrightarrow{r}$ es función de $\overrightarrow{x}$, escribiremos $\overrightarrow{r}(\overrightarrow{x})$.

El sistema de ecuaciones (1) se escribe en forma matricial como el siguiente:

$$A\overrightarrow{x} + \overrightarrow{r}(\overrightarrow{x}) = \overrightarrow{w}. \tag{2}$$

De (2) se obtiene

$$\overrightarrow{r}(\overrightarrow{x}) = \overrightarrow{w} - A\overrightarrow{x}.$$

El problema consiste en determinar $\widehat{x} \in \mathbb{R}^4$ solución de

$$\min_{\overrightarrow{x}\in\mathbb{R}^4} \|\overrightarrow{r}(\overrightarrow{x})\|^2$$

o lo que es lo mismo

$$\|\overrightarrow{w} - A\widehat{x}\|^2 = \min_{\overrightarrow{x}\in\mathbb{R}^4} \|\overrightarrow{w} - A\overrightarrow{x}\|^2. \tag{3}$$

i) Denotamos con $R(A)$ el rango de la matriz A. Se ha demostrado que si $R(A) = 4$, existe un único $\widehat{x} \in \mathbb{R}^4$ solución de

$$A^T A\overrightarrow{x} = A^T\overrightarrow{w}, \tag{4}$$

y que minimiza el funcional $J(\overrightarrow{x}) = \|\overrightarrow{w} - A\overrightarrow{x}\|^2$, $\overrightarrow{x} \in \mathbb{R}^4$.

Observación. Si $A = (a_{ij}) \in M_{m\times n}[\mathbb{R}]$ y $R(A) = n, \exists \widehat{x} \in \mathbb{R}^n$ tal que $A^T A\widehat{x} = A^T\overrightarrow{b}$. En nuestro caso $n = 4$.

Con la información suministrada en la tabla, se tiene

$$A = \begin{bmatrix} 1 & 0 & 0{,}25 & 4 \\ 1 & 1 & 0{,}5 & 6 \\ 1 & 2 & 1. & 10 \\ 1 & 3 & 1{,}5 & 20 \\ 1 & 4 & 2. & 25 \end{bmatrix}, \quad \overrightarrow{w} = \begin{bmatrix} 14{,}5 \\ 21 \\ 33 \\ 63 \\ 78 \end{bmatrix}.$$

Ponemos $B = A^T A$. Entonces

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0{,}25 & 0{,}5 & 1 & 1{,}5 & 2 \\ 4 & 6 & 10 & 20 & 25 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0{,}25 & 4 \\ 1 & 1 & 0{,}5 & 6 \\ 1 & 2 & 1. & 10 \\ 1 & 3 & 1{,}5 & 20 \\ 1 & 4 & 2. & 25 \end{bmatrix} = \begin{bmatrix} 5 & 10 & 5{,}25 & 65 \\ 10 & 30 & 15 & 186 \\ 5{,}25 & 15 & 7{,}5625 & 94 \\ 65 & 186 & 94 & 1177 \end{bmatrix}$$

Sea $\overrightarrow{b} = A^T\overrightarrow{w}$. Se tiene

$$\overrightarrow{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0{,}25 & 0{,}5 & 1 & 1{,}5 & 2 \\ 4 & 6 & 10 & 20 & 25 \end{bmatrix} \begin{bmatrix} 14{,}5 \\ 21 \\ 33 \\ 63 \\ 78 \end{bmatrix} = \begin{bmatrix} 209{,}5 \\ 588 \\ 297{,}625 \\ 3724 \end{bmatrix}$$

El sistema normal de ecuaciones es $B\overrightarrow{x} = \overrightarrow{b}$ o en forma explícita:

$$\begin{bmatrix} 5 & 10 & 5{,}25 & 65 \\ 10 & 30 & 15 & 186 \\ 5{,}25 & 15 & 7{,}5625 & 94 \\ 65 & 186 & 94 & 1177 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 209{,}5 \\ 588 \\ 297{,}625 \\ 3724 \end{bmatrix}$$

ii) Apliquemos ahora el método de Householder.

Recordemos que la matriz de Householder tiene la forma $H = I - 2\overrightarrow{u}\,\overrightarrow{u}^T$, con $\overrightarrow{u} \in \mathbb{R}^n$ y $\|\overrightarrow{u}\| = 1$. Esta matriz es simétrica y ortogonal. Por otro lado, dado $\overrightarrow{v} \in \mathbb{R}^n$ con $\overrightarrow{v} \neq 0$, existen $\alpha \in \mathbb{R}$ y H tales que $H\overrightarrow{v} = \alpha\overrightarrow{e}_1$.

Consideramos nuevamente el sistema normal de ecuaciones arriba propuesto:

$$\begin{bmatrix} 5 & 10 & 5{,}25 & 65 \\ 10 & 30 & 15 & 186 \\ 5{,}25 & 15 & 7{,}5625 & 94 \\ 65 & 186 & 94 & 1177 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 209{,}5 \\ 588 \\ 297{,}625 \\ 3724 \end{bmatrix}.$$

Etapa 1

Sea $\overrightarrow{v}^T = (5,\ 10,\ 5{,}25,\ 65)$. Entonces $\|\overrightarrow{v}\| = 66{,}16315062$, $\alpha = -66{,}16315062$,

$r = \alpha\,(\alpha - v_1) = -66{,}16315062\,(-66{,}16315062 - 5) = 4708{,}378253$.

Luego

$$\overrightarrow{w} = \overrightarrow{v} - \alpha\overrightarrow{e}_1 = \begin{bmatrix} 5 \\ 10 \\ 2{,}25 \\ 65 \end{bmatrix} + 66{,}16315062 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 71{,}16315062 \\ 10 \\ 5{,}25 \\ 65 \end{bmatrix},$$

$$\begin{aligned} H^{(1)} &= I - \frac{1}{r}\overrightarrow{w}\,\overrightarrow{w}^T \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - 0{,}0002123873543 \begin{bmatrix} 71{,}16315062 \\ 10 \\ 5{,}25 \\ 65 \end{bmatrix} (71{,}16315062, 10, 5{,}25, 65) \\ &= \begin{bmatrix} -0{,}075570787 & -0{,}1511415328 & -0{,}07934930475 & -0{,}9824199634 \\ -0{,}1511415328 & 0{,}9787612646 & -0{,}0111503361 & -0{,}1380517803 \\ -0{,}07934930475 & -0{,}0111503361 & 0{,}9941460736 & -0{,}07247718465 \\ -0{,}9824199634 & -0{,}1380517803 & -0{,}07247718465 & 0{,}1026634281 \end{bmatrix} \end{aligned}$$

Se pone $B^{(1)} = H^{(1)}B$. Resulta

$$B^{(1)} = \begin{bmatrix} -66{,}16315063 & -189{,}2103064 & -95{,}6114252 & -1196{,}791557 \\ 0 & 2{,}006536434 & 0{,}8267341518 & 8{,}690318587 \\ 0 & 0{,}3034316251 & 0{,}1215354329 & 0{,}912417267 \\ 0 & 4{,}042300817 & 1{,}87388202 & 24{,}48707075 \end{bmatrix}$$

Además, $\overrightarrow{b}^{(1)} = H^{(1)}\overrightarrow{b}$, se tiene

$$\overrightarrow{b}^{(1)} = \begin{bmatrix} -3786{,}851578 \\ 26{,}42402386 \\ 2{,}797612564 \\ 73{,}75615504 \end{bmatrix}.$$

Etapa 2

Sea $\overrightarrow{v}^T = (0, 2{,}006536434, 0{,}3034316251, 4{,}042300817)$. Obtenemos $\|\overrightarrow{v}\| = 4{,}523102377$,

$\alpha = -4{,}523102377$, $r = \alpha(\alpha - v_1) = -4{,}523102377\,(-4{,}523102377 - 2{,}006536434) = 29{,}53422483$,

$$\vec{w} = \vec{v} - \alpha\vec{e}_2 = \begin{bmatrix} 0 \\ 2{,}006536434 \\ 0{,}3034316251 \\ 4{,}042300817 \end{bmatrix} + 4{,}523102377 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 6{,}529638811 \\ 0{,}3034316251 \\ 4{,}042300817 \end{bmatrix}$$

Se pone $H^{(2)} = I - \frac{1}{r}\vec{w}\,\vec{w}^T$. Entonces

$$\begin{aligned} H^{(2)} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - 0{,}03385902307 \begin{bmatrix} 0 \\ 6{,}529638811 \\ 0{,}3034316251 \\ 4{,}042300817 \end{bmatrix} \\ &\qquad (0, 6{,}529638811, 0{,}3034316251, 4{,}042300817) \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -0{,}4436195038 & -0{,}06708484571 & -0{,}8937009334 \\ 0 & -0{,}06708484571 & 0{,}9968825743 & -0{,}04153018787 \\ 0 & -0{,}8937009334 & -0{,}04153018787 & 0{,}4467369301 \end{bmatrix}. \end{aligned}$$

