Universidad de Extremadura
Departamento de Matemáticas

Apuntes de Cálculo Numérico

Pedro Martín Jiménez

http://cum.unex.es/profes/profes/pjimenez/

Badajoz, septiembre 2005

Índice

Introducción

El texto que sigue no es definitivo ni exhaustivo. Se irá reformando a medida que se detecten fallos o se modifique el contenido para mejorarlo. Debeis emplearlo como una ayuda para preparar la asignatura Análisis Matemático I y como complemento de las notas que tomeis en clase.

Espero que os sirva.

Pedro Martín.

Capítulo 1

Errores, redondeo, estabilidad, condicionamiento.

1.1 Cifras significativas. Exactitud y precisión. Errores.

Cualquier número real x puede representarse en forma decimal con un número finito o infinito de dígitos:

$$x = \alpha_m \alpha_{m-1} \dots \alpha_1 \alpha_0 \quad . \quad \beta_1 \beta_2 \dots \beta_n \dots \quad \text{con } \alpha_i, \beta_j \in 0, 1, 2, \dots, 9$$

Con la expresión anterior queremos representar que

$$x = \alpha_m 10^m + \alpha_{m-1} 10^{m-1} + \dots + \alpha_1 10 + \alpha_0 10^0 + \beta_1 10^{-1} + \beta_2 10^{-2} + \dots + \beta_n 10^{-n} + \dots$$

El Cálculo Numérico consiste en desarrollar métodos para encontrar soluciones lo más aproximadas posibles a la verdadera solución de un cierto problema. Se trabaja, por tanto, con números aproximados en lugar de números exactos, por lo que se utilizan aproximaciones con un número finito de decimales.

Toda expresión aproximada tiene un número determinado de **cifras significativas**. El concepto de cifra significativa intenta trasmitir la idea de cuando una cifra que aparece en la expresión decimal trasmite información "esencial"y "de fiar"sobre el número que se intenta aproximar. La significación de una cifra depende del contexto y no solo de la expresión decimal en la que aparece.

Ejemplo: Cuando decimos "circulamos a 48.5 Km/hora"porque vemos que la aguja del velocímetro del coche están entre 48 y 49, en nuestra aproximación (48.5) al valor verdadero de la velocidad, estamos manejando solo dos cifras significativas, las que permiten situar la velocidad entre 48 y 49. La tercera cifra (.5) es una estimación que podríamos evitar, pues no es "de fiar". No es significativa. El velocímetro (para velocidades menores de 100) solo permite aproximaciones con dos cifras significativas. Si hacemos observaciones con una cifra decimal, esta podría variar si es otro el observador.

□

Los ceros que se añaden a la izquierda de una expresión decimal para situar la primera cifra distinta de cero no son significativos:

Ejemplo: Los números

$$0.00001845 \quad 0.0001845 \quad 0.001845$$

tienen todos cuatro cifras significativas.

□

Los ceros que se añaden al final pueden ser o no significativos:

Ejemplo: En la expresión "se manifestaron 45000 personas", los tres últimos ceros parecen ser no significativos. Sin embargo en la expresión "me costó 45000 euros exactos"sí son significativos.

□

Para evitar equívocos acerca del número de cifras significativas, se emplea la **notación científica** (en el ejemplo anterior, $4.5 \times 10^4, 4.5000 \times 10^4$).

Cualquier método del Cálculo Numérico que se emplee en la resolución de un problema, debe ser suficientemente **exacto**, es decir, los valores aproximados que se obtengan en diferentes intentos de solución deben estar cerca del valor verdadero. Cuanto mayor sea el número de cifras significativas de la aproximación, mayor será la exactitud. También deben ser **preciso**, es decir los diferentes valores obtenidos para la resolución de un problema deben estar cercanos entre sí.

Con el uso de aproximaciones para representar operaciones y cantidades matemáticas se generan **errores numéricos**. Se denomina error E a la cantidad:

$$E = \text{valor verdadero} - \text{valor aproximado}$$

A menudo se trabaja con el error absoluto ($|E|$). El error relativo es

$$e = \frac{E}{\text{valor verdadero}}.$$

Este último compara la magnitud del error cometido con la magnitud del valor que se pretende estimar y puede interpretarse en términos de %.

Las causas de los errores pueden ser:

• Truncamiento: surgen al *sustituir un procedimiento* matemático exacto por uno aproximado. Ejemplo: utilizar un polinomio de Taylor para encontrar el valor de una función en un punto.

• Redondeo: surgen al *sustituir un número* exacto por uno aproximado. Ejemplo: sustituir 1/3 por 0.33333333. Para minimizar este tipo de error se recurre a las **reglas de redondeