

Métodos Numéricos

Agustín Mista

5 de Febrero de 2018

Índice general

Sucesiones y Series	3
Sucesiones Numéricas	3
Sucesión Convergente	3
Sucesiones Acotadas	3
Monotonía	3
Operaciones con Sucesiones	3
Series Numéricas	4
Convergencia y Divergencia de Series	4
Condición Necesaria de Convergencia	4
Reindexado de Términos	5
Propiedad de Linearidad	5
Propiedad Telescópica	5
Criterios de Convergencia de Series de Términos Positivos	5
Criterios de Convergencia de Series Alternadas	6
Convergencia Condicional y Absoluta	6
Errores	6
Representación Computacional de Números	6
Enteros Binarios	6
Fracciones Binarias	7
Representación Computacional de Números en Punto Flotante	7
Norma IEEE754 para Números en Punto Flotante	7
Error Absoluto y Relativo	8
Cifras Significativas	8
Fuentes de Errores en Problemas Matemáticos de Ingeniería	8
Polinomio de Taylor	9
Error del Polinomio de Taylor	9
Solución de Ecuaciones No Lineales	9
Algoritmo	9
Criterios de Terminación	9
Orden de Convergencia	10
Solución de Ecuaciones No Lineales	10
Raíces o Ceros	10
Método de la Bisección	10
Método de Newton (Newton-Raphson)	11
Método de la Secante	12
Método de la Falsa Posición	12

Métodos Iterativos de Punto Fijo	12
Punto fijo	12
Existencia de Soluciones de $x = g(x)$	13
Sistemas de Ecuaciones No Lineales	14
Solución de Ecuaciones Lineales	14
Métodos Directos	14
Matriz Inversa	15
Eliminación de Gauss	15
Pivoteo	16
Factorización LU	16
Factorización LU con Matriz de Permutación	16
Matrices Simétricas	17
Matrices No Simétricas Reales	17
Ortogonalización de Gram-Schmidt	17
Factorización QR	18
Problema de Mínimos Cuadrados	18
Normas Vectoriales y Matriciales	19
Estabilidad de la Resolución de Sistemas Lineales	19
Métodos Iterativos	20
Esquema general de los Métodos Iterativos	20
Estabilidad Asintótica de un Sistema Lineal Discreto	21
Método de Sobrerrelajación (SOR)	21
Aproximación de Autovalores	22
Círculos de Gerschgorín	22
Método de la Potencia	23
Interpolación y Ajuste de Curvas	23
Interpolación Polinomial	23
Método de Interpolación de Lagrange	24
Método de Interpolación por Diferencias Divididas de Newton	24
Error en la Interpolación Polinomial	25
Acotación del Error (Caso General)	25
Aproximación con Menor V.A. Máximo	25
Polinomio de Chebychev	26
Aproximación de Mínimos Cuadrados	26
Integración Numérica	26
Integración Numérica Basada en Polinomios Interpolantes	26
Regla del Trapecio	27
Método Compuesto del Trapecio	27
Regla de Simpson	28
Método Compuesto de Simpson	29
Integración Numérica en Dominio Bidimensional	29

Sucesiones y Series

Sucesiones Numéricas

Definición: una sucesión es una función $f : \mathbb{N} \rightarrow \mathbb{R}$. Una sucesión genera una lista infinita de números $f(1), f(2), \dots, f(n), \dots$. También puede notarse $f_1, f_2, \dots, f_n, \dots$ ó $\{f(n)\}$ ó $\{f_n\}$ ó $\{f_n\}_{n=1}^{\infty}$.

Sucesión Convergente

Una sucesión f_n es convergente si existe un número real L tal que para cada $\epsilon > 0$ se puede encontrar un número natural $N(\epsilon)$ tal que $\forall n > N$ se verifique $|f_n - L| < \epsilon$. Se dice entonces que L es el límite de la sucesión f_n , y se escribe $L = \lim_{n \rightarrow \infty} f_n$ ó $f_n \rightarrow L$. Decimos entonces que f_n converge a L . Una sucesión no convergente se llama divergente.

Teorema: (Unicidad del límite) una sucesión convergente tiene un único límite.

Sucesiones Acotadas

Definición: una sucesión f_n se dice que está acotada superiormente si existe un número $c \in \mathbb{R}$ tal que $\forall n \in \mathbb{N}, f_n \leq c$. Se dice que está acotada inferiormente si existe un número $k \in \mathbb{R}$ tal que $\forall n \in \mathbb{N}, f_n \geq k$. Se dice que está acotada si lo está superior e inferiormente, es decir que existe $M > 0$ tal que $\forall n \in \mathbb{N}, |f_n| \leq M$.

Monotonía

Definición: una sucesión f_n es:

- Monótona creciente si $\forall n \in \mathbb{N}, f_n \leq f_{n+1}$.
- Monótona decreciente si $\forall n \in \mathbb{N}, f_{n+1} \leq f_n$.
- Estrictamente creciente si $\forall n \in \mathbb{N}, f_n < f_{n+1}$.
- Estrictamente decreciente si $\forall n \in \mathbb{N}, f_{n+1} < f_n$.

Teorema:

1. Toda sucesión monótona creciente y acotada superiormente converge.
2. Toda sucesión monótona decreciente y acotada inferiormente converge.

Operaciones con Sucesiones

Teorema: sean a_n y b_n sucesiones convergentes con límites a y b respectivamente. Luego se cumplen las siguientes reglas:

1. $\lim_{n \rightarrow \infty} (a_n + b_n) = a + b$
2. $\lim_{n \rightarrow \infty} (a_n - b_n) = a - b$
3. $\lim_{n \rightarrow \infty} (a_n \cdot b_n) = a \cdot b$
4. $\lim_{n \rightarrow \infty} (c \cdot a_n) = c \cdot a$ ($c \in \mathbb{R}$)
5. $\lim_{n \rightarrow \infty} (a_n / b_n) = a / b$ (si $b \neq 0$)

Teorema: (del Sándwich) sean a_n , b_n y c_n sucesiones. Si $a_n \leq b_n \leq c_n$ para todo $n > N \in \mathbb{N}$ y $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = L$ entonces $\lim_{n \rightarrow \infty} b_n = L$.

Teorema: suponga que $f(x)$ es una función definida para todo $x \geq N$ y que f_n es una sucesión de números reales tal que $f_n = f(n)$ para todo $n \geq N$, entonces $\lim_{x \rightarrow \infty} f(x) = L \Rightarrow \lim_{n \rightarrow \infty} f_n = L$. Por lo tanto podemos usar la regla de L'Hopital para calcular límites de sucesiones.

Series Numéricas

Definición: dada una sucesión a_n formamos otra S_n para la cual $S_n = a_1 + a_2 + \dots + a_n = \sum_{n=1}^{\infty} a_n$. Luego la sucesión S_n es llamada serie infinita o serie y se indica como $\sum_{n=1}^{\infty} a_n$.

Una serie $\sum_{n=1}^{\infty} a_n$ se dice:

- de términos positivos si $\forall n, a_n > 0$.
- alternada si $a_n = (-1)^n c_n$ para alguna sucesión c_n tal que $\forall n, c_n > 0$.

Ejemplo:

- Serie armónica: $\sum_{n=1}^{\infty} 1/n$
- Serie geométrica (de razón r): $\sum_{n=1}^{\infty} r^n$
 - Si $r > 0$ es una serie de términos positivos.
 - Si $r < 0$ es una serie alternada.
 - Si $r \neq 1$ entonces $\sum_{n=1}^{\infty} r^n = \frac{1-r^{n+1}}{1-r}$

Convergencia y Divergencia de Series

Definición: se dice que la serie $\sum_{n=1}^{\infty} a_n$ es convergente cuando la sucesión S_n tiene límite finito, y notamos $\lim_{n \rightarrow \infty} (\sum_{k=1}^{\infty} a_k) = \lim_{n \rightarrow \infty} S_n = S$.

Ejemplo: la serie armónica $\sum_{n=1}^{\infty} 1/n$ no converge.

Sabemos que $\frac{1}{2} \leq S_{2n} - S_n$, luego suponemos que existe s tal que $\lim_{n \rightarrow \infty} S_n = s$:

$$\frac{1}{2} \leq \lim_{n \rightarrow \infty} S_{2n} - \lim_{n \rightarrow \infty} S_n = s - s = 0 \quad (ABS!)$$

Condición Necesaria de Convergencia

Teorema: si la serie $\sum_{n=1}^{\infty} a_n$ converge, entonces $\lim_{n \rightarrow \infty} a_n = 0$. En cambio, si $\lim_{n \rightarrow \infty} a_n \neq 0$ entonces $\sum_{n=1}^{\infty} a_n$ es divergente.

Demostración: suponemos que la serie $\sum_{n=1}^{\infty} a_n$ converge a s , luego:

$$a_k = S_k - S_{k-1} \Rightarrow \lim_{k \rightarrow \infty} S_k - \lim_{k \rightarrow \infty} S_{k-1} = s - s = 0$$

Ejemplo: Serie geométrica $\sum_{n=0}^{\infty} r^n = \sum_{n=1}^{\infty} r^{n-1}$.

- Si $|r| \geq 1$ entonces $\lim_{n \rightarrow \infty} r^n \neq 0 \Rightarrow$ la serie diverge.
- Si $|r| < 1$ entonces $\lim_{n \rightarrow \infty} r^n = \lim_{n \rightarrow \infty} \frac{1-r^{n+1}}{1-r} = \frac{1}{1-r}$.

Reindexado de Términos

Si a_n es una serie, entonces.

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1+h}^{\infty} a_{n-h} = \sum_{n=1-h}^{\infty} a_{n+h}$$

Si $\sum_{n=1}^{\infty} a_n$ converge, entonces $\sum_{n=k}^{\infty} a_n$ converge para cualquier $k > 1$.

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + \cdots + a_{k-1} + \sum_{n=k}^{\infty} a_n$$

En particular para la serie geométrica $\sum_{n=1}^{\infty} r^n = \frac{r}{1-r}$ si $|r| < 1$.

Propiedad de Linealidad

Teorema: sean $\sum_{n=1}^{\infty} a_n$ y $\sum_{n=1}^{\infty} b_n$ series convergentes con sumas a y b respectivamente. Si α y β son constantes entonces $\sum_{n=1}^{\infty} (\alpha a_n + \beta b_n)$ es convergente con suma $\alpha a + \beta b$.

Corolario: si $\sum_{n=1}^{\infty} a_n$ converge y $\sum_{n=1}^{\infty} b_n$ diverge entonces $\sum_{n=1}^{\infty} (a_n + b_n)$ diverge.

Demostración: si $\sum_{n=1}^{\infty} (a_n + b_n)$ fuera convergente, entonces también lo sería $\sum_{n=1}^{\infty} b_n$ ya que $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} (a_n + b_n) + \sum_{n=1}^{\infty} (-a_n)$.

Propiedad Telescópica

Definición: una serie $\sum_{n=1}^{\infty} a_n$ es telescópica cuando se puede escribir como $\sum_{n=1}^{\infty} (b_n - b_{n+1})$ para alguna sucesión b_n tal que $a_n = b_n - b_{n+1}$. Luego tenemos que $\sum_{k=1}^n (b_k - b_{k+1}) = b_1 - b_{n+1}$.

Teorema: sean a_n y b_n sucesiones tales que $a_n = b_n - b_{n+1}$. Luego $\sum_{n=1}^{\infty} a_n$ converge \iff la sucesión b_n converge. En cuyo caso $\sum_{n=1}^{\infty} a_n = b_1 - \lim_{n \rightarrow \infty} b_n$.

Criterios de Convergencia de Series de Términos Positivos

- **Criterio de Acotación:** si $\forall n \geq 1, a_n \geq 0$, entonces $\sum_{n=1}^{\infty} a_n$ converge \implies la sucesión de sus sumas parciales está acotada superiormente.
- **Criterio de Comparación:** si $a_n \geq 0, b_n \geq 0$ y existe $c > 0$ tal que $a_n \leq cb_n$ si $n \geq N$ para algún N , entonces:
 - Si $\sum_{n=1}^{\infty} b_n$ converge $\implies \sum_{n=1}^{\infty} a_n$ converge.
 - Si $\sum_{n=1}^{\infty} a_n$ diverge $\implies \sum_{n=1}^{\infty} b_n$ diverge.
- **Criterio del Límite:** sean a_n y b_n sucesiones tales que $a_n \geq 0$ y $b_n \geq 0$ y sea $\lambda = \lim_{n \rightarrow \infty} \frac{a_n}{b_n}$. Si λ es finito y $\lambda \neq 0$, entonces $\sum_{n=1}^{\infty} b_n$ converge $\iff \sum_{n=1}^{\infty} a_n$ converge.
- **Criterio de la Raíz:** sea a_n una sucesión tal que $a_n > 0$ y sea $\alpha = \lim_{n \rightarrow \infty} \sqrt[n]{a_n}$, entonces:
 - $\alpha < 1 \implies \sum_{n=1}^{\infty} a_n$ converge.
 - $\alpha > 1 \implies \sum_{n=1}^{\infty} a_n$ diverge.
 - $\alpha = 1 \implies$ el criterio no decide.
- **Criterio del Cociente:** sea a_n una sucesión tal que $a_n > 0$ y sea $\alpha = \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n}$, entonces:

- $\alpha < 1 \implies \sum_{n=1}^{\infty} a_n$ converge.
- $\alpha > 1 \implies \sum_{n=1}^{\infty} a_n$ diverge.
- $\alpha = 1 \implies$ el criterio no decide.

- **Criterio de la integral:** sea f una función **positiva** y **estrictamente decreciente** definida en $[1, \infty)$ tal que $\forall n \in \mathbb{N}, f(n) = a_n$. La serie $\sum_{n=1}^{\infty} a_n$ converge $\iff \int_1^{\infty} f(x)dx$ converge.

Criterios de Convergencia de Series Alternadas

- **Criterio de Leibniz:** si a_n es una sucesión **monótona decreciente** con límite 0, entonces la serie alternada $\sum_{n=1}^{\infty} (-1)^{n-1} a_n$ converge.

Convergencia Condicional y Absoluta

Teorema: si $\sum_{n=1}^{\infty} |a_n|$ converge, entonces $\sum_{n=1}^{\infty} a_n$ converge y además tenemos que $|\sum_{n=1}^{\infty} a_n| \leq \sum_{n=1}^{\infty} |a_n|$.

Demostración: definimos la sucesión de términos positivos $b_n = a_n + |a_n|$. Luego resulta $b_n = 0$ ó $b_n = 2|a_n|$ y siempre vale que $0 \leq b_n \leq 2|a_n|$. Como además sabemos que $\sum_{n=1}^{\infty} |a_n|$ converge y $2\sum_{n=1}^{\infty} |a_n|$ domina a $\sum_{n=1}^{\infty} b_n$ luego $\sum_{n=1}^{\infty} b_n$ converge. Como $\sum_{n=1}^{\infty} b_n$ converge y $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} b_n - \sum_{n=1}^{\infty} |a_n|$ entonces $\sum_{n=1}^{\infty} a_n$ converge por el teorema de linealidad.

Definición: una serie $\sum_{n=1}^{\infty} a_n$ es absolutamente convergente si $\sum_{n=1}^{\infty} |a_n|$ es convergente.

Definición: una serie $\sum_{n=1}^{\infty} a_n$ es condicionalmente convergente si $\sum_{n=1}^{\infty} a_n$ es convergente pero $\sum_{n=1}^{\infty} |a_n|$ es divergente.

Errores

Representación Computacional de Números

Enteros Binarios

$$x = (a_n a_{n-1} \dots a_1 a_0)_2 \quad \text{con} \quad a_i \in \{0, 1\}$$

Binario \rightarrow Decimal:

$$x = a_n 2^n + a_{n-1} 2^{n-1} + \dots + a_1 2^1 + a_0 = \sum_{n=0}^n a_n 2^n$$

Decimal \rightarrow Binario:

x número decimal \rightarrow dividir x por 2 \rightarrow llamar x al cociente y a_0 al resto \rightarrow repetir

Fracciones Binarias

$$x = (.a_1a_2...a_m...)_2 \quad \text{con} \quad a_i \in \{0,1\}$$

Binario \rightarrow Decimal:

$$x = a_12^{-1} + a_22^{-2} + \dots + a_m2^{-m} + \dots = \sum_{n=0}^{\infty} a_n2^{-n}$$

Decimal \rightarrow Binario:

x número decimal \rightarrow multiplicar x por 2 \rightarrow llamar x a la parte decimal y a_1 a la parte entera \rightarrow repetir

Representación Computacional de Números en Punto Flotante

Sea $x \in \mathbb{R}$, x se representa en punto flotante ($fl(x)$) como:

$$fl(x) = \delta(.a_1a_2...a_n)_\beta \cdot \beta^E$$

Donde:

- β : base de representación
- δ : signo (0 = positivo, 1 = negativo)
- E : exponente ($E \in \mathbb{Z}$)

Norma IEEE754 para Números en Punto Flotante

Sea $x \in \mathbb{R}$, x se representa en punto flotante IEEE754 ($fl(x)$) como:

$$fl(x) = \delta(1.a_1a_2...a_n)_2 \cdot 2^E$$

Precisión Simple (32 bits) \rightarrow

1	8	23
signo	exponente	mantisa

Precisión Doble (64 bits) \rightarrow

1	11	52
signo	exponente	mantisa

Máximo valor del exponente con precisión simple: $(11111111)_2 = 2^8 - 1 = 255$.

En IEEE754 se representan del 1 al 254, 0 y 255 está reservados para casos especiales. Utilizando un sesgo de 127 se pueden representar exponentes en el rango $-126 \leq E \leq 127$.

¿Cuál es el mayor entero M tal que todo entero x tal que $0 \leq x \leq M$ se puede representar en forma exacta en punto flotante?

$$(1.11...1)_2 \cdot 2^{23} = 2^{23} + 2^{22} + \dots + 2^1 + 2^0 = 2^{24} + 1$$

¿Cuál es el menor número y representable que es mayor que 1?

$$(1.00...01)_2 \cdot 2^0 = 1 + 2^{-23}$$

Luego el epsilon de la máquina es $\epsilon = y - 1 = 2^{-23}$.

Corte o Truncamiento: si x tiene una mantisa \underline{x} que no cabe en el espacio disponible de n bits, el truncamiento consiste en cortar los dígitos a_{n+1}, a_{n+2}, \dots . Luego el error es $x - fl(x)$ y es siempre positivo, lo cual puede generar propagación de errores.

Redondeo en Decimal:

$$fl(x) = \begin{cases} \delta(.a_1 a_2 \dots a_n)_{10} \cdot 10^E, & a_{n+1} < 5 \quad (\text{truncar}) \\ \delta[(.a_1 a_2 \dots a_n)_{10} + (.00...01)_{10}] \cdot 10^E, & a_{n+1} \geq 5 \quad (\text{redondear}) \end{cases}$$

Redondeo en Binario:

$$fl(x) = \begin{cases} \delta(1.a_1 a_2 \dots a_n)_2 \cdot 2^E, & a_{n+1} = 0 \\ \delta[(1.a_1 a_2 \dots a_n)_2 + (.00...01)_2] \cdot 2^E, & a_{n+1} = 1 \end{cases}$$

Usando redondeo el mayor error posible es la mitad que usando truncamiento. Tiene en promedio la mitad de las veces un signo y la mitad el otro, por lo que reduce la propagación de errores.

Error Absoluto y Relativo

Error Absoluto: $|\text{valor verdadero} - \text{valor aproximado}| = |x_v - x_a|$

Error Relativo: $\frac{\text{error absoluto}}{|\text{valor verdadero}|} = \frac{|x_v - x_a|}{|x_v|}$

Cifras Significativas

Definición: se dice que x_a tiene m cifras significativas con respecto a x_v si $|error(x_a)| \leq 5$ unidades en el dígito $m + 1$ de x_v , contando de izquierda a derecha desde el primer dígito distinto de 0.

Ejemplo: $x_a = 0.02138$, $x_v = 0.02144$, luego $|x_v - x_a| = 0.0006$ y x_a tiene 2 cifras significativas.

Se puede demostrar que si el error relativo de x_a respecto de x_v es menor a $5 \times 10^{-m-1}$ luego x_a tiene m cifras significativas respecto a x_v .

Fuentes de Errores en Problemas Matemáticos de Ingeniería

- Errores de Modelado Matemático
- Equivocaciones
- Errores Observacionales
- Errores de Redondeo/Truncamiento
- Errores de Aproximación Matemática

Polinomio de Taylor

Definición: sea $f(x)$ una función derivable alrededor de $x = a$, con tantas derivadas como sea necesario. Buscamos un polinomio $P(x)$ tal que:

$$\begin{aligned}P(a) &= f(a) \\P'(a) &= f'(a) \\&\vdots \\P^{(n)}(a) &= f^{(n)}(a)\end{aligned}$$

La fórmula general para dicho polinomio es:

$$P_n(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \dots + \frac{1}{n!}f^{(n)}(a)(x-a)^n = \sum_{i=0}^n \frac{f^{(i)}(a)}{i!}(x-a)^i$$

Luego $f(x) \approx P_n(x)$ alrededor de a .

Error del Polinomio de Taylor

Teorema: suponga que $f(x)$ tiene $n+1$ derivadas continuas en un intervalo $\alpha \leq x \leq \beta$ y que el punto a pertenece a dicho intervalo. El error del polinomio de Taylor está dado por

$$R_n(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(cx)}{(n+1)!}(x-a)^{n+1}$$

donde cx pertenece al intervalo entre x y a .

Solución de Ecuaciones No Lineales

Algoritmo

Definición: dado un punto inicial $x_0 \in \mathbb{R}^n$, un algoritmo genera la secuencia x_1, x_2, \dots donde $x_{k+1} \in A(x_k)$ para cada k . La transformación de x_k a x_{k+1} constituye una iteración del algoritmo.

Criterios de Terminación

- (1) $\|x_{k+n} - x_k\| < \epsilon$
- (2) $\frac{\|x_{k+1} - x_k\|}{\|x_k\|} < \epsilon$
- (3) $|f(x_{k+n}) - f(x_k)| < \epsilon$

Orden de Convergencia

Definición: el orden de convergencia de una sucesión $x_k \rightarrow \underline{x}$ es el mayor número $\rho > 0$ tal que

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - \underline{x}\|}{\|x_k - \underline{x}\|^\rho} = \beta < \infty$$

- Si $\rho = 1$ y $\beta \in (0, 1)$, la convergencia es lineal y β es la velocidad de convergencia.
- Si $\rho = 2$ y $\beta < \infty$, la convergencia es cuadrática.
- Si $\rho > 1$ ó $\rho = 1$ y $\beta = 0$, la convergencia es superlineal.

Definición: la convergencia es superlineal si

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - \underline{x}\|}{\|x_k - \underline{x}\|} = 0$$

Además, la convergencia cuadrática es superlineal.

Por otro lado, suponemos que la convergencia es de orden ρ , y sea $\alpha > 0$ y β tal que $0 < \beta < \infty$:

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - \underline{x}\|}{\|x_k - \underline{x}\|^{\rho+\alpha}} = \frac{\beta}{\lim_{k \rightarrow \infty} \|x_k - \underline{x}\|^\alpha} \rightarrow \infty$$

Por lo tanto, si se tiene convergencia de orden ρ , no se tiene convergencia de orden $\rho + \alpha$ con $\alpha > 0$.

Solución de Ecuaciones No Lineales

Raíces o Ceros

Definición: sea $f : \mathbb{R} \rightarrow \mathbb{R}$ una función no lineal. Se llama raíz o cero de la ecuación $f(x) = 0$ a todo valor α tal que $f(\alpha) = 0$.

Teorema: (de Bolzano) sea f una función continua en $[a, b] \subset \mathbb{R}$ tal que $f(a)f(b) < 0$, luego existe un $c \in [a, b]$ tal que $f(c) = 0$.

Método de la Bisección

Suponemos que $f(x)$ es continua en $[a, b]$ y que $f(a)f(b) < 0$, luego f tiene al menos una raíz en el intervalo. Dada una tolerancia del error $\epsilon > 0$:

1. Defina $c = (a + b)/2$.
2. Si $b - c \leq \epsilon$, aceptar c como raíz y detenerse.
3. Si $b - c > \epsilon$, comparar el signo de $f(c)$ con el de $f(a)$ y $f(b)$.
 - Si $f(b)f(c) \leq 0$, reemplazar a con c .
 - En caso contrario, reemplazar b con c .
4. Regresar al paso 1.

Acotación del Error Tenemos que $b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k)$, luego por inducción tenemos que $b_k - a_k = (\frac{1}{2})^{k-1}(b_1 - a_1)$. Si α es una solución del sistema, luego $|\alpha - c_k| \leq (\frac{1}{2})^{k-1}(b_1 - a_1)$ y c_k converge a α cuando $k \rightarrow \infty$.

Si queremos obtener un error $|\alpha - c_k| \leq \epsilon$, esto se cumple cuando:

$$k \geq \frac{\ln(\frac{b-a}{\epsilon})}{\ln(2)}$$

Ventajas

1. Siempre converge
2. Acotación de error garantizado
3. Velocidad de convergencia garantizada.

Desventaja

1. La convergencia es lenta en comparación con otros métodos.

Método de Newton (Newton-Raphson)

Sea α una raíz de $f(x) = 0$. Supongamos que $f \in \mathbb{C}^2$ en $[a, b]$ y sea $x_0 \in [a, b]$ una estimación “cercana” de α . Consideramos el polinomio de Taylor $f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2}f''(cx)$ con cx entre x y x_0 . Obtenemos una nueva estimación de α haciendo $P_1(x) = 0 = f(x_0) + (x_1 - x_0)f'(x)$ de donde $x_1 = x_0 - (f(x_0)/f'(x_0))$. Repitiendo el proceso obtenemos que $x_{n+1} = x_n - (f(x_n)/f'(x_n))$.

Análisis del Error

$$\begin{aligned} 0 = f(\alpha) &= f(x_n) + (\alpha - x_n)f'(x_n) + (\alpha - x_n)^2 \frac{f''(cx)}{2} \\ &= \frac{f(x_n)}{f'(x_n)} + (\alpha - x_n) + (\alpha - x_n)^2 \frac{f''(cx)}{2f'(x_n)} \\ &= (x_n - x_{n+1}) + \alpha - x_n + (\alpha - x_n)^2 \frac{f''(cx)}{2f'(x_n)} \\ \Rightarrow \alpha - x_{n+1} &= (\alpha - x_n)^2 \frac{f''(cx)}{2f'(x_n)} \quad (\text{Error}) \\ \Rightarrow \frac{|\alpha - x_{n+1}|}{(\alpha - x_n)^2} &= \left| \frac{f''(cx)}{2f'(x_n)} \right| \quad (\text{Orden de convergencia}) \end{aligned}$$

Es decir, suponiendo que el método converge, éste lo hace de manera cuadrática. Sin embargo, la convergencia no está garantizada a partir de cualquier valor inicial x_0 .

Ventajas

1. Converge rápidamente en la mayoría de los casos.
2. Formulación sencilla.

Desventajas

1. Puede no converger.
2. Puede ocurrir que $f'(\alpha) = 0$.
3. Necesitamos conocer tanto $f(x)$ como $f'(x)$.

Método de la Secante

$$x_{n+1} = x_n - f(x_n) \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

Si x_n converge a α entonces $\rho \approx 1.62$.

Ventajas

1. Converge más rápido que la convergencia lineal
2. No requiere $f'(x)$.
3. Requiere una única evaluación de $f(x)$ por iteración.

Desventajas

1. Puede no converger.
2. Puede tener dificultades si $f'(\alpha) = 0$.
3. El método de Newton se puede generalizar más fácilmente a sistemas de ecuaciones.

Método de la Falsa Posición

Elegimos las aproximaciones iniciales a_1 y b_1 con $f(a_1)f(b_1) < 0$, luego obtenemos c_1 aplicando el método de la secante sobre a_1 y b_1 .

- Si $f(a_1)f(c_1) < 0$, luego $a_2 = a_1$ y $b_2 = c_1$
- Si $f(b_1)f(c_1) < 0$, luego $a_2 = c_1$ y $b_2 = b_1$
- Si $f(c_1) = 0$, entonces $\alpha = c_1$

Luego c_2 aplicando el método de la secante a a_2 y b_2 , y repetimos el proceso.

Ventajas

1. La convergencia está garantizada.

Desventajas

1. Más lento que el método de la secante.

Métodos Iterativos de Punto Fijo

Fórmula General $x_{n+1} = g(x_n)$ donde $g(x)$ es una función apropiada.

Punto fijo

Definición: dada una función $g : \mathbb{R} \rightarrow \mathbb{R}$ continua diremos que α es un punto fijo de g si $g(\alpha) = \alpha$.

- El método puede converger o diverger dependiendo de x_0 .
- El método puede converger a una raíz u otra dependiendo de la elección de $g(x)$.
- La convergencia puede ser más rápida o más lenta dependiendo de $g(x)$.

Ejemplo: el método de Newton es un método iterativo de punto fijo con $g(x) = x - (f(x)/f'(x))$.

Existencia de Soluciones de $x = g(x)$

Lema: sea $g(x)$ una función continua en $[a, b]$ y suponga que g satisface que $a \leq x \leq b \implies a \leq g(x) \leq b$. Luego $x = g(x)$ tiene al menos una solución en $[a, b]$.

Demostración: consideremos la función continua $f(x) = x - g(x)$ con $a \leq x \leq b$. Evaluando los puntos extremos, $f(a) \leq 0$ y $f(b) > 0$. Luego por teorema de Bolzano existe $\alpha \in [a, b]$ tal que $f(\alpha) = 0$ y por ende $\alpha = g(\alpha)$.

Teorema: sea $g : \mathbb{R} \rightarrow \mathbb{R}$ tal que $g \in \mathbb{C}^1$ en $[a, b]$. Suponga que g satisface $a \leq x \leq b \implies a \leq g(x) \leq b$. Si $\lambda := \sup|g'(x)| < 1$ con $x \in [a, b]$. Luego se cumplen:

1. Existe una solución única α de la ecuación $x = g(x)$ en $[a, b]$.
2. Para cualquier valor inicial $x_0 \in [a, b]$, la iteración $x_{n+1} = g(x_n)$ converge a α .
3. $|\alpha - x_n| \leq \lambda^n(x_0 - x_1)/(1 - \lambda)$
4. $\lim_{n \rightarrow \infty} (\alpha - x_{n+1})/(\alpha - x_n) = g'(\alpha)$. Por lo tanto para x_n cercano a α tenemos que $\alpha - x_{n+1} \approx g'(\alpha)(\alpha - x_n)$,

Demostración: las hipótesis sobre g permiten aplicar el lema anterior para afirmar que existe al menos una solución de $x = g(x)$ en $[a, b]$. Luego por el teorema del valor medio tenemos que para $w, z \in [a, b]$ se cumple $g(w) - g(z) = g'(c)(w - z)$ para algún c entre w y z . Luego $|g(w) - g(z)| = |g'(c)||w - z| \leq \lambda|w - z|$.

-
1. Por contradicción. Supongo que existen dos soluciones α y β . Luego $\alpha = g(\alpha)$ y $\beta = g(\beta)$. Restando miembro a miembro tengo que $\alpha - \beta = g(\alpha) - g(\beta)$ lo cual implica que $|\alpha - \beta| \leq \lambda|\alpha - \beta|$. Luego $(1 - \lambda)|\alpha - \beta| \leq 0$, y como $0 < \lambda < 1$ tenemos que $\alpha = \beta$ y $x = g(x)$ tiene única solución en $[a, b]$.
 2. La propiedad $a \leq x \leq b \implies a \leq g(x) \leq b$ implica que dado $x_0 \in [a, b]$, las iteraciones $x_k \in [a, b]$. Luego para demostrar que las iteraciones convergen, restar $x_{n+1} = g(x_n)$ de $\alpha = g(\alpha)$, obteniendo $\alpha - x_{n+1} = g(\alpha) - g(x_{n+1}) = g'(cn)(\alpha - x_n)$ para algún cn entre α y x_n . Luego $|\alpha - x_{n+1}| \leq \lambda|\alpha - x_n|$ **(A)**. Por inducción obtenemos que $|\alpha - x_n| \leq \lambda^n|\alpha - x_0|$ **(B)**. Como $\alpha < 1$, $\alpha^n \rightarrow 0$ cuando $n \rightarrow \infty$, y tenemos que $x_n \rightarrow \alpha$ cuando $n \rightarrow \infty$.
 3. Por desigualdad triangular tenemos que $|\alpha - x_0| \leq |\alpha - x_1| + |x_1 - x_0|$, luego aplicamos **(A)** con $n = 0$ obtenemos $|\alpha - x_0| \leq \lambda|\alpha - x_0| + |x_1 - x_0|$ de donde despejamos $|\alpha - x_0| \leq |x_1 - x_0|/(1 - \lambda)$. Multiplicando ambos lados por λ^n obtenemos $|\alpha - x_0| \leq \lambda^n(x_0 - x_1)/(1 - \lambda)$ y aplicando **(B)** obtenemos el resultado buscado.
 4. Vimos que $\alpha - x_{n+1} = g'(cn)(\alpha - x_n)$ para algún cn entre α y x_n . Luego $\lim_{n \rightarrow \infty} (\alpha - x_{n+1})/(\alpha - x_n) = \lim_{n \rightarrow \infty} g'(cn) = g'(\alpha)$ ya que $x_n \rightarrow \alpha$ cuando $n \rightarrow \infty$.

Corolario: suponga que $x = g(x)$ tiene una solución α y suponga que g y g' son continuas en un intervalo alrededor de α . Luego:

1. Si $|g'(\alpha)| < 1$, entonces la iteración $x_{n+1} = g(x_n)$ converge a α para x_0 suficientemente cercano a α .
2. Si $|g'(\alpha)| > 1$, entonces la iteración $x_{n+1} = g(x_n)$ no converge a α .
3. Si $|g'(\alpha)| = 1$, no se pueden sacar conclusiones.

Demostración:

1. Vimos que $\alpha - x_{n+1} = g'(cn)(\alpha - x_n)$ para algún cn entre α y x_n . Luego $|\alpha - x_{n+1}| = |g'(cn)||\alpha - x_n|$. Siendo $g'(x)$ continua y $|g'(\alpha)| < 1$, existe $\epsilon > 0$ tal que $|g'(x)| < 1$ para todo $x \in [\alpha - \epsilon, \alpha + \epsilon]$.

Luego cn también pertenece a $[\alpha - \epsilon, \alpha + \epsilon]$ y $|g'(cn)| < 1$. Por lo tanto x_{n+1} está más próximo a α que x_n , por lo que la iteración converge a α .

2. Vimos que $\alpha - x_{n+1} = g'(cn)(\alpha - x_n)$ para algún cn entre α y x_n . Luego $|\alpha - x_{n+1}| = |g'(cn)| |\alpha - x_n|$. Siendo $g'(x)$ continua y $|g'(\alpha)| > 1$, existe $\epsilon > 0$ tal que $|g'(x)| > 1$ para todo $x \in [\alpha - \epsilon, \alpha + \epsilon]$. Luego cn también pertenece a $[\alpha - \epsilon, \alpha + \epsilon]$ y $|g'(cn)| > 1$. Por lo tanto x_{n+1} está más alejado a α que x_n , por lo que la iteración no converge a α .

Sistemas de Ecuaciones No Lineales

$$S = \begin{cases} f_1(x_1, x_2) = 0 \\ f_2(x_1, x_2) = 0 \end{cases}$$

En forma vectorial, $f(\underline{x}) = \underline{0}$ con $f = [f_1 \ f_2]^T$, $\underline{x} = [x_1 \ x_2]^T$ y $\underline{0} = [0 \ 0]^T$.

Punto Fijo: reescribimos S como

$$S = \begin{cases} x_1^{(n+1)} = g_1(x_1^{(n)}, x_2^{(n)}) \\ x_2^{(n+1)} = g_2(x_1^{(n)}, x_2^{(n)}) \end{cases}$$

Usando el método de Newton tenemos:

$$\begin{aligned} 0 &= f_1(x_1^{(0)}, x_2^{(0)}) + (x_1^{(1)} - x_1^{(0)}) \frac{\delta f_1}{\delta x_1}(x_1^{(0)}, x_2^{(0)}) + (x_2^{(1)} - x_2^{(0)}) \frac{\delta f_1}{\delta x_2}(x_1^{(0)}, x_2^{(0)}) \\ 0 &= f_2(x_1^{(0)}, x_2^{(0)}) + (x_1^{(1)} - x_1^{(0)}) \frac{\delta f_2}{\delta x_1}(x_1^{(0)}, x_2^{(0)}) + (x_2^{(1)} - x_2^{(0)}) \frac{\delta f_2}{\delta x_2}(x_1^{(0)}, x_2^{(0)}) \end{aligned}$$

Denotamos al Jacobiano de f como

$$J = \begin{bmatrix} \frac{\delta f_1}{\delta x_1} & \frac{\delta f_1}{\delta x_2} \\ \frac{\delta f_2}{\delta x_1} & \frac{\delta f_2}{\delta x_2} \end{bmatrix}$$

Luego, en forma matricial, $\underline{x}_0 = [x_1^{(0)} \ x_2^{(0)}]^T$ y $\underline{0} = f(\underline{x}_0) + J(\underline{x}_0)(\underline{x}_1 - \underline{x}_0)$, de donde obtenemos que $\underline{x}_1 = \underline{x}_0 - [J(\underline{x}_0)]^{-1} f(\underline{x}_0)$ si $J(\underline{x}_0)$ es no singular. Finalmente, la iteración general del método resulta:

$$\underline{x}_{n+1} = \underline{x}_n - [J(\underline{x}_n)]^{-1} f(\underline{x}_n)$$

Solución de Ecuaciones Lineales

Métodos Directos

Consideramos sistemas de n ecuaciones con n incógnitas:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

En forma matricial es $Ax = b$ con $A(ij) = a_{ij}$.

Definición: una matriz se dice plena si la mayoría de sus elementos son no nulos. En cambio, se dice rala si la mayoría de sus elementos son nulos.

Definición: decimos que una matriz es p-banda si existe $p \in \mathbb{Z}$ tal que $|i - j| \geq p \implies a_{ij} = 0$. Si $p = 1$, entonces la matriz es diagonal, si $p = 2$ la matriz se dice tri-diagonal.

Definición: dado un sistema $Ax = b$, si $b = 0$, decimos que el sistema es homogéneo. En caso contrario el sistema es no homogéneo.

Teorema: los siguientes enunciados son equivalentes:

1. Para cada b , existe una única solución x .
2. Para cada b , existe una solución x .
3. El sistema homogéneo $Ax = 0$ tiene única solución $x = 0$.
4. $\det(A) \neq 0$
5. Existe A^{-1} .

Regla de Cramer considere el sistema $Ax = b$ con $\det(A) \neq 0$. Luego $x_i = \det(A_i)/\det(A)$ donde $A_i = [a_1 | \dots | a_{i-1} | b | a_{i+1} | \dots | a_n]$.

Matriz Inversa

Definición: sea $A \in \mathbb{R}^{n \times n}$. Decimos que A^{-1} es la inversa de A si $AA^{-1} = A^{-1}A = I$. Si A^{-1} existe, es única. Luego, un sistema $Ax = b$ puede resolverse haciendo $x = A^{-1}b$, aunque esto es muy ineficiente.

Eliminación de Gauss

Consiste en 2 pasos: eliminación de incógnitas y sustitución regresiva.

$$[A^{(1)}|b^{(1)}] = \left[\begin{array}{ccc|c} a_{11}^{(1)} & \dots & a_{1n}^{(1)} & b_1^{(1)} \\ \vdots & & \vdots & \vdots \\ a_{n1}^{(1)} & \dots & a_{nn}^{(1)} & b_n^{(1)} \end{array} \right] \xrightarrow{\text{reduce en (n-1) pasos a}} [A^{(n)}|b^{(n)}] = \left[\begin{array}{ccc|c} a_{11}^{(1)} & \dots & a_{1n}^{(1)} & b_1^{(1)} \\ & \ddots & \vdots & \vdots \\ 0 & & a_{nn}^{(n)} & b_n^{(n)} \end{array} \right]$$

Denotamos $U = A^{(n)}$ y $g = b^{(n)}$, luego $Ux = g$.

Paso k-ésimo: suponga que para $i = 1, \dots, k-1$ los x_i ya han sido eliminados de las ecuaciones $i+1, \dots, n$. Eliminamos x_k de las ecuaciones $k+1, \dots, n$.

$$E_i^{(k+1)} = E_i^{(k)} - m_{ik} E_k^{(k)} \quad (\text{con } i = k+1, \dots, n)$$

Donde $m_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}$. Se llama a a_{kk} el elemento pivote.

Una vez obtenida la matriz triangular superior $A^{(n)} = U$ resolvemos por sustitución regresiva.

$$x_n = \frac{b_n^{(n)}}{a_{nn}^{(n)}} \\ x_i = \frac{1}{a_{ii}^{(i)}} \left(b_i^{(i)} - \sum_{j=i+1}^n a_{ij}^{(i)} x_j \right)$$

Pivoteo

Si $a_{kk}^{(k)} = 0$, se debe examinar los elementos $a_{ik}^{(k)}$ en las filas $E_i^{(k)}$ con $i = k + 1, \dots, n$. Siendo A no singular, al menos uno de dichos elementos es no nulo. Luego esta fila puede intercambiarse con $E_k^{(k)}$.

Pivoteo Parcial en el paso k calcular $c = \max_{k \leq i \leq n} |a_{ik}^{(k)}|$. Si $|a_{kk}^{(k)}| < c$ luego intercambiar la ecuación $E_k^{(k)}$ por aquella correspondiente a c . Esto reduce errores debido a la supresión de cifras significativas.

Método de Gauss-Jordan Transforma $[A|b] \rightarrow [I|x]$. Requiere un mayor número de operaciones que la eliminación de Gauss.

Factorización LU

Queremos resolver $Ax = b$. Mediante eliminación de Gauss sin pivoteo obtenemos $Ux = g$, donde

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix}, \quad \text{con } u_{ij} = a_{ij}^{(i)}$$

Además introducimos la matriz auxiliar L triangular inferior basada en los coeficientes $m_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}$.

$$L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ m_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & 1 \end{bmatrix}$$

Teorema: sea $A \in \mathbb{R}^{n \times n}$ una matriz no singular, y sean L y U las matrices definidas anteriormente. Luego, si U es generada sin pivoteo se tiene $A = LU$.

—
Luego resolver $Ax = b$ es igual a resolver $LUx = b$, lo cual es equivalente a resolver dos sistemas triangulares.

$Lg = b \rightarrow$ sist. triangular inferior \rightarrow sustitución hacia adelante

$Ux = g \rightarrow$ sist. triangular superior \rightarrow sustitución hacia atrás

Factorización LU con Matriz de Permutación

Definición: una matriz de permutación $P \in \mathbb{R}^{n \times n}$ es una matriz con exactamente una entrada unitaria en cada fila y cada columna, siendo el resto de las entradas nulas.

Propiedad: si P es una matriz de permutación, entonces existe P^{-1} y $P^{-1} = P^T$.

—
Ahora para resolver $Ax = b$ donde A requiere pivoteo, podemos resolver $PA = LU$ donde P incluye los intercambios de filas requeridos por A , y luego resolver $Lg = Pb$ y $Ux = g$ como se mostró antes.

Unicidad de la Factorización LU: si A es tal que la eliminación de Gauss debe realizarse sin pivoteo, luego A puede factorizarse en $A = LU$ y dicha factorización es única.

Definición: una matriz $A \in \mathbb{R}^{n \times n}$ es estrictamente diagonal dominante si

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}|, \quad \forall i \in 1..n$$

Teorema: una matriz A diagonal dominante es no singular. Luego, el sistema $Ax = b$ puede resolverse por eliminación de Gauss sin pivoteo.

Matrices Simétricas

Definición: una matriz A es simétrica si $A = A^T$. Toda matriz simétrica posee autovalores reales.

Definición: una matriz es definida positiva si es simétrica y sus autovalores son todos positivos. Una matriz es semidefinida positiva si es simétrica y sus autovalores son todos no negativos.

Teorema: para matrices reales simétricas los siguientes enunciados son equivalentes y sirven como definición de matriz definida positiva.

- Para todo $x \neq 0$, $x^T Ax > 0$.
- Todos los autovalores de A son positivos.
- $A = B^T B$ para alguna matriz B no singular. B no es única pero existe una única matriz triangular superior R con elementos diagonales positivos tal que $A = R^T R$ (factorización de Cholesky).

Matrices No Simétricas Reales

Toda matriz no simétrica A puede expresarse como $A = M + C$ donde:

- $M = \frac{1}{2}(A + A^T)$ es una matriz simétrica.
- $C = \frac{1}{2}(A - A^T)$ es una matriz antisimétrica.

Nota para toda matriz antisimétrica C y $x \in \mathbb{R}^n$ se cumple $x^T C x = 0$.

Definición: (extensión de matriz definida positiva a matrices no simétricas)

A es definida positiva \iff la matriz simétrica $M = \frac{1}{2}(A + A^T)$ es definida positiva.

Ortogonalización de Gram-Schmidt

Sea $\beta = \{x_1, x_2, \dots, x_n\}$ una base arbitraria, no necesariamente ortogonal de un espacio n -dimensional S . El objetivo es construir una base ortonormal $O = \{u_1, u_2, \dots, u_n\}$ de S . La estrategia consta de construir O secuencialmente de manera que $O_k = \{u_1, u_2, \dots, u_k\}$ es una base ortonormal de $S_k = \text{span}\{x_1, x_2, \dots, x_k\}$ para $k = 1, \dots, n$.

Algoritmo:

$$u_1 = \frac{x_1}{\|x_1\|}, \quad w_k = x_k - \sum_{i=1}^{k-1} (x_k^T u_i) u_i, \quad u_k = \frac{w_k}{\|w_k\|}$$

Factorización QR

Sea $A \in \mathbb{R}^{m \times n}$, $A = [a_1 | a_2 | \dots | a_n]$ una matriz con columnas linealmente independientes. Aplicando Gram-Schmidt resulta una base ortonormal q_1, q_2, \dots, q_n de $\text{span}\{A\}$ donde:

$$q_k = \frac{a_k - \sum_{i=1}^{k-1} (a_k^T q_i) q_i}{\|a_k - \sum_{i=1}^{k-1} (a_k^T q_i) q_i\|}$$

En forma matricial tenemos:

$$\underbrace{[a_1 | a_2 | \dots | a_n]}_A = \underbrace{[q_1 | q_2 | \dots | q_n]}_Q \underbrace{\begin{bmatrix} \sqcup_1 & a_2^T q_1 & a_3^T q_1 & \dots & a_n^T q_1 \\ 0 & \sqcup_2 & a_3^T q_2 & \dots & a_n^T q_2 \\ 0 & 0 & \sqcup_3 & \dots & a_n^T q_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sqcup_n \end{bmatrix}}_R$$

Donde:

- $A \in \mathbb{R}^{m \times n}$ es una matriz de columnas LI.
- $Q \in \mathbb{R}^{m \times n}$ es una base ortonormal de $\text{span}\{A\}$.
- $R \in \mathbb{R}^{n \times n}$ es una matriz triangular superior con elementos diagonales positivos.

Toda matriz $A \in \mathbb{R}^{m \times n}$ con columnas LI puede factorizarse de manera única como $A = QR$. Además, si $A \in \mathbb{R}^{n \times n}$ es no singular, entonces $Q^T = Q^{-1}$. Luego $Ax = b \iff QRx = b \iff Rx = Q^T b$, y este último es un sistema que se resuelve por sustitución regresiva.

Problema de Mínimos Cuadrados

En puntos discretos t_i obtenemos observaciones b_i , resultando un conjunto de pares ordenados $\{(t_1, b_1), \dots, (t_m, b_m)\}$. Supongamos que queremos aproximar los datos mediante una ecuación lineal $y = f(t) = \alpha + \beta t$. Tenemos que el error de aproximar cada punto es $\epsilon_i = |f(t_i) - b_i| = |\alpha + \beta t_i - b_i|$. Luego en forma matricial $\epsilon = Ax - b$ (con $A = [1 | t]$ y $x = [\alpha \ \beta]^T$) y queremos hallar los valores de α y β que minimicen el error al cuadrado $\sum_{i=1}^m \epsilon_i^2 = \epsilon^T \epsilon$.

Problema de mínimos cuadrados general: para $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$, sea $\epsilon = \epsilon(x) = Ax - b$. El problema de los mínimos cuadrados es el de hallar un vector x que minimice $\sum_{i=1}^m \epsilon_i^2 = \sum_{i=1}^m \epsilon_i \epsilon_i^T = (Ax - b)^T (Ax - b)$.

Teorema: el conjunto solución del problema de mínimos cuadrados es el conjunto de soluciones del sistema $A^T A x = A^T b$. Además, existe una única solución $\iff \text{rank}(A) = n$, en cuyo caso $x = (A^T A)^{-1} A^T b$.

Aplicando factorización QR al problema de mínimos cuadrados: El conjunto solución de mínimos cuadrados es el conjunto solución del sistema $A^T A x = A^T b$. Suponga que $\text{rank}(A^{m \times n}) = n$ y sea $A = QR$ Luego $A^T A = (QR)^T QR = R^T Q^T QR = R^T Q^{-1} QR = R^T R$. Por lo tanto $R^T R x = R^T Q^T b$. Siendo R no singular nos queda $R x = Q^T b$ lo cual se resuelve por sustitución regresiva. Siendo $x = R^{-1} Q^T b = (A^T A)^{-1} A^T b$. Si el sistema es consistente, la ecuación anterior da como resultado la solución $Ax - b = 0$. Si no lo es, da la solución de mínimos cuadrados.

Normas Vectoriales y Matriciales

Definición: dado un espacio vectorial V , una función $\|\cdot\| : V \rightarrow \mathbb{R}$ es una norma vectorial si satisface las siguientes propiedades:

1. $\forall x \in V, \|x\| \geq 0 \quad (\|x\| = 0 \iff x = 0)$
2. $\forall x \in V, \lambda \in \mathbb{R}, \|\lambda x\| = |\lambda| \|x\|$
3. $\forall x, y \in V, \|x + y\| \leq \|x\| + \|y\|$

Norma Euclídea: $\|x\|_2 = \sqrt{x^T x}$

Norma Infinito: $\|x\|_\infty = \max |x_i|$

Norma l1: $\|x\|_1 = \sum |x_i|$

Normas Matriciales para $A \in \mathbb{R}^{m \times n}$:

- $\|A\| = 0 \iff A = 0$
- $\forall \lambda \in \mathbb{R}, \|\lambda A\| = |\lambda| \|A\|$
- $\forall A, B, \|A + B\| \leq \|A\| + \|B\|$

Definición: para una matriz cuadrada A , se dice que la norma matricial $\|\cdot\|$ es submultiplicativa si $\forall A, B, \|AB\| \leq \|A\| \cdot \|B\|$.

Definición: dada una norma vectorial se define la norma matricial inducida para $A \in \mathbb{R}^{m \times n}$ como $\|A\| = \sup\{\|Ax\|/\|x\| : x \neq 0, x \in \mathbb{R}^n\}$.

Teorema: la norma matricial inducida es submultiplicativa.

Teorema: sea $\|\cdot\|$ una norma vectorial, luego $\forall x \in \mathbb{R}^n, \|Ax\| \leq \|A\| \cdot \|x\|$.

Demostración: para $x = 0$ se verifica trivialmente. Luego supongo $x \neq 0$ y sea $v = x/\|x\|$ y $\|v\| = 1$. Se tiene

$$\|Ax\| = \left\| Ax \frac{\|x\|}{\|x\|} \right\| = \|x\| \cdot \|Av\| \leq \|x\| \cdot \|v\| \cdot \|A\| = \|x\| \cdot \|A\|$$

Estabilidad de la Resolución de Sistemas Lineales

Considerar el sistema $Ax = b$ y el sistema perturbado $A\tilde{x} = \tilde{b}$.

Teorema: sea $A \in \mathbb{R}^{n \times n}$ no singular. Luego las soluciones de $Ax = b$ y $A\tilde{x} = \tilde{b}$ satisfacen

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \frac{\|b - \tilde{b}\|}{\|b\|}$$

Demostración: tenemos que $Ax - A\tilde{x} = b - \tilde{b}$, y como A es no singular obtenemos $x - \tilde{x} = A^{-1}(b - \tilde{b})$. Luego, usando la propiedad anterior obtenemos $\|x - \tilde{x}\| \leq \|A^{-1}\| \cdot \|b - \tilde{b}\|$. Dividimos cada lado por $\|x\|$ obteniendo $\|x - \tilde{x}\|/\|x\| \leq \|A^{-1}\| \cdot \|b - \tilde{b}\|/\|x\|$. Luego $\|x - \tilde{x}\|/\|x\| \leq \|A\| \cdot \|A^{-1}\| \cdot \|b - \tilde{b}\|/(\|A\| \cdot \|x\|)$ y como $\|b\| = \|Ax\| \leq \|A\| \cdot \|x\|$ llegamos al resultado buscado.

Nota: El número $\|A\| \cdot \|A^{-1}\|$ se conoce como el número de condición de A ($K(A)$), y de lo anterior se desprende que:

$$\frac{\|\Delta x\|}{\|x + \Delta x\|} \leq K(A) \frac{\|\Delta A\|}{\|A\|}$$

Lema: $K(A) \geq 1$

Demostración: $1 = \|I\| = \|AA^{-1}\| \leq \|A\| \cdot \|A^{-1}\| = K(A)$.

Métodos Iterativos

- Generan una sucesión $\{x^{(k)}\}$ que converge a la solución.
- Para n grande, la eliminación de Gauss requiere $\frac{2}{3}n^3$ operaciones, mientras que los métodos iterativos requieren $\approx n^2$.

Esquema general de los Métodos Iterativos

Sea $A \in \mathbb{R}^{n \times n}$, $Ax = b$ el sistema a resolver. Sea $N \in \mathbb{R}^{n \times n}$ no singular. Luego $Nx = Nx - Ax + b$, de donde obtenemos un proceso iterativo de la forma $Nx^{(k+1)} = (N - A)x^{(k)} + b$, de donde obtenemos $x^{(k+1)} = (I - N^{-1}A)x^{(k)} + N^{-1}b$ y solución del sistema cumple que $x = (I - N^{-1}A)x + N^{-1}b$.

Si $A = L + D + U$, luego:

- El método de Jacobi utiliza $N = D$.
- El método de Gauss-Seidel utiliza $N = L + D$.

Error de aproximación: $e^{(k)} = x - x^{(k)}$ y restando $x = (I - N^{-1}A)x + N^{-1}b$ y $x^{(k+1)} = (I - N^{-1}A)x^{(k)} + N^{-1}b$ obtenemos $e^{(k+1)} = (I - N^{-1}A)e^{(k)}$.

Teorema: si $\|I - N^{-1}A\| < 1$ entonces la sucesión $\{x^{(k)}\}$ converge a la solución del sistema $Ax = b$ para cualquier estimación inicial $x^{(0)}$.

Demostración:

$$\begin{aligned} \|e^{(k+1)}\| &= \|(I - N^{-1}A)e^{(k)}\| \leq \|I - N^{-1}A\| \cdot \|e^{(k)}\| \\ &\leq \|I - N^{-1}A\| \cdot \|(I - N^{-1}A)e^{(k-1)}\| = \|I - N^{-1}A\|^2 \cdot \|e^{(k-1)}\| \\ &\leq \dots \leq \|I - N^{-1}A\|^{k+1} \cdot \|e^{(0)}\| \end{aligned}$$

Como además sabemos que $\|I - N^{-1}A\| < 1$, luego $\|I - N^{-1}A\|^{k+1} \rightarrow 0$ cuando $k \rightarrow \infty$. Finalmente $\lim_{k \rightarrow \infty} \|e^{(k+1)}\| = 0 \implies x^{(\infty)} \rightarrow x$.

Estabilidad Asintótica de un Sistema Lineal Discreto

Teorema: sea $B \in \mathbb{R}^{n \times n}$. El proceso iterativo $x^{(k+1)} = Bx^{(k)}$ converge a $\underline{x} = \underline{0}$ para todo valor inicial $x^{(0)} \iff \rho(B) < 1$, donde $\rho(B)$ es el radio espectral de B ($\max|\lambda_i|$).

Corolario: la fórmula de iteración $Nx^{(k+1)} = (N - A)x^{(k)} + b$ dará lugar a una sucesión que converge a la solución del sistema $Ax = b$ para cualquier $x^{(0)} \iff \rho(I - N^{-1}A) < 1$.

Teorema: si la matriz A es diagonal dominante, luego la sucesión $\{x^{(k)}\}$ generada por el método de Jacobi converge a la solución del sistema $Ax = b$ para todo $x^{(0)}$ inicial.

Teorema: si la matriz A es diagonal dominante, luego la sucesión $\{x^{(k)}\}$ generada por el método de Gauss-Seidel converge a la solución del sistema $Ax = b$ para todo $x^{(0)}$ inicial.

Método de Sobrerelajación (SOR)

Este método permite mejorar la convergencia usando relajación. La relajación representa una ligera modificación del método de Gauss-Seidel y ésta permite mejorar la convergencia en algunos casos. Después de que se calcula cada nuevo valor de x , éste valor se modifica mediante un promedio ponderado de los resultados de las iteraciones anterior y actual.

$$x_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \frac{\omega}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{i-1} a_{ij}x_j^{(k)} \right)$$

Donde ω es el factor de relajación:

- Si $\omega = 1$, tenemos el método de Gauss-Seidel
- Si $0 < \omega < 1$, tenemos el método de Subrelajación (útil cuando Gauss-Seidel no converge).
- Si $\omega > 1$, tenemos el método de Sobrerelajación (acelera la velocidad de convergencia cuando Gauss-Seidel converge).

—

Si reescribimos la ecuación anterior, obtenemos:

$$a_{ii}x_i^{(k+1)} + \omega \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} = (1 - \omega)a_{ii}x_i^{(k)} - \omega \sum_{j=i+1}^{i-1} a_{ij}x_j^{(k)} + \omega b_i$$

En forma matricial:

$$(D + \omega L)x^{(k+1)} = [(1 - \omega)D - \omega U]x^{(k)} + \omega b$$

Luego, si existe $(D + \omega L)^{-1}$, entonces $x^{(k+1)} = T_\omega x^{(k)} + C_\omega$, donde $T_\omega = (D + \omega L)^{-1}[(1 - \omega)D - \omega U]$ y $C_\omega = \omega(D + \omega L)^{-1}b$. Luego el error está dado por $e^{(k+1)} = T_\omega e^{(k)}$, de donde se desprende que el método SOR converge a la solución del $Ax = b$ para todo valor inicial $\iff \rho(T_\omega) < 1$.

Teorema: sea $A \in \mathbb{R}^{n \times n}$, luego $\rho(A) \leq \|A\|$ para cualquier norma matricial submultiplicativa.

Demostración: sea (λ, v) un par autovalor-autovector de A y $\underline{X} = [v \mid 0 \mid \dots \mid 0] \in \mathbb{R}^{n \times n}$. Luego $\lambda \underline{X} = A \underline{X}$, de donde tenemos que $|\lambda| \cdot \|\underline{X}\| = \|\lambda \underline{X}\| = \|A \underline{X}\| \leq \|A\| \cdot \|\underline{X}\|$. Luego $|\lambda| \leq \|A\|$ para todo $\lambda \in \sigma(A)$ (espectro de A), por lo tanto $\rho(A) \leq \|A\|$.

Aproximación de Autovalores

Definición: sea $A \in \mathbb{R}^{n \times n}$. Si existe un número λ y un vector $v \neq 0$ tales que $Av = \lambda v$ decimos que λ es un autovalor de A y que v es un autovector de A . Además, si v es un autovector de A , luego αv es un autovector de A para cualquier $\alpha \in \mathbb{R} \neq 0$.

¿Cómo calcular λ y v ? Sabemos que $Av = \lambda v$, luego $(\lambda I - A)v = 0$ (con $v \neq 0$). Llamamos polinomio característico de A a $f(\lambda) = \det(\lambda I - A)$. Además, si $A \in \mathbb{R}^{n \times n}$, luego $f(\lambda)$ es un polinomio de grado n .

Teorema: sea $A \in \mathbb{R}^{n \times n}$ una matriz simétrica real. Luego existe un conjunto de pares autovalor-autovector $\lambda_i, v^{(i)}, i = 1, \dots, n$ que satisfacen:

1. Los autovalores $\lambda_1, \dots, \lambda_n$ son las raíces del polinomio característico de A . Todos son números reales.
2. Los autovectores $v^{(1)}, \dots, v^{(n)}$ ortogonales entre si, y pueden elegirse de longitud 1. Es decir, $v^{(i)T} v^{(j)} = 0$ y $v^{(i)T} v^{(i)} = 1$.
3. Para cada vector $x = [x_1, x_2, \dots, x_n]^T$ existe un único vector $c = [c_1, c_2, \dots, c_n]$ tal que $x = c_1 v^{(1)} + \dots + c_n v^{(n)}$. Las constantes están dadas por $c_i = \sum_{j=1}^n x_j v_j^{(i)} = x^T v^{(i)}$ donde $\text{long}(v^{(i)}) = 1$.
4. Definir la matriz $U = [v^{(1)} \ v^{(2)} \ \dots \ v^{(n)}]$. Luego $U^T A U = D$ matriz diagonal con $D_{ii} = \lambda_i$. Además $U U^T = U^T U = I$, y $A = U D U^T$.

Círculos de Gerschgorín

Definición: sea $A \in \mathbb{R}^{n \times n}$, luego para $i = 1, \dots, n$ sean:

$$r_i = \sum_{j=1, j \neq i}^n |a_{ij}|, \quad c_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i\}$$

Teorema: sea $A \in \mathbb{R}^{n \times n}$ y sea λ un autovalor de A . Luego $\lambda \in C_i$ para algún $i = 1, \dots, n$.

Demostración: sea λ un autovalor de A y v el autovector asociado. Sea k la componente de v tal que $|v_k| = \|v\|_\infty$. Luego:

$$\begin{aligned}
Av = \lambda v &\implies \sum_{j=1}^n a_{kj}v_j = \lambda v_k \quad (\text{k-ésimo componente}) \\
&\implies \sum_{j=1, j \neq k}^n a_{kj}v_j + a_{kk}v_k = \lambda v_k \\
&\implies (\lambda - a_{kk})v_k = \sum_{j=1, j \neq k}^n a_{kj}v_j \\
&\implies |\lambda - a_{kk}| \cdot |v_k| \leq \sum_{j=1, j \neq k}^n |a_{kj}| \cdot |v_j| \leq r_k \|v\|_\infty \\
&\implies |\lambda - a_{kk}| \leq r_k
\end{aligned}$$

—

Método de la Potencia

Sea $A \in \mathbb{R}^{n \times n}$ una matriz simétrica y sean $\lambda_1, \dots, \lambda_n$ sus autovalores tal que $|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$, es decir, suponemos que existe un único autovalor de módulo máximo. Luego sea $\{v^{(1)}, \dots, v^{(n)}\}$ la base de autovectores correspondiente. Sea $z^{(1)}$ una estimación inicial de $v^{(1)}$. Definimos:

$$\omega^{(n+1)} = Az^{(n)}, \quad z^{(n+1)} = \frac{\omega^{(n+1)}}{\|\omega^{(n+1)}\|}$$

Entonces resulta que $z^{(n)} \rightarrow v^{(1)} / \|v^{(1)}\|_\infty$ cuando $n \rightarrow \infty$, y eligiendo una componente k no nula de $\omega^{(n-1)}$ luego $\lambda^{(n)} = \omega^{(n)} / z_k^{(n-1)} \rightarrow \lambda_1$ cuando $n \rightarrow \infty$.

Interpolación y Ajuste de Curvas

Dados $n + 1$ puntos distintos $a \leq x_1 < \dots < x_{n+1} \leq b$ de un intervalo $[a, b]$, llamados nodos de la interpolación, y $n + 1$ números reales y_1, \dots, y_{n+1} llamados valores de la interpolación, se trata de hallar una función P tal que $P(x_i) = y_i$, $i = 1, \dots, n + 1$.