Se define $B^{(2)} = H^{(2)}B^{(1)}$. Se tiene

$$B^{(2)} = \begin{bmatrix} -66{,}16315063 & -189{,}2103064 & -95{,}6114252 & -1196{,}791557 \\ 0 & -4{,}523102374 & -2{,}049500383 & -25{,}80052218 \\ 0 & 0 & -0{,}01212288182 & -0{,}6903684562 \\ 0 & 0 & 0{,}09318268742 & 3{,}13484012 \end{bmatrix}$$

Además, $\vec{b}^{(2)} = H^{(2)}\vec{b}^{(1)}$. Obtenemos

$$\vec{b}^{(2)} = \begin{bmatrix} -3786{,}851578 \\ -77{,}82583436 \\ -2{,}046867324 \\ 9{,}218238109 \end{bmatrix}.$$

Etapa 3

Sea $\vec{v} = \begin{bmatrix} 0 \\ 0 \\ -0{,}01212288182 \\ -0{,}09318268742 \end{bmatrix}$, $\|\vec{v}\| = 0{,}09396795996$, $\alpha = 0{,}09396795996$,

$r = \alpha(\alpha - v_1) = 0{,}09396795996\,(0{,}09396795996 + 0{,}01212288182) = 0{,}009969139979$.

$$\vec{w} = \vec{v} - \alpha\vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ -0{,}01212288182 \\ 0{,}09318268742 \end{bmatrix} - 0{,}09396795996 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -0{,}1060908418 \\ 0{,}09318268742 \end{bmatrix}$$

Se define $H^{(3)} = I - \frac{1}{r}\vec{w}\,\vec{w}^T$. Luego

$$H^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -0{,}1290108007 & 0{,}991643188 \\ 0 & 0 & 0{,}991643188 & 0{,}129010802 \end{bmatrix}.$$

Además, $B^{(3)} = H^{(3)}B^{(2)}$, con lo que

$$B^{(3)} = \begin{bmatrix} -66{,}16315063 & -189{,}2103064 & -95{,}6114252 & -1196{,}791557 \\ 0 & -4{,}523102374 & -2{,}049500383 & -25{,}80052218 \\ 0 & 0 & 0{,}09396795991 & 3{,}197707838 \\ 0 & 0 & 0 & -0{,}2801709388 \end{bmatrix},$$

$$\overrightarrow{b}^{(3)} = H^{(3)} \overrightarrow{b}^{(2)} = \begin{bmatrix} -3786{,}851578 \\ -77{,}82583436 \\ 9{,}405271019 \\ -0{,}8405097475 \end{bmatrix}$$

Resolución del sistema triangular superior:

$$\begin{aligned} d &= \frac{-0{,}8405097475}{-0{,}2801709388} = 2{,}999989046, \\ c &= \frac{9{,}405271019 - 3{,}197707838 \times d}{0{,}09396795991} = -1{,}998739434, \\ b &= \frac{-77{,}82583436 + 2{,}049500383 \times c + 25{,}80052218}{-4{,}523102374} = 0{,}9995280443, \\ a &= \frac{-3786{,}851578 + 189{,}2103064 \times b + 95{,}6114252 \times c + 1196{,}791557}{-66{,}16315063} = 2{,}999726189 \end{aligned}$$

La solución es:

$$\overrightarrow{x}^T = (2{,}999726189, 0{,}9995280443, -1{,}998739434, 2{,}999989046)\,.$$

9.6. Ajuste de datos con funciones dependientes de un parámetro

Sea $S = \left\{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots n\right\}$ un conjunto de datos experimentales. Supongamos que se busca una función real f dependiente de la variable $x \in \mathbb{R}$ y del parámetro $a \in \mathbb{R}$ a determinar según el conjunto de datos S, esto es,

$$y_i = f(x_i, a) + r_i \quad i = 1, \ldots, n,$$

donde r_i es una perturbación de y_i que depende del parámetro a, escribiremos $r_i(a)$. En forma matricial este sistema de ecuaciones se escribe como

$$\overrightarrow{y} = \overrightarrow{b}(a) + \overrightarrow{r}(a)\,,$$

donde $\overrightarrow{y}, \overrightarrow{b}(a), \overrightarrow{r}(a) \in \mathbb{R}^n$ definimos a continuación

$$\overrightarrow{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \overrightarrow{b}(a) = \begin{bmatrix} f(x_1, a) \\ \vdots \\ f(x_n, a) \end{bmatrix}, \quad \overrightarrow{r}(a) = \begin{bmatrix} r_1(a) \\ \vdots \\ r_n(a) \end{bmatrix}.$$

Resulta

$$\overrightarrow{r}(a) = \overrightarrow{y} - \overrightarrow{b}(a) = \begin{bmatrix} y_1 - f(x_1, a) \\ \vdots \\ y_n - f(x_n, a) \end{bmatrix}.$$

La determinación del parámetro a y del vector $\overrightarrow{r}(a)$, en general es complicada. recurrimos entonces a aproximar el parámetro a mediante el método de mínimos cuadrados. Para el efecto definimos una función real E como sigue:

$$E(a) = \|\overrightarrow{r}(a)\|^2 = \left\|\overrightarrow{y} - \overrightarrow{b}(a)\right\|^2 = \sum_{i=1}^{n} (y_i - f(x_i, a))^2 \quad a \in \mathbb{R}$$

y consideramos el problema siguiente:

$$\text{hallar } \widehat{a} \in \mathbb{R} \text{ solución de } \min_{a \in \mathbb{R}} E(a)\,,$$

que a su vez es equivalente al siguiente:

$$\text{hallar } \widehat{a} \in \mathbb{R} \text{ tal que } E(\widehat{a}) \leq E(a) \quad \forall a \in \mathbb{R}.$$

Escribiremos $E(\widehat{a}) = \min\limits_{a\in\mathbb{R}} E(a)$.

Supondremos que la función f es diferenciable y que $\dfrac{\partial f}{\partial a}(x,a)$ es continua. De las condiciones necesarias de extremo se tiene que $E'(a)=0$. Luego

$$E'(a) = \frac{d}{da}\sum_{i=1}^{n}(y_i - f(x_i,a))^2 = \sum_{i=1}^{n} -2(y_i - f(x_i,a))\frac{\partial f}{\partial a}(x_i,a)$$

con lo que

$$E'(a)=0 \Leftrightarrow \sum_{i=1}^{n}(y_i - f(x_i,a))\frac{\partial f}{\partial a}(x_i,a) = 0.$$

Note que la función f depende de x y de a.

Se define $\varphi(a) = \sum\limits_{i=1}^{n}(y_i - f(x_i,a))\dfrac{\partial f}{\partial a}(x_i,a)$ y consideramos la ecuación:

$$\text{hallar } \widehat{a}\in\mathbb{R} \text{ tal que } \varphi(\widehat{a}) = 0.$$

La ecuación precedente, en general, es no lineal. Su resolución es a menudo muy complicada y se recurre a la aplicación de métodos iterativos de resolución de ecuaciones no lineales que han sido estudiados anteriormente, estre ellos, el método de bisección, punto fijo modificado y método de Newton son los más utilizados y que están relacionados con la regularidad de la función f.

A continuación presentamos algunas funciones f que son muy utilizadas:

1. $f(x,a) = \lambda\exp(ax) \quad x\in\mathbb{R}$,
2. $f(x,a) = \lambda\,\mathrm{sen}(ax) \quad x\in\mathbb{R}$,
3. $f(x,a) = \lambda\cos(ax) \quad x\in\mathbb{R}$,
4. $f(x,a) = \lambda + x^a \quad x\in\mathbb{R}$,
5. $f(x,a) = \lambda\arctan(ax) \quad x\in\mathbb{R}$,

donde $\lambda\in\mathbb{R}$ es una constante conocida y $a\in\mathbb{R}$ es el parámetro a determinar.

Para calcular una aproximación del parámetro a, para algunas funciones f como las indicadas, recurriremos a proponer un problema alternativo en el que se ha linealizado mediante algún procedimiento particular. Aclararemos esta situación con ejemplos.

Ejemplos

1. Con el propósito de realizar comparaciones consideramos la función g definida como $g(x) = 5{,}2\exp(-0{,}282x) \quad x\geq 0$ y con esta función g obtenemos el conjunto de datos S siguiente:

   $$S = \{(0,5{,}2),(1{,}5,3{,}406),(4,1{,}683),(8{,}5,0{,}473),(10,0{,}31)\}.$$

   Buscamos una función f definida como $f(x,a) = 5{,}2\exp(ax) \quad x\geq 0$, con a un parámetro a determinar como solución de la ecuación

   $$\begin{aligned}\sum_{i=1}^{n}(y_i - f(x_i,a))\frac{\partial f}{\partial a}(x_i,a) &= 0 \Longleftrightarrow\\ \sum_{i=1}^{n}(y_i - 5{,}2\exp(ax_i))(5{,}2x_i\exp(ax_i)) &= 0 \Longleftrightarrow\\ \sum_{i=1}^{n}(y_i - 5{,}2\exp(ax_i))\,x_i\exp(ax_i) &= 0.\end{aligned}$$

Este tipo de ecuaciones son muy laboriosas. Otra alternatica ampliamente utilizada es la siguiente. Ponemos $y = 5{,}2 \exp(ax)$ entonces $\ln(y) = ax + \ln(5{,}2)$. Denotamos con $u = \ln(y)$ y $b = \ln(5{,}2) \simeq 1{,}648658626$, $u_i = \ln(y_i) \quad i = 1, \ldots, n$.

El problema alternatico a resolver es el siguiente. Se define

$$E_1(a) = \sum_{i=1}^{n} (u_i - ax_i - b)^2$$

y de las condiciones de extremo se tiene

$$\begin{aligned} E_1'(a) &= 0 \Leftrightarrow \sum_{i=1}^{n} (u_i - ax_i - b) x_i = 0 \Leftrightarrow \sum_{i=1}^{n} u_i x_i - a \sum_{i=1}^{n} x_i^2 - b \sum_{i=1}^{n} x_i = 0 \\ &\Leftrightarrow \quad a = \frac{\sum\limits_{i=1}^{n} y_i x_i - b \sum\limits_{i=1}^{n} x_i}{\sum\limits_{i=1}^{n} x_i^2}. \end{aligned}$$

Calculamos $u_1 = \ln(5{,}2)$, $u_2 = \ln(3{,}406)$, $u_3 = \ln(1{,}683)$, $u_4 = \ln(0{,}473)$ y $u_5 = \ln(0{,}31)$. Además $\sum\limits_{i=1}^{n} x_i = 24$, $\sum\limits_{i=1}^{n} x_i^2 = 190{,}5$, $\sum\limits_{i=1}^{n} u_i x_i = -14{,}15481935$, resulta

$$a = \frac{\sum\limits_{i=1}^{n} y_i x_i - b \sum\limits_{i=1}^{n} x_i}{\sum\limits_{i=1}^{n} x_i^2} = \frac{-14{,}15481935 - 1{,}648658626 \times 24}{190{,}5} = -0{,}2820085373$$

que redondeando se tiene $a = -0{,}282$.

La función buscada es por lo tanto $f(x, a) = 5{,}2 \exp(-0{,}282) = g(x) \quad x \geq 0$. En la figura siguiente se muestran los puntos del conjunto S y la gráfica de la función f.

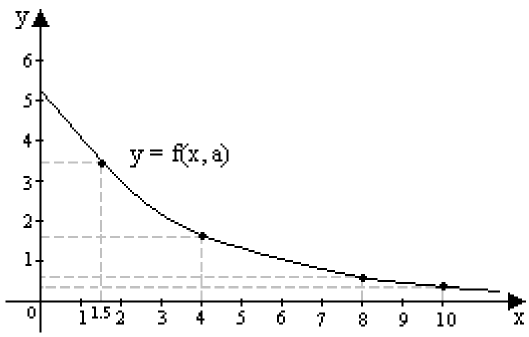

Figura 85

2. Nuevamente, con el propósito de realizar comparaciones, consideramos la función g definida como $g(x) = 3 \operatorname{sen}(2{,}5x) \quad x \in [0, \pi]$ y S el conjunto de datos siguiente:

$$S = \{(0,0), (0{,}52, 2{,}9), (0{,}78, 2{,}79), (1, 1{,}8), (1{,}57, -2{,}12), (2{,}1, -2{,}57), (2{,}62, 0{,}8), (\pi, 0)\}.$$

Buscamos una función f de la forma $f(x, a) = 3 \operatorname{sen}(ax) \quad x \in [0, \pi]$. De acuerdo a los resultados obtenidos anteriormente, para determinar el parámetro $a \in \mathbb{R}$ con el método de mínimos cuadrados debemos resolver la ecuación

$$\sum_{i=1}^{n} (y_i - f(x_i, a)) \frac{\partial f}{\partial a}(x_i, a) = 0.$$

Como $f(x,a) = 3\,\text{sen}(ax)$ se sigue que $\frac{\partial f}{\partial a}(x,a) = 3x\cos(ax)$, con lo que la ecuación a resolver.se escribe como

$$\sum_{i=1}^{n} (y_i - f(x_i,a))\, x_i \cos(ax_i) = 0,$$

ecuación que es muy laboriosa de resolver.

Por otro lado, fijado $a \in \mathbb{R}$, la función f es solución de la ecuación diferencial $u''(x) + a^2 u(x) = 0$ $\forall x \in \mathbb{R}$ pues $\frac{\partial^2 f}{\partial x^2}(x,a) = -a^2 f(x,a)$. Es esta ecuación diferencial aparece axplícitamente el parámetro $a \in \mathbb{R}$ y no figuran las funciones $\text{sen}(ax)$ y $\cos(ax)$. Por lo tanto el problema alternatico a resolver se construye como sigue:

$$y_i'' - a^2 y_i = r_i \quad i = 2, \ldots, n-1$$

donde y_i'' es la aproximación de la derivada segunda de f respecto de x, aproximación que se le calcula con diferencias finitas centrales estudiadas en el capítulo II.

Se define

$$\overrightarrow{w} = \begin{bmatrix} y_2'' \\ \vdots \\ y_{n-1}'' \end{bmatrix}, \quad \overrightarrow{y} = \begin{bmatrix} y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}, \quad \overrightarrow{r}(a) = \begin{bmatrix} r_2(a) \\ \vdots \\ r_{n-1}(a) \end{bmatrix},$$

con lo que la relación precedente se escribe como

$$\overrightarrow{w} - a^2 \overrightarrow{y} = r(\overrightarrow{a}),$$

y definimos la función $E_1(a)$ como

$$E_1(a) = \|\overrightarrow{r}(a)\|^2 = \left\|\overrightarrow{w} - a^2\overrightarrow{y}\right\|^2 = \sum_{i=2}^{n-1} \left(y_i'' - a^2 y_i\right)^2,$$

donde $\|\cdot\|$ es la norma euclídea en $\mathbb{R}^{n-3}$.

El problema alternativo que se propone es el siguiente:

$$\text{hallar } \widehat{a} \in \mathbb{R} \text{ tal que } E_1(\widehat{a}) = \min_{a \in \mathbb{R}} E_1(a),$$

que por las condiciones necesarias de extremo conduce a resolver la ecuación $E_1'(a) = 0$. Así,

$$E_1'(a) = \sum_{i=3}^{n-2} \left(y_i'' - a^2 y_i\right) a y_i = 0 \Leftrightarrow \sum_{i=3}^{n-2} y_i'' y_i - a^2 y_i^2 = 0 \Leftrightarrow a^2 = \frac{\sum_{i=3}^{n-2} y_i'' y_i}{\sum_{i=3}^{n-2} y_i^2}.$$

Queda calcular y_i'' $\; i = 2, \ldots, n-1$, $\sum_{i=3}^{n-2} y_i'' y_i$, $\sum_{i=3}^{n-2} y_i^2$.

Recordemos que la derivada primera se aproxima con diferencias finitas centrales mediante el cociente

$$y_i' = \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}} \quad i = 2, \ldots, n-1,$$

y la derivada segunda se aproxima con diferencias finitas centrales mediante el cociente

$$y_i'' = \frac{y_{i+1}' - y_{i-1}'}{x_{i+1} - x_{i-1}} \quad i = 3, \ldots, n-2..$$

Se propone como ejercicio realizar los cálculos para obtener el valor de a.