Interpolación Polinomial

Dados $n + 1$ puntos $\{(x_i, y_i) : y_i = f(x_i), i = 0, \dots, n\}$, buscamos encontrar un polinomio $P(x)$ que interpole los datos, es decir que $P(x_i) = y_i$, $i = 1, \dots, n$.

Teorema: (Existencia y Unicidad del Polinomio Interpolante) dados $n + 1$ puntos distintos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ con x_0, \dots, x_n números distintos, existe un polinomio $P(x)$ de grado menor o igual a n que interpola dichos puntos. Además, dicho polinomio es único en el conjunto de polinomios de grado menor o igual a n .

Método de Interpolación de Lagrange

Consideramos el polinomio de grado máximo n que pasa por los puntos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$. Para $k = 0, \dots, n$ definimos:

$$L_k(x) = \frac{(x - x_0) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n)}{(x_k - x_0) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n)} = \prod_{i=0, i \neq k}^n \frac{(x - x_i)}{(x_k - x_i)}$$

Donde $L_k(x)$ satisface $L_k(x_k) = 1$ y $L_k(x_i) = 0$ si $i \neq k$. Luego el polinomio interpolador de Lagrange está dado por:

$$P(x) = L_0(x)y_0 + L_1(x)y_1 + \cdots + L_n(x)y_n = \sum_{k=0}^n L_k(x)y_k$$

Notar que $P(x_i) = y_i$, $i = 0, \dots, n$.

Método de Interpolación por Diferencias Divididas de Newton

Dados los $n + 1$ puntos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, expresar el polinomio interpolador de la forma:

$$P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \cdots + a_n(x - x_0)(x - x_1) \cdots (x - x_{n-1})$$

Luego vemos que:

$$\begin{aligned} P_1(x) &= a_0 + a_1(x - x_0) \\ P_1(x) &= \underbrace{a_0 + a_1(x - x_0)}_{P_1(x)} + a_2(x - x_0)(x - x_1) \\ &\vdots \\ P_n(x) &= P_{n-1}(x) + a_n(x - x_0) \cdots (x - x_{n-1}) \end{aligned}$$

Imponiendo las Condiciones de Interpolación: queremos que $P_n(x_i) = f(x_i) = y_i$, $i = 1, \dots, n$. Luego:

$$\begin{aligned} P_n(x_0) &= a_0 = y_0 \\ P_n(x_1) &= y_0 + a_1(x_1 - x_0) \implies a_1 = \frac{y_1 - y_0}{x_1 - x_0} \\ &\vdots \end{aligned}$$

Los coeficientes a_i se pueden calcular introduciendo la idea de diferencia dividida.

- Diferencia dividida de primer orden: $f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{(x_1 - x_0)} = (y_1 - y_0)/(x_1 - x_0)$
- Diferencia dividida de segundo orden: $f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{(x_2 - x_0)}$
- Diferencia dividida de orden k : $f[x_i, \dots, x_{i+k}] = \frac{f[x_{i+1}, \dots, x_{i+k}] - f[x_i, x_{i+k-1}]}{(x_{i+k} - x_i)}$

Propiedad: sea (i_0, i_1, \dots, i_n) una permutación de los enteros $(0, 1, \dots, n)$, luego se demuestra que $f[x_{i_0}, x_{i_1}, \dots, x_{i_n}] = f[x_0, x_1, \dots, x_n]$.

Fórmula de la Interpolación por Diferencias Divididas de Newton

$$P_n(x) = f(x_0) + (x - x_0)f[x_0, x_1] + \cdots + (x - x_0) \cdots (x - x_{n-1})f[x_0, \dots, x_n] = \sum_{i=0}^n \left(\prod_{j=0}^{i-1} (x - x_j) \right) f[x_0, \dots, x_n]$$

Teorema: suponga que f está definida en $[a, b]$ y que $\{x_0, x_1, \dots, x_n\}$ son valores distintos en $[a, b]$. El polinomio de grado menor o igual a k que interpola $f(x)$ en $\{x_i, \dots, x_{i+k}\} \subseteq \{x_0, \dots, x_n\}$ está dado por:

$$P_{i,k}(x) = f(x_i) + (x - x_i)f[x_i, x_{i+1}] + \dots + (x - x_i) \cdots (x - x_{i+k-1})f[x_i, \dots, x_{i+k}]$$

Teorema: (de Rolle) sea f continua en $[a, b]$ y diferenciable en (a, b) . Si $f(a) = f(b)$, entonces existe $c \in (a, b)$ tal que $f'(c) = 0$.

Teorema: (generalizado de Rolle) sea f continua en $[a, b]$ y diferenciable n veces en (a, b) . Si $f(x)$ se anula en los $n + 1$ números distintos $x_0, x_1, \dots, x_n \in [a, b]$, entonces existe $c \in (a, b)$ tal que $f^{(n)}(c) = 0$.

Error en la Interpolación Polinomial

Teorema: sean $x_0, x_1, \dots, x_n \in [a, b]$ números distintos, y sea $f(x)$ diferenciable n veces en $[a, b]$. Luego para todo $x \in [a, b]$ existe $\xi(x) \in (a, b)$ tal que:

$$f(x) - P(x) = \frac{(x - x_0)(x - x_1) \cdots (x - x_n)}{(n + 1)!} f^{(n+1)}(\xi(x))$$

donde $P(x)$ es un polinomio interpolante de grado menor o igual a n .

Acotación del Error (Caso General)

$$f(x) - P(x) = \frac{(x - x_0)(x - x_1) \cdots (x - x_n)}{(n + 1)!} f^{(n+1)}(cx) = \frac{\psi_n(x)}{(n + 1)!} f^{(n+1)}(cx)$$

Para x_0, x_1, \dots, x_n distintos en $[a, b]$ y $x \in [a, b]$ la cota de error $|f(x) - P_n(x)|$ está dada por:

$$|f(x) - P_n(x)| \leq \max_{a \leq x \leq b} |f(x) - P_n(x)| \leq \frac{1}{(n + 1)!} \max_{a \leq x \leq b} |\psi_n(x)| \max_{a \leq x \leq b} |f^{(n+1)}(x)|$$

Error para Nodos Uniformemente Espaciados

$$h = \frac{b - a}{n} \quad x_i = a + ih$$

Luego $\psi_n(x) = x(x - h)(x - 2h) \cdots (x - b) = \prod_{i=0}^n (x - ih)$. Usando nodos uniformemente espaciado, en general, el error no está uniformemente distribuido. Además, se tiene que el error no siempre tiende a cero cuando se aumenta la cantidad de puntos.

Teorema: (de Aproximación de Weierstrass) sea $f(x)$ una función continua en $[a, b]$ y sea $\epsilon > 0$. Luego existe un polinomio $P_n(x)$ tal que $\max_{a \leq x \leq b} |f(x) - P(x)| \leq \epsilon$.

Aproximación con Menor V.A. Máximo

Dada una función $f(x)$ continua en $[a, b]$ queremos aproximarla con un polinomio $P(x)$ tal que minimice el error de aproximación $E(P) = \max_{a \leq x \leq b} |f(x) - P(x)|$. Definimos el error minimax como $\zeta(f) = \min_{gr(P) \leq n} E(P)$ (esto requiere optimización no lineal). Luego denotamos este polinomio minimax como $m_n(x)$, y se tiene $E(m_n) = \zeta(f)$.

Polinomio de Chebychev

Para $n \geq 0$ definimos la función $T_n(x) = \cos(n \cdot \cos^{-1}(x))$ para $-1 \leq x \leq 1$. Luego, T_n verifica que $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$.

Propiedades de T_n :

- $|T_n(x)| \leq 1$, $-1 \leq x \leq 1$
- $T_n(x) = 2^{n-1} + \text{términos de menor grado}$

Las raíces de T_n se usan para encontrar los valores de x_0, \dots, x_n que minimizan el error de interpolación de grado menor o igual a $n - 1$.

Aproximación de Mínimos Cuadrados

Sea $y = g(x)$ una relación desconocida entre las variables x e y . Experimentalmente se obtienen $\{(x_1, y_1), \dots, (x_m, y_m)\}$ donde $y_i = g(x_i) + v_i$ con v_i el error de medición. A partir de los datos queremos aproximar $g(x)$ mediante una función $f(x)$ de la forma $f(x) = a_1\varphi_1(x) + a_2\varphi_2(x) + \dots + a_p\varphi_p(x)$ donde a_i son números y φ_i son funciones dadas. Es decir, queremos hallar los valores de a_1, \dots, a_p que minimizan $G(a_1, \dots, a_p) = \sum_{j=1}^m [(a_1\varphi_1(x_j) + \dots + a_p\varphi_p(x_j)) - y_j]^2$. El mínimo se satisface cuando $\delta G / \delta a_i = 0$, $i = 1, \dots, p$.

Aproximación Polinomial de Mínimos Cuadrados: $\varphi_i = x^{i-1}$.

Integración Numérica

Dada una función $f : [a, b] \rightarrow \mathbb{R}$ se quiere calcular la integral definida $I(f) = \int_a^b f(x)dx = F(b) - F(a)$, donde $F(x)$ es cualquier antiderivada de $f(x)$. Una aproximación de la misma resulta $\sum_{i=0}^n a_i f(x_i)$.

Integración Numérica Basada en Polinomios Interpolantes

Sean $\{x_0, \dots, x_n\}$ $n + 1$ nodos distintos en $[a, b]$. Tenemos que:

$$f(x) = P_n(x) + \prod_{i=0}^n (x - x_i) \frac{f^{(n+1)}(\xi(x))}{(n+1)!}$$

Donde $P_n(x) = \sum_{i=0}^n f(x_i)L_i(x)$. Integrando en $[a, b]$ tenemos:

$$\begin{aligned} \int_a^b f(x)dx &= \int_a^b \sum_{i=0}^n f(x_i)L_i(x) + \int_a^b \prod_{i=0}^n (x - x_i) \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \\ &= \sum_{i=0}^n a_i f(x_i) + \underbrace{\frac{1}{(n+1)!} \int_a^b \prod_{i=0}^n (x - x_i) f^{(n+1)}(\xi(x))}_{\text{error de integración}} \end{aligned}$$

Donde $\xi(x) \in [a, b]$ para todo $x \in [a, b]$ y $a_i = \int_a^b L_i(x)dx$.

Teorema: (del Valor Intermedio Ponderado para Integrales) si $f \in [a, b]$, y g es integrable en $[a, b]$ y no cambia de signo en $[a, b]$, entonces existe un número $c \in (a, b)$ tal que $\int_a^b f(x)g(x)dx = f(c) \int_a^b g(x)dx$.