9.7. Mínimos cuadrados continuos.

Sean $V = C([a,b])$ el espacio de funciones continuas provisto del producto escalar

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx \quad \forall f,g \in C([a,b]).$$

Sea $\mathbb{K}_m[\mathbb{R}]$ el espacio de polinomios de grado $\leq m$. Se tiene $\dim \mathbb{K}_m[\mathbb{R}] = m+1$ y $\mathbb{K}_m[\mathbb{R}]$ es un subespacio cerrado de $C([a,b])$. Sea $\{\varphi_0, \ldots, \varphi_m\}$ una base de $\mathbb{K}_m[\mathbb{R}]$. Entonces

$$\mathbb{K}_m[\mathbb{R}] = \left\{\sum_{i=o}^{m} \alpha_i\varphi_i \mid \alpha_i \in \mathbb{R},\ i = 0,\ldots,m\right\}.$$

Sea $f \in C([a,b])$. Existe $\hat{g} \in \mathbb{K}_m[\mathbb{R}]$ tal que

$$\|f-\hat{g}\| = \underset{g\in\mathbb{K}_m[\mathbb{R}]}{\mathrm{Min}} \|f-g\|, \qquad \langle f-\hat{g}, w\rangle = 0 \quad \forall w \in \mathbb{K}_m[\mathbb{R}].$$

Note que $\hat{g} \in \mathbb{K}_m[\mathbb{R}] \Longleftrightarrow \exists \hat{a}_0, \ldots, \hat{a}_m \in \mathbb{R}$ tales que $\hat{g} = \sum\limits_{i=o}^{m} \hat{a}_i\varphi_i$. Además,

$$\|f-\hat{g}\|^2 = \underset{g\in\mathbb{K}_m[\mathbb{R}]}{\mathrm{Min}} \|f-g\|^2,$$

$$\langle f-\hat{g}, w\rangle = 0 \quad \forall w \in \mathbb{K}_m[\mathbb{R}] \Longleftrightarrow \int_a^b \left(f(x) - \sum_{i=o}^{m} \hat{a}_i\varphi_i(x)\right) w(x)dx = 0 \quad \forall a_0,\ldots,a_m \in \mathbb{R}.$$

Poniendo sucesivamente $w = \varphi_i \quad i = 1, \cdots, n$ se obtiene el sistema de ecuaciones siguiente:

$$\int_a^b f(x)\varphi_j(x)dx - \sum_{i=o}^{m} \hat{a}_i \int_a^b \varphi_i(x)\varphi_j(x)dx = 0 \quad j = 0,\ldots,m,$$

o en forma explícita dicho sistema tiene la forma

$$\begin{cases} \sum\limits_{i=o}^{m} \hat{a}_i \int_a^b \varphi_i(x)\varphi_0(x)dx = \int_a^b f(x)\varphi_0(x)dx, \\ \qquad\qquad\vdots \\ \sum\limits_{i=o}^{m} \hat{a}_i \int_a^b \varphi_i(x)\varphi_m(x)dx = \int_a^b f(x)\varphi_m(x)dx, \end{cases}$$

que en forma matricial se expresa como $A\overrightarrow{x} = \overrightarrow{b}$, donde $A = (a_{ij})$ es la matriz definida como $a_{ij} = \int_a^b \varphi_i(x)\varphi_j(x)dx$, $\overrightarrow{b} = (b_0,\ldots,b_m)$ con $b_j = \int_a^b f(x)\varphi_j(x)dx \quad j = 0,\ldots,m$, $\overrightarrow{x}^T = (\hat{a}_0,\ldots,\hat{a}_m)$ es el vector de las incógnitas..La matriz A es simétrica, definida positiva por lo tanto invertible.

9.8. Aproximación numérica de series de Fourier

9.8.1. Preliminares

Sean $L > 0$, a_0, a_k, $b_k \in \mathbb{R}$ con $k = 1, 2, \cdots$. Las series de funciones de la forma

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right)\right)$$

se llaman series de Fourier y aparecen en algunas aplicaciones de la matemática tales como el procesamiento de la señal y de imágenes, en la resolución de algunas clases de ecuaciones en derivadas parciales. Nos interesamos en la aproximación numérica de esta clase de series de funciones. Para ello primeramente introducimos algunas notaciones.

Sean $L > 0$. Se denota con $C([-L, L])$ al espacio vectorial de las funciones continuas en $[-L, L]$. Proveemos a $C([-L, L])$ del producto escalar notado $\langle \cdot, \cdot \rangle$ y definido como (véase en el apéndice los espacios con producto interior)

$$\langle u, v\rangle = \int_{-L}^{L} u(x)\, v(x)\, dx \quad \forall u, v \in C([-L, L]),$$

y la norma asociada a este producto escalar se nota $\| \cdot \|$ y se define como

$$\| u \| = \left(\int_{-L}^{L} | u(x) |^2\, dx \right)^{\frac{1}{2}} \quad \forall u \in C([-L, L]).$$

Definición 1 *Sean $L > 0$ y u una función real definida en todo $\mathbb{R}$. Se dice que u es periódica de período $2L$ si y solo si se verifica*

$$u(x + 2L) = u(x) \quad \forall x \in \mathbb{R}.$$

En la figura siguiente se muestra la gráfica de una función u periódica de período $2L$.

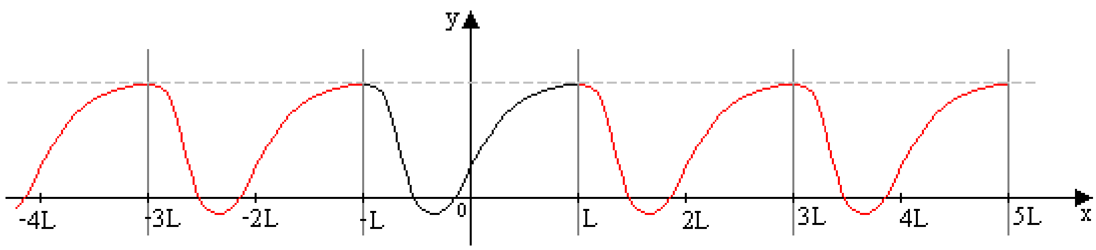

Figura 86

Observe que para $x = -L$ se tiene $u(L) = u(-L)$.

Sea $u \in C([-L, L])$. A la función u lo extendemos a todo $\mathbb{R}$ por periodicidad y por abuso de lenguaje lo notamos aún con u, es decir que

$$u(x + 2L) = u(x) \quad \forall x \in \mathbb{R}.$$

Definición 2 *Sea $u \in C([-L, L])$.*

i) Se dice que u es par si y solo si u verifica la propiedad $u(-x) = u(x) \quad \forall x \in [-L, L]$.

ii) Se dice que u es impar si y solo si verifica la propiedad $u(-x) = -u(x) \quad \forall x \in [-L, L]$.

En la izquierda de la figura siguiente se muestra la gráfica de una función par; y, a la derecha está representada la gráfica de una función impar.

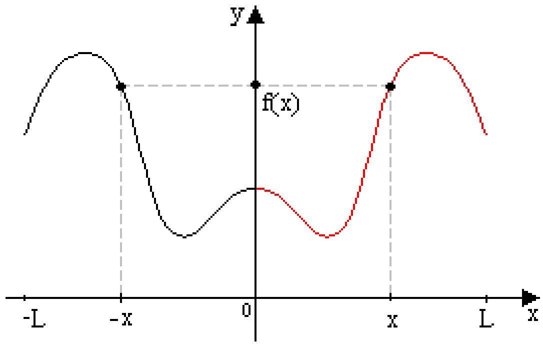

Figura 87

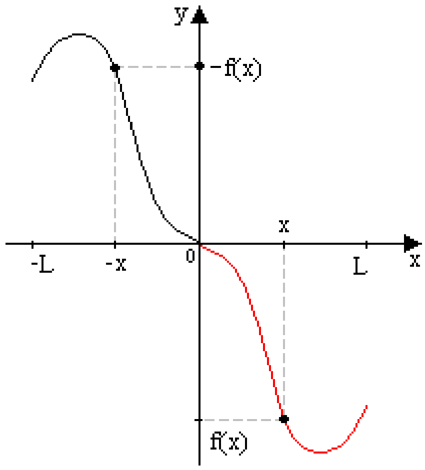

Figura 88

Sea $u \in C([-L, L])$. Se verifica inmediatamente las siguientes propiedades.

i) Si u es impar, $\displaystyle\int_{-L}^{L} u(x)\,dx = 0.$

ii) Si u es par, $\displaystyle\int_{-L}^{L} u(x)\,dx = 2\int_{0}^{L} u(x)\,dx.$

iii) La función f definida como $f(x) = \dfrac{1}{2}(u(x) + u(-x)) \quad x \in [-L, L]$ es par.

iv) La función g definida como $g(x) = \dfrac{1}{2}(u(x) - u(-x)) \quad x \in [-L, L]$ es impar.

Además $u(x) = f(x) + g(x) \quad \forall x \in [-L, L]$.

Definición 3

i) Sean $u, v \in C([-L, L])$. Se dice que u es ortogonal o perpendicular a v, que se nota $u \perp v$, si y solo si $\langle u, v\rangle = 0$.

ii) Sean $u \in C([-L, L])$, $M \subset C([-L, L])$ con $M \neq \phi$. Se dice que u es ortogonal a M, que se escribe $u \perp M$, si y solo si $\langle u, v\rangle = 0 \quad \forall v \in M$.

iii) Sean M, N dos subconjuntos no vacíos de $C([-L, L])$. Se dice que M es ortogonal a N, que se nota $M \perp N$, si y solo si $\langle u, v\rangle = 0 \quad \forall u \in M, \forall v \in N$.

iv) Sea $M \subset C([-L, L])$ con $M \neq \emptyset$. Se dice que M es ortogonal si y solo si $\langle u, v\rangle = 0 \quad \forall u, v \in M,\ u \neq v$.

v) Se dice que M es ortonomal si y solo si M es ortogonal, y $\forall u \in M$, $\|u\| = 1$.

Los siguientes conjuntos de funciones son muy importantes en el desarrollo en series de Fourier de funciones reales periódicas de período $2L$ y continuas a trozos en el intervalo $[-L, L]$. Sean M, N los subconjuntos de $C([-L, L])$ definidos como $M = \{\varphi_k \mid k \in \mathbb{N}\}$, $N = \{\psi_k \mid k \in \mathbb{Z}^+\}$, donde

$$\begin{aligned}
\varphi_0(x) &= 1 \quad x \in [-L, L], \\
\varphi_k(x) &= \cos\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L], \quad k = 1, 2, \ldots, \\
\psi_k(x) &= \operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L], \quad k = 1, 2, \ldots
\end{aligned}$$

Se tiene

i) M es un conjunto ortogonal.

ii) N es un conjunto ortogonal.

iii) $M \perp N$.

Sea $f \in C([-L,L])$. Supongamos que f se prolonga por periodicidad a todo $\mathbb{R}$ con periódo $2L$, esto es,

$$f(x+2L) = f(x) \quad \forall x \in \mathbb{R}.$$

Admitimos que la función f se representa como una serie de Fourier, es decir que se verifica

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right)\right) \quad x \in [-L,L],$$

donde $a_0,\ a_k,\ b_k \in \mathbb{R} \quad k = 1,2,\dots$, son los coeficientes de Fourier definidos a continuación:

$$\begin{aligned} a_0 &= \frac{1}{L}\int_{-L}^{L} f(x)\,dx, \\ a_k &= \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{k\pi x}{L}\right)dx \quad k = 1,2,\dots, \\ b_k &= \frac{1}{L}\int_{-L}^{L} f(x)\operatorname{sen}\left(\frac{k\pi x}{L}\right)dx \quad k = 1,2,\dots. \end{aligned}$$

Si la función f es par, se tiene que $f(x)\operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L,L]$ es impar y en consecuencia $b_k = 0$ $k = 1,2,\dots$. Resulta que la serie de Fourier de f se escribe como

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi x}{L}\right) \quad x \in [-L,L],$$

y

$$a_k = \frac{2}{L}\int_0^L f(x)\cos\left(\frac{k\pi x}{L}\right)dx \quad k = 0,1,\dots.$$

Si la función f es impar, se tiene $f(x)\cos\left(\frac{k\pi x}{L}\right)dx \quad x \in [-L,L]$ impar. Luego $a_k = 0 \quad k = 0,1,\dots,$ con lo que la serie de Fourier de f se escribe como sigue:

$$f(x) = \sum_{k=1}^{\infty} b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L,L],$$

con

$$b_k = \frac{2}{L}\int_0^L f(x)\operatorname{sen}\left(\frac{k\pi x}{L}\right)dx \quad k = 1,2,\dots.$$

Definición 4 *Sean $m \in \mathbb{Z}^+$. La función $Q_m \in C([-L,L])$ definida como*

$$Q_m(x) = \frac{a_0}{2} + \sum_{k=1}^{m} \left(a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right)\right) \quad x \in [-L,L]$$

se llama polinomio trigonométrico, donde $a_0, a_k, b_k \in \mathbb{R} \quad k = 1,\dots,m$.

Sean $u_0(x) = \frac{a_0}{2}$, $u_k(x) = a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L,L]$. A las funciones u_0, u_k $k = 1,\dots,m$ se les denomina armónicos de Q_m.

En el caso en que a_0, a_k, b_k son los coeficientes de Fourier, el polinomio trogonométrico Q_m se le denomina polinomio trogonométrico de Fourier.

En lo que sigue, nos interesamos en los polinomios trigonométricos de Fourier siguientes:

i) Si $f \in C([-L,L])$ es par, $Q_m(x) = \dfrac{a_0}{2} + \sum\limits_{k=1}^{m} a_k \cos\left(\dfrac{k\pi x}{L}\right) \quad x \in [-L,L]$,

donde $a_k = \dfrac{2}{L}\displaystyle\int_0^L f(x)\cos\left(\dfrac{k\pi x}{L}\right)dx \quad k = 0,1,\ldots,m.$

ii) Si $f \in C([-L,L])$ es impar, $Q_m(x) = \sum\limits_{k=1}^{m} b_k \operatorname{sen}\left(\dfrac{k\pi x}{L}\right) \quad x \in [-L,L]$,

donde $b_k = \dfrac{2}{L}\displaystyle\int_0^L f(x)\operatorname{sen}\left(\dfrac{k\pi x}{L}\right)dx \quad k = 1,\ldots,m.$

iii) Si $f \in ([-L,L])$ es arbitraria

$$Q_m(x) = \frac{a_0}{2} + \sum_{k=1}^{m}\left(a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right)\right) \quad x \in [-L,L]$$

y a_0, a_k, $b_k \quad k = 1,\ldots,m$ los coeficientes de Fourier arriba definidos.

Teorema 3 *Sean* $f \in C([-L,L])$, $m \in \mathbb{Z}^+$. *Se definen*

$$\begin{aligned} Q_m(x) &= \sum_{k=0}^{m} a_k \cos\left(\frac{k\pi x}{L}\right) \quad x \in [-L,L], \\ E(a_0,\ldots,a_m) &= \|f - Q_m\|^2 = \int_{-L}^{L}\left[f(x) - \left(a_0 + a_1\cos\left(\frac{\pi x}{L}\right) + \cdots + a_m \cos\left(\frac{m\pi x}{L}\right)\right)\right]^2 dx. \end{aligned}$$

Existen $\widehat{a}_0,\ldots,\widehat{a}_m \in \mathbb{R}$ *tales que* $E(\widehat{a}_0,\ldots,\widehat{a}_m) = \min\limits_{(a_0,\ldots,a_m)\in\mathbb{R}^{m+1}} E(a_0,\ldots,a_m)$, *y* $\widehat{a}_j \quad j = 1,\ldots,m$ *los coeficientes de Fourier definidos como* $a_0 = \dfrac{1}{2L}\int_{-L}^{L} f(x)\,dx$ *y* $a_j = \dfrac{1}{L}\int_{-L}^{L} f(x)\cos\left(\dfrac{j\pi x}{L}\right)dx$ $j = 1,\ldots,m$.

Demostración. De las condiciones necesarias de extremo, se tiene

$$\nabla E(a_0,\ldots,a_m) = \overrightarrow{0} \Leftrightarrow \frac{\partial E}{\partial a_j}(a_0,\ldots,a_m) = 0 \quad j = 0,1,\ldots,m.$$

Además, de la definición del funcional E, se tiene

$$\begin{aligned} \frac{\partial E}{\partial a_j}(a_0,\ldots,a_m) &= \frac{\partial}{\partial a_j}\int_{-L}^{L}\left[f(x) - \left(a_0 + a_1\cos\left(\frac{\pi x}{L}\right) + \cdots + a_m\cos\left(\frac{m\pi x}{L}\right)\right)\right]^2 dx \\ &= \int_{-L}^{L} 2\left[f(x) - \left(a_0 + a_1\cos\left(\frac{\pi x}{L}\right) + \cdots + a_m\cos\left(\frac{m\pi x}{L}\right)\right)\left(-\cos\left(\frac{j\pi x}{L}\right)\right)\right]dx \\ &= -2\int_{-L}^{L}\left[f(x)\cos\left(\frac{j\pi x}{L}\right) - \left(a_0\cos\left(\frac{j\pi x}{L}\right) + \cdots + a_m\cos\left(\frac{m\pi x}{L}\right)\cos\left(\frac{j\pi x}{L}\right)\right)\right]dx \end{aligned}$$

Luego,

$$\begin{aligned} \frac{\partial E}{\partial a_j}(a_0,\ldots,a_m) &= 0 \Leftrightarrow a_0\int_{-L}^{L}\cos\left(\frac{j\pi x}{L}\right)dx + \cdots + a_m\int_{-L}^{L}\cos\left(\frac{m\pi x}{L}\right)\cos\left(\frac{j\pi x}{L}\right)dx = \\ &= \int_{-L}^{L} f(x)\cos\left(\frac{j\pi x}{L}\right)dx \quad j = 0,1,\ldots,m. \end{aligned}$$

Por otro lado, como el conjunto de funciones $\{\varphi_j \mid j = 0,1,\ldots,m\}$ con $\varphi_0(x) = 1$, $\varphi_j(x) = \cos\left(\dfrac{j\pi x}{L}\right)$ $x \in [-L,L]$, $\quad j = 1,\ldots,m$, es un conjunto ortogonal, entonces $\langle\varphi_k,\varphi_j\rangle = 0$ si $j \neq k$ y $\langle\varphi_k,\varphi_j\rangle = \|\varphi_k\|^2$

si $j = k$. Además $\|\varphi_0\|^2 = 2L$, y $\|\varphi_j\|^2 = L \quad j = 1, \ldots, m$. Por lo tanto, el sistema de ecuaciones precedente se reduce al siguiente:

$$\begin{aligned}
\text{para } j &= 0, \quad 2La_0 = \int_{-L}^{L} f(x)\, dx \Rightarrow a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)\, dx, \\
\text{para } j &= 1, \quad a_1 L = \int_{-L}^{L} f(x) \cos\left(\frac{\pi x}{L}\right) dx \Rightarrow a_1 = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{\pi x}{L}\right) dx,
\end{aligned}$$

así sucesivamente, para $j = m$ obtenemos $a_m = \dfrac{1}{L}\displaystyle\int_{-L}^{L} f(x)\cos\left(\frac{m\pi x}{L}\right) dx$. ■

Claramente, los $a_k \quad k = 0, 1, \ldots, m$ son los coeficientes de Fourier, por lo tanto Q_m es el polinomio trigonométrico de Fourier de cosenos.