Regla del Trapecio

Aproximamos $f(x)$ mediante un polinomio lineal. Sean $x_0 = a$, $x_1 = b$ y $h = b - a$. Por Lagrange tenemos que $P_1(x) = \frac{x-x_1}{x_0-x_1}f(x_0) + \frac{x-x_0}{x_1-x_0}f(x_1)$. Luego:

$$\int_a^b f(x)dx = \int_{x_0}^{x_1} \left[\frac{(x-x_1)}{(x_0-x_1)}f(x_0) + \frac{(x-x_0)}{(x_1-x_0)}f(x_1) \right] dx + \frac{1}{2} \int_{x_0}^{x_1} f''(\xi(x))(x-x_0)(x-x_1)dx$$

Donde:

$$\begin{aligned} \int_{x_0}^{x_1} \left[\frac{(x-x_1)}{(x_0-x_1)}f(x_0) + \frac{(x-x_0)}{(x_1-x_0)}f(x_1) \right] dx &= \left[\frac{(x-x_1)^2}{2(x_0-x_1)}f(x_0) + \frac{(x-x_0)^2}{2(x_1-x_0)}f(x_1) \right]_{x_0}^{x_1} \\ &= \frac{(x_1-x_0)}{2}(f(x_0) + f(x_1)) = \frac{h}{2}(f(x_0) + f(x_1)) \end{aligned}$$

Además, como $(x-x_0)(x-x_1)$ no cambia de signo en $[x_0, x_1]$, luego podemos aplicar el teorema anterior para obtener:

$$\begin{aligned} \frac{1}{2} \int_{x_0}^{x_1} f''(\xi(x))(x-x_0)(x-x_1)dx &= \frac{1}{2}f''(c) \int_{x_0}^{x_1} (x-x_0)(x-x_1)dx \quad \text{p.a. } c \in [x_0, x_1] \\ &= \frac{1}{2}f''(c) \left[\frac{x^3}{3} - \frac{(x_1+x_0)x^2}{2} + x_0x_1x \right]_{x_0}^{x_1} \\ &= f''(c) \left[\left(\frac{x_1^3}{3} - \frac{x_1^2(x_1+x_0)}{2} + x_0x_1^2 \right) - \left(\frac{x_0^3}{3} - \frac{x_0^2(x_1+x_0)}{2} + x_0^2x_1 \right) \right] \\ &= \frac{1}{2}f''(c) \left[\left(-\frac{x_1^3}{6} + \frac{x_0x_1^2}{2} \right) - \left(-\frac{x_0^3}{6} + \frac{x_0^2x_1}{2} \right) \right] \\ &= \frac{f''(c)}{12} [x_1^3 - 3x_0x_1^2 + 3x_0^2x_1 - x_0^3] \\ &= -\frac{f''(c)}{12}(x_1-x_0)^3 = -\frac{h^3}{12}f''(c) \end{aligned}$$

Finalmente, la regla del trapecio queda dada por:

$$\int_a^b f(x)dx = \underbrace{\frac{h}{2}[f(x_0) + f(x_1)]}_{\text{integral aproximada}} - \underbrace{\frac{h^3}{12}f''(c)}_{\text{error de aproximación}}$$

Método Compuesto del Trapecio

Utiliza varios subintervalos de igual longitud. Sea n el número de subintervalos, luego $h = (b-a)/n$ y $x_j = a + jh$. Tenemos que la aproximación por n trapecios T_n resulta:

$$\begin{aligned} T_n(f) &= h \left(\frac{f(x_0) + f(x_1)}{2} \right) + h \left(\frac{f(x_1) + f(x_2)}{2} \right) + \dots + h \left(\frac{f(x_{n-1}) + f(x_n)}{2} \right) \\ &= h \left[\frac{1}{2}f(x_0) + f(x_1) + \dots + f(x_{n-1}) + \frac{1}{2}f(x_n) \right] \end{aligned}$$

Teorema: (Error de la Integración Trapezoidal) sea $f : \mathbb{R} \rightarrow \mathbb{R}$, derivable 2 veces en $[a, b]$, luego el error de integración usando n subintervalos resulta:

$$E_n^T(f) = \int_a^b f(x)dx - T_n(f) = -\frac{h^2(b-a)}{12}f''(cn) \text{ p.a. } cn \in [a, b]$$

Demostración:

- Si $n = 1 \implies a = x_0, b = x_1$ y $h = b - a$. Luego $E_1^T(f) = -\frac{h^3}{12}f''(c)$.
- Si $n > 1 \implies h = (b-a)/n, x_j = a + jh$. Luego $E_n^T(f) = -\frac{h^3}{12}f''(\gamma_1) - \frac{h^3}{12}f''(\gamma_2) - \dots - \frac{h^3}{12}f''(\gamma_n)$, donde $x_{j-1} \leq \gamma_j \leq x_j$. Reagrupando y multiplicando por $\frac{n}{n}$ obtenemos:

$$E_n^T(f) = -\frac{h^3}{12}n \left[\frac{f''(\gamma_1) + \dots + f''(\gamma_n)}{n} \right]$$

Si llamamos $\xi_n = (f''(\gamma_1) + \dots + f''(\gamma_n))/n$ luego se cumple que $\min_{a \leq x \leq b} f''(x) \leq \xi_n \leq \max_{a \leq x \leq b} f''(x)$. Por hipótesis, $f''(x)$ es continua en $[a, b]$, luego existe $cn \in [a, b]$ tal que $f''(cn) = \xi_n$. Como además $hn = b - a$ tenemos que $E_n^T(f) = -\frac{h^2(b-a)}{12}f''(cn)$.

Estimación del Error Trapezoidal: Sabemos que $E_n^T(f) = -\frac{h^2}{12}[f''(\gamma_1)h + \dots + f''(\gamma_n)h]$, donde $[f''(\gamma_1)h + \dots + f''(\gamma_n)h]$ es una aproximación de $\int_a^b f''(x)dx = f'(b) - f'(a)$. Luego

$$E_n^T(f) \approx -\frac{h^2}{12}[f'(b) - f'(a)]$$

Regla de Simpson

Aproximamos $f(x)$ mediante el polinomio de interpolación de Lagrange de grado 2 con los nodos $x_0 = a, x_1 = a + h$ y $x_2 = b$, donde $h = (b - a)/2$. Luego resulta:

$$\begin{aligned} \int_a^b f(x)dx &= \int_{x_0}^{x_2} \underbrace{\left[\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f(x_2) \right]}_{P_2(x)} dx \\ &+ \int_{x_0}^{x_2} \underbrace{\frac{(x-x_0)(x-x_1)(x-x_2)}{6}f^{(3)}(\xi(x))}_{\text{aproximación del error}} dx \end{aligned}$$

Luego, resulta de operar algebraicamente que:

$$\begin{aligned} \int_{x_0}^{x_2} P_2(x)dx &= \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] \\ \int_{x_0}^{x_2} \frac{(x-x_0)(x-x_1)(x-x_2)}{6}f^{(3)}(\xi(x))dx &= -\frac{h^5}{90}f^{(4)}(\xi) \text{ p.a. } \xi \in [x_0, x_2] \end{aligned}$$

Y resulta:

$$\int_a^b f(x)dx = \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] - \frac{h^5}{90}f^{(4)}(\xi) \text{ p.a. } \xi \in [x_0, x_2]$$

Método Compuesto de Simpson

Utiliza varios subintervalos de igual longitud. Sea n el número de subintervalos, luego $h = (b - a)/n$ y $x_j = a + jh$. Tenemos que la aproximación S_n resulta:

$$\begin{aligned}\int_a^b f(x)dx &= \int_{x_0}^{x_2} f(x)dx + \cdots + \int_{x_{n-2}}^{x_n} f(x)dx \\ &= \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] + \frac{h}{3}[f(x_2) + 4f(x_3) + f(x_4)] + \cdots + \frac{h}{3}[f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)] \\ &= \frac{h}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)]\end{aligned}$$

Teorema: (Error de la Integración Compuesta de Simpson) sea $f : \mathbb{R} \rightarrow \mathbb{R}$, derivable 4 veces en $[a, b]$, luego el error de integración usando $n \in \mathbb{P}^+$ subintervalos resulta:

$$E_n^S(f) = \int_a^b f(x)dx - S_n(f) = -\frac{h^4(b-a)}{180}f^{(4)}(cn) \text{ p.a. } cn \in [a, b]$$

Demostración: Sabemos que $E_n^S(f) = -\frac{h^5}{90}f^{(4)}(\gamma_1) - \frac{h^5}{90}f^{(4)}(\gamma_2) - \cdots - \frac{h^5}{90}f^{(4)}(\gamma_n)$, donde $x_{j-1} \leq \gamma_j \leq x_j$. Reagrupando y multiplicando por $\frac{(n/2)}{(n/2)}$ obtenemos:

$$E_n^S(f) = -\frac{h^5}{90} \left(\frac{n}{2}\right) \left[\frac{f^{(4)}(\gamma_1) + \cdots + f^{(4)}(\gamma_n)}{(n/2)}\right]$$

Si llamamos $\xi_n = (f^{(4)}(\gamma_1) + \cdots + f^{(4)}(\gamma_n))/(n/2)$ luego se cumple que $\min_{a \leq x \leq b} f^{(4)}(x) \leq \xi_n \leq \max_{a \leq x \leq b} f^{(4)}(x)$. Por hipótesis, $f^{(4)}(x)$ es continua en $[a, b]$, luego existe $cn \in [a, b]$ tal que $f^{(4)}(cn) = \xi_n$. Como además $hn = b - a$ tenemos que $E_n^S(f) = -\frac{h^4(b-a)}{180}f^{(4)}(cn)$.

Estimación del Error de Simpson: Sabemos que $E_n^S(f) = -\frac{h^4}{90} \frac{1}{2}[2nf^{(4)}(\gamma_1) + \cdots + 2nf^{(4)}(\gamma_n)]$, donde $[2nf^{(4)}(\gamma_1) + \cdots + 2nf^{(4)}(\gamma_n)]$ es una aproximación de $\int_a^b f^{(4)}(x)dx = f^{(3)}(b) - f^{(3)}(a)$. Luego

$$E_n^S(f) \approx -\frac{h^4}{180}[f^{(3)}(b) - f^{(3)}(a)]$$

Integración Numérica en Dominio Bidimensional

Sea desea calcular la integral de una función $f(x, y)$ en un dominio bidimensional $Q = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c(x) \leq y \leq d(x)\}$. Es decir:

$$I = \int_a^b \int_{c(x)}^{d(x)} f(x, y)dydx$$

Para esto, definimos $G(x) = \int_{c(x)}^{d(x)} f(x, y)dy$, luego $I = \int_a^b G(x)dx$, lo cual puede aproximarse como $I \approx \sum_{i=1}^n \omega_i G(x_i)$, donde ω_i es el factor de ponderamiento del método utilizado, y x_i son los nodos equidistantes en x . Por otra parte tenemos que $G(x_i) \approx \sum_{j=1}^m a_{ij} f(x_i, y_j)$.

Cota del Error:

$$|E_n^S| \leq \frac{hx^4(b-a)hy^4(d-c)}{180^2} \cdot \max_{(x,y) \in Q} \left| \frac{\delta^4 f}{\delta x^4} \right| \cdot \max_{(x,y) \in Q} \left| \frac{\delta^4 f}{\delta y^4} \right|$$