Sea $f \in C([-L, L])$, m$\in \mathbb{Z}^+$ y consideramos el conjunto de funciones $M = \{\psi_k \mid k = 1, \ldots, m\}$ con $\psi_k(x) = \operatorname{sen}\left(\dfrac{k\pi x}{L}\right) \quad x \in [-L, L], \quad k = 1, \ldots, m$. Este conjunto M es ortogonal. Se definen

$$\begin{aligned}
Q_m(x) &= \sum_{k=1}^{m} b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L], \\
E(b_1, \ldots, b_m) &= \|f - Q_m\|^2 = \int_{-L}^{L} (f(x) - Q_m(x))^2\, dx = \int_{-L}^{L} \left[f(x) - \sum_{k=1}^{m} b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right)\right]^2 dx
\end{aligned}$$

Procediendo en forma similar a la prueba del teorema precedente, se obtienen los coeficientes de Fourier $\widehat{b}_j \quad j = 1, \ldots, m$ definidos como $\widehat{b}_j = \dfrac{1}{L}\displaystyle\int_{-L}^{L} f(x) \operatorname{sen}\left(\frac{j\pi x}{L}\right) dx$, y

$$E\left(\widehat{b}_1, \ldots, \widehat{b}_m\right) = \min_{(b_1, \ldots, b_m) \in \mathbb{R}^m} E(b_1, \ldots, b_m).$$

Más aún, si combinamos los dos resultados precedentes con $f \in C([-L, L])$ y se definen

$$\begin{aligned}
Q_m(x) &= \sum_{k=0}^{m} a_k \cos\left(\frac{k\pi x}{L}\right) + \sum_{k=1}^{m} b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L], \\
E(a_0, \ldots, a_m, b_1, \ldots, b_m) &= \|f - Q_m\|^2 = \int_{-L}^{L} (f(x) - Q_m(x))^2\, dx
\end{aligned}$$

resulta que los coeficientes de Fourier

$$a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)\, dx,\ \widehat{a}_k = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{k\pi x}{L}\right) dx,\ \widehat{b}_k = \frac{1}{L}\int_{-L}^{L} f(x) \operatorname{sen}\left(\frac{k\pi x}{L}\right) dx \quad k = 1, 2, \ldots, m$$

minimizan el funcional $E(a_0, \ldots, a_m, b_1, \ldots, b_m)$.

Teorema 4 (Igualdad de Bessel) *Sean* $M = \{\varphi_k \mid k = 0, 1, \ldots, m\}$, *donde* $\varphi_k(x) = \cos\left(\dfrac{k\pi x}{L}\right)$ $x \in [-L, L], \quad k = 0, 1, \ldots, m$, *y* $f \in C([-L, L])$. *Entonces, para cada* $m \in \mathbb{Z}^+$ *se tiene*

$$\left\| f - \sum_{k=0}^{m} \frac{\langle f, \varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k \right\|^2 = \|f\|^2 - \sum_{k=0}^{m} \frac{|\langle f, \varphi_k\rangle|^2}{\|\varphi_k\|^2}.$$

Demostración. El conjunto M es ortogonal y de la definición de la norma asociada al producto escalar, se tiene

$$\left\| f - \sum_{k=0}^{m} \frac{\langle f, \varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k \right\|^2 = \left\langle f - \sum_{k=0}^{m} \frac{\langle f, \varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k, f - \sum_{k=0}^{m} \frac{\langle f, \varphi_j\rangle}{\|\varphi_j\|^2}\varphi_j \right\rangle$$

y de la linealidad del producto escalar respecto de cada variable (véase el apéndice, espacios con producto interior) resulta

$$
\begin{aligned}
\left\| f-\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k\right\|^2 &= \langle f,f\rangle-2\left\langle f,\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k\right\rangle+\left\langle \sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k,\sum_{j=1}^{m}\frac{\langle f,\varphi_j\rangle}{\|\varphi_j\|^2}\varphi_j\right\rangle\\
&= \|f\|^2-2\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\langle f,\varphi_k\rangle+\sum_{k=0}^{m}\sum_{j=1}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\frac{\langle f,\varphi_j\rangle}{\|\varphi_j\|^2}\langle\varphi_k,\varphi_j\rangle .
\end{aligned}
$$

Por hipótesis $\langle\varphi_k,\varphi_j\rangle=0$ si $k\neq j$ y $\langle\varphi_k,\varphi_k\rangle=\|\varphi_k\|^2$, entonces

$$
\sum_{k=0}^{m}\sum_{j=1}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\frac{\langle f,\varphi_j\rangle}{\|\varphi_j\|^2}\langle\varphi_k,\varphi_j\rangle=\sum_{k=0}^{m}\left(\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\right)^2\|\varphi_k\|^2=\sum_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^2}.
$$

Por lo tanto,

$$
\begin{aligned}
\left\| f-\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k\right\|^2 &= \|f\|^2-2\sum_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^2}+\sum_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^2}\\
&= \|f\|^2-\sum_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^2}.
\end{aligned}
$$

■

Observaciones

1. Puesto que $\left\| f-\sum\limits_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k\right\|^2\geq 0$ se sigue que $\|f\|^2-\sum\limits_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^2}\geq 0\Leftrightarrow\sum\limits_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^2}\leq\|f\|^2$ (desigualdad de Bessel).

   Por el teorema de Pitágoras, se tiene

   $$\left\|\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k\right\|^2=\sum_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^4}\|\varphi_k\|^2=\sum_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^2}.$$

   Por lo tanto, la igualdad y desigualdad de Bessel se escriben como sigue:

   $$
   \begin{aligned}
   \left\| f-\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k\right\|^2 &= \|f\|^2-\left\|\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k\right\|^2,\\
   \sum_{k=0}^{m}\frac{|\langle f,\varphi_k\rangle|^2}{\|\varphi_k\|^2} &= \left\|\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k\right\|^2\leq\|f\|^2 .
   \end{aligned}
   $$

   Note que

   $$Q_m(x)=\sum_{k=0}^{m}\widehat{a}_k\cos\left(\frac{k\pi x}{L}\right)dx=\sum_{k=0}^{m}\frac{\langle f,\varphi_k\rangle}{\|\varphi_k\|^2}\varphi_k(x)\qquad x\in[-L,L],$$

   el polinomio trigonométrico de Fourier de senos, con $\widehat{a}_k\quad k=1,\ldots,m$ los coeficientes de Fourier.

2. En forma similar a la precedente se obtiene con el conjunto ortogonal $N=\{\psi_k\mid k=1,\ldots,m\}$ con $\psi_k(x)=\operatorname{sen}\left(\frac{k\pi x}{L}\right)\quad x\in[-L,L],\quad k=1,\ldots,m$. Se tiene

   $$\left\| f-\sum_{k=1}^{m}\frac{\langle f,\psi_k\rangle}{\|\psi_k\|^2}\psi_k\right\|^2=\|f\|^2-\sum_{k=1}^{m}\frac{|\langle f,\psi_k\rangle|^2}{\|\psi_k\|^2}.$$

   En este caso

   $$\sum_{k=1}^{m}\frac{\langle f,\psi_k\rangle}{\|\psi_k\|^2}\psi_k=\sum_{k=1}^{m}\widehat{b}_k\operatorname{sen}\left(\frac{k\pi x}{L}\right)=Q_m(x),$$

   el polinomio trigonométrico de Fourier de senos, con $\widehat{b}_k\quad k=1,\ldots,m$ los coeficientes de Fourier.

9.8.2. Aproximación numérica

Para calcular valores aproximados de los coeficientes de Fourier, aplicamos el método de los trapecios.

Sea $n \in \mathbb{Z}^+$ y $\tau(n)$ una partición uniforme del intervalo $[0, L]$, esto es $h = \dfrac{L}{h}$ y $x_j = jh \quad j = 0, 1, \ldots, n$.

i) Supongamos que la función $f \in C([-L, L])$ es par, entonces el polinomio trigonométrico de Fourier Q_m está definido como

$$Q_m(x) = \frac{a_0}{2} + \sum_{k=1}^{m} a_k \cos\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L],$$

donde

$$a_k = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{k\pi x}{L}\right)dx = \frac{2}{L}\int_{0}^{L} f(x)\cos\left(\frac{k\pi x}{L}\right)dx \quad k = 0, 1, \ldots, m,$$

que se aproxima con la fórmula de los trapecios. Tenemos

$$\begin{aligned} a_k &= \frac{2}{L}\int_0^L f(x)\cos\left(\frac{k\pi x}{L}\right)dx \\ &\simeq \frac{2}{L}\left[\frac{h}{2}\left(f(0)\cos(0) + f(L)\cos(k\pi)\right) + h\sum_{j=1}^{n-1} f(x_j)\cos\left(\frac{k\pi x_j}{L}\right)\right] \\ &= \frac{1}{n}\left(f(0) + f(L)\cos(k\pi)\right) + \frac{2}{n}\sum_{j=1}^{n-1} f(x_j)\cos\left(\frac{k\pi j}{n}\right). \end{aligned}$$

Ponemos $y_j = f(x_j) \quad j = 0, 1, \ldots, n$. Entonces

$$\widehat{a}_k = \frac{1}{n}\left(y_0 + y_n\cos(k\pi)\right) + \frac{2}{n}\sum_{j=1}^{n-1} y_j\cos\left(\frac{k\pi j}{L}\right) \quad k = 0, 1, \ldots, m,$$

y se define $\widehat{Q}_m(x)$ como sigue:

$$\widehat{Q}_m(x) = \frac{a_0}{2} + \sum_{k=1}^{m} \widehat{a}_k\cos\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L],$$

$\widehat{Q}_m(x)$ es una aproximación de $Q_m(x)$ en $x \in [-L, L]$.

ii) Supongamos que la función $f \in C([-L, L])$ es impar, el polinomio trogonométrico de Fourier está definido como

$$Q_m(x) = \sum_{k=1}^{m} b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L],$$

con

$$b_k = \frac{1}{L}\int_{-L}^{L} f(x)\operatorname{sen}\left(\frac{k\pi x}{L}\right)dx = \frac{2}{L}\int_{0}^{L} f(x)\operatorname{sen}\left(\frac{k\pi x}{L}\right)dx \quad k = 1, 2, \ldots, m,$$

los mismos que se aproximan con la fórmula de los trapecios.

Resulta

$$\begin{aligned} b_k &= \frac{2}{L}\int_0^L f(x)\operatorname{sen}\left(\frac{k\pi x}{L}\right)dx \\ &\simeq \frac{2}{L}\left[\frac{h}{2}\left(f(0)\operatorname{sen}(0) + f(L)\operatorname{sen}(k\pi)\right) + h\sum_{j=0}^{n-1} f(x_j)\operatorname{sen}\left(\frac{k\pi x_j}{L}\right)\right]. \end{aligned}$$

Puesto que $\operatorname{sen}(k\pi) = 0 \quad k = 1, 2, \ldots, m$, se sigue que

$$b_k \simeq \frac{2}{n}\sum_{j=1}^{n-1} f(x_j)\operatorname{sen}\left(\frac{k\pi x_j}{L}\right) = \frac{2}{n}\sum_{j=1}^{n-1} f(x_j)\operatorname{sen}\left(\frac{k\pi j}{n}\right).$$

Nuevamente, se define $y_j = f(x_j) \quad j = 1, \ldots, n-1$, $b_k = \frac{2}{n}\sum\limits_{j=1}^{n-1} y_j \operatorname{sen}\left(\frac{k\pi j}{n}\right)$ y con estos coeficientes, el polinomio trigonométrico $\widehat{Q}_m(x) = \sum\limits_{k=1}^{m} \widehat{b}_k \operatorname{sen}\left(\frac{k\pi x}{L}\right) \quad x \in [-L, L]$.

Entonces $\widehat{Q}_m(x)$ es una aproximación de $Q_m(x)$

iii) Sea $f \in C([-L, L])$. Se define las funciones $u, v \in C([-L, L])$ como sigue:

$$\begin{cases} u(x) = \frac{1}{2}(f(x) + f(-x)) \\ v(x) = \frac{1}{2}(f(x) - f(-x)) \end{cases} \quad x \in [-L, L],$$

entonces u es par y v es impar. Se tiene $f = u + v$.

El polinomio trogonométrico de Fourier

$$\begin{aligned} Q_m(x) &= \frac{a_0}{2} + \sum_{k=1}^{m}\left(a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right)\right) \\ &= \frac{a_0}{2} + \sum_{k=1}^{m} a_k \cos\left(\frac{k\pi x}{L}\right) + \sum_{k=1}^{m} b_k \operatorname{sen}\left(\frac{k\pi x}{L}\right) \\ &= P_m(x) + R_m(x) \quad x \in [-L, L] \end{aligned}$$

donde $P_m(x) = \frac{1}{2}(Q_m(x) + Q_m(-x))$, $R_m(x) = \frac{1}{2}(Q_m(x) - Q_m(-x))$ con P_m par y R_m impar que se aproximan como en i) y ii) precedentes.

Para calcular $\widehat{Q}_m(x)$ se requiere de la siguiente información: número de términos del polinomio trigonométrico $Q_m(x)$, extremo derecho $L > 0$ del intervalo $[-L, L]$ par. Para calcular aproximaciones de los coeficientes mediante el método de los trapecios se requiere conocer el número de puntos n del intervalos $[0, L]$. Adicionalmente requerimos de una aproximación de π. Con esta información, proponemos el siguiente algoritmo de cálculo de $Q_m(x)$

Algoritmo

Datos de entrada: $m, n \in \mathbb{Z}^+$, L, $x \in \mathbb{R}$, función f.

Datos de salida: $\widehat{Q}_m(x)$.

1. $pi = 3{,}1415926536$.
2. $h = \frac{L}{n}$.
3. $y_0 = f(0)$.
4. $y_1 = f(L)$.
5. Para $k = 0, 1, \ldots, m$

    $S = 0$.

    Si $k = 0$ entonces

    para $j = 1, \ldots, n-1$

    $x_j = jh$

    $S = S + f(x_j)$

    fin bucle j

    $a_0 = \frac{1}{n}(y_0 + y_1) + \frac{2}{n}S$

$S = 0$

Si $a < k \leq n$ entonces

para $j = 1, \ldots, n-1$

$$S = s + f(x_j) * \cos\left(\frac{k\pi j}{n}\right)$$

fin de bucle j

$\widehat{a}_k = \frac{1}{n}(y_0 + y_1 \cos(k\pi)) + \frac{2}{n}S$

Fin de bucle k

6. $S = 0$
7. Para $k = 1, \ldots, m$

$$S = S + \widehat{a}_k * \cos\left(\frac{k\pi x}{L}\right)$$

Fin de bucle k

8. $Q_m(x) = \frac{1}{2}\widehat{a}_0 + S$
9. Imprimir $x,\ Q_m(x)$
10. Fin

Para los otros casos se elaboran algoritmos muy similares, por lo que se propone como ejercicio

9.9. Ejercicios

1. Para la función f definida en $[0,1]$ que en cada item se propone, hallar el polinomio P que mejor se aproxima en mínimos cuadrados a la función f.

   a) $f(x) = e^x$, y $P(x) = a + bx \quad x \in [0,1]$ **b)** $f(x) = e^{-x}$, y $P(x) = a + bx + cx^2 \quad x \in [0,1]$

   c) $f(x) = \operatorname{sen} x$, y $P(x) = a + bx \quad x \in [0,1]$. **d)** $f(x) = \sqrt{x+1}$, y $P(x) = a + bx + cx^2 \quad x \in [0,1]$
   .

2. Supongamos que se dispone del conjunto de n pares de datos experimentales $S = \{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, ..., n\}$ que en cada item se indica. Encontrar un polinomio P de grado n que se indica y que mejor se ajusta al conjunto de datos.

   a) $S = \{(0,1), (1,-1), (2,1), (3,3)\}$ y $P(x) = a + bx \quad x \in \mathbb{R}$.

   b) $S = \{(0,0), (1,-1), (2,0), (3,3), (4,8)\}$ y $P(x) = a + bx + cx^2 \quad x \in \mathbb{R}$.

   c) $S = \{(0,0), (1,-1), (2,0), (3,3), (4,8)\}$ y $P(x) = a + bx + cx^2 \quad x \in \mathbb{R}$.

   d) $S = \{(0,1), (1,4), (2,9), (3,16), (4,25)\}$ y $P(x) = a + bx + cx^2 \quad x \in \mathbb{R}$.

3. Sea f una función real dependiente de un parámetro c. Escribiremos $t = f(x,c)$. Suponga que $\frac{\partial f}{\partial c}$ es continua.

   Se dispone de un conjunto de datos experimentales

   $$S = \left\{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots, n\right\}$$

   y se asume que cada $y_i = f(x_i, c) + r_i(c)$, donde $r_i(c)$ denota el error en la observación y_i, $i = 1, \ldots, n$.

En el método de mínimos cuadrados se considera el problema siguiente:

$$\min_{c\in\mathbb{R}} \sum_{i=1}^{n} r_i^2(c).$$

Se define $E(c) = \sum\limits_{i=1}^{n} r_i^2(c) = \sum\limits_{i=1}^{n} (y_i - f(x_i, c))^2$.

a) Elaborar un algoritmo para aproximar $\widehat{c} \in \mathbb{R}$ tal que

$$E(\widehat{c}) = \min_{c\in\mathbb{R}} E(c).$$

b) Se considera la siguiente información experimental:

$$S = \{(1, 1{,}35), (1{,}5, 0{,}498), (2, 0{,}183), (2{,}2, 0{,}123)\}$$

Aplique el método de mínimos cuadrados para calcular la constante $\widehat{c} > 0$ tal que $f(t) = 10e^{-\widehat{c}t}$.

c) Se considera el siguiente conjunto de datos

$$S = \{(0{,}26, 5), (0{,}785, 5), (0{,}5, 8{,}7), (1{,}05, 8{,}7)\}$$

Aplique el método de mínimos cuadrados para calcular la constante $\widehat{c}$ tal que $f(t) = 10\sin(ct)$.

4. Sea F una función real dependiente de dos parámetros a y b.

Escribiremos $y = f(x, a, b)$. Se supone que $\dfrac{\partial^2 f}{\partial a^2}, \dfrac{\partial^2 f}{\partial b^2}, \dfrac{\partial^2 f}{\partial a \partial b}$ son continuas.

Se dispone de un conjunto de datos experimentales

$$S = \left\{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots, n\right\}, \quad n \geq 3.$$

Y se asume que cada $y_i = f(x_i, a, b) + r_i(a, b)$, donde $r_i(a, b)$ denota el error en la observación $y_i, i = 1, \ldots, n$.

En el método de mínimos cuadrados se considera el problema

$$\min_{(c)a,b\in\mathbb{R}^2} \sum_{i=1}^{n} r_i^2(a, b). \tag{P}$$

A fin de calcular los parámetros a y b usando la información experimental, definimos

$$E(a, b) = \sum_{i=1}^{n} r_i^2(a, b) = \sum_{i=1}^{n} (y_i - f(x_i, a, b))^2.$$

a) Utilice el método de Newton y elabore un algoritmo para aproximar $\widehat{a}, \widehat{b} \in \mathbb{R}$ tales que

$$E\left(\widehat{a}, \widehat{b}\right) = \min_{(a,b)\in\mathbb{R}} E(a, b).$$

b) Considere la siguiente información experimental

$$S = \{(2, 20), (10, 20{,}2), (50, 21{,}03), (100, 22{,}1), (500, 33)\}$$

Aplique el método de mínimos cuadrados para calcular los parámetros $\widehat{a}, \widehat{b}$ tales que $f(t) = \widehat{a}e^{\widehat{b}t}$ $t \geq 0$.

c) Se dispone del conjunto de datos

$$S = \{(0{,}1, 50), (2, 3{,}85), (4, 1{,}02), (5, 0{,}66)\}$$

Aplique el método de mínimos cuadrados para calcular los parámetros $\widehat{a}, \widehat{b}$ tales que $f(x) = \frac{a}{1+bx^2}$ $x \geq 0$.

9.10. Lecturas complementarias y bibliografía

1. Tom M. Apostol, Análisis Matemático, Segunda Edición, Editorial Reverté, Barcelona, 1982.
2. N. Bakhvalov, Metodos Numéricos, Editorial Paraninfo, Madrid, 1980.
3. Åke Björck, Numerical Methods for Least Squares Problems, Editorial Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1996.
4. E. K. Blum, Numerical Analysis and Computation. Theory and Practice, Editorial Addison-Wesley Publishing Company, Reading, Massachusetts, 1972.
5. John P. Boyd, Chebyshev and Fourier Spectral Methods, Second Edition (Revised), Editorial Dover Publications, Inc.,Mineola, 2001.
6. Richard L. Burden, J. Douglas Faires, Análisis Numérico, Séptima Edición, International Thomson Editores, S. A., México,2002.
7. Steven C. Chapra, Raymond P. Canale, Numerical Methods for Engineers, Third Edition, Editorial McGraw-Hill, Boston, 1998.
8. P. G. Ciarlet, Introduction á L' Analyse Numérique Matricielle et á L' Optimisation, Editorial Masson, París, 1990.
9. S. D. Conte, Carl de Boor, Análisis Numérico, Segunda Edición, Editorial Mc Graw-Hill, México, 1981.
10. B. P. Demidovich, I. A. Maron, E. Cálculo Numérico Fundamental, Editorial Paraninfo, Madrid, 1977.
11. B. P. Demidovich, I. A. Maron, E. S. Schuwalowa, Métodos Numéricos de Análisis, Editorial Paraninfo, Madrid, 1980.
12. J. E. Dennis, Jr., Robert B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Editorial Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1996.
13. Ferruccio Fontanella, Aldo Pasquali, Calcolo Numerico. Metodi e Algoritmi, Volumi I, II Pitagora Editrice Bologna, 1983.
14. John E. Freund, Ronald E. Walpole, Estadística Matemática con Aplicaciones, Cuarta Edición, Editorial Prentice-Hall Hispanoamericana, S. A., México, 1990.
15. Clude Gasquet, Patrick Witomski, Analyse de Fourier et Applications: Filtrage, Calcul Numérique et Ondeletles, Editorial-Dunod, París, 2000.
16. Curtis F. Gerald, Patrick O. Wheatley, Análisis Numérico con Aplicaciones, Sexta Edición, Editorial Pearson Educación de México, México, 2000.
17. Gene H. Golub, Charles F. Van Loan, Matrix Computations, Second Edition, The Johns Hopkins University Press, Baltimore, 1989.
18. Kenneth Hoffman, Ray Kunze, Algebra Lineal, Editorial Prentice-Hall Hispanoamericana, S. A., México, 1987.
19. Franz E. Hohn, Algebra de Matrices, Editorial Trillas, México, 1979.
20. Robert W. Hornbeck, Numerical Methods, Quantum Publishers, Inc., New York, 1975.
21. David Kincaid, Ward Cheney, Análisis Numérico, Editorial Addison-Wesley Iberoamericana, Wilmington, 1994.
22. Erwin Kreyszig, Introducción a la Estadística Matemática, Editorial Limusa, México, 1981.

23. Charles L. Lawson, Richard J. Hanson, Solving Least Squares Problems, Editorial Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1995.

24. L. Lebart, A. Morineau, J.-P. Fénelon, Tratamiento Estadístico de Datos, Editorial Marcombo Boixareu Editores, Barcelona, 1985.

25. Thomas M. Little, F. Jackson Hills, Métodos Estdísticos para la Investigación en la Agricultura, Editorial Trillas, México, 2002.

26. Shoichiro Nakamura, Métodos Numérico Aplicados con Software, Editorial Prentice-Hall Hispanoamericana, S. A., México, 1992.

27. Antonio Nieves, Federico C. Dominguez, Métodos Numéricos Aplicados a la Ingeniería, Tercera Reimpresión, Compañía Editorial Continental, S. A. De C. V., México, 1998.

28. Anthony Ralston, Introducción al Análisis Numérico, Editorial Limusa, México, 1978.

29. Fazlollah Reza, Los Espacios Lineales en la Ingeniería, Editorial Reverté, S. A., Barcelona, 1977.

30. Francis Scheid, Theory and Problems of Numerical Analysis, Schaum's Outline Series, Editorial McGraw-Hill, New York, 1968.

31. M. Sibony, J. Cl. Mardon, Analyse Numérique I, Sustèmes Linéaires et non Linéaires, Editorial Hermann, París, 1984.

32. J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, Editorial Springer-Verlag, 1980.

33. Gilbert Strang, Algebra Lineal y sus Aplicaciones, editorial Fondo Educativo Interamericano, México, 1982.

34. V. Voïévodine, Principes Numériques D' Algèbre Linéaire, Editions Mir, Moscú, 1976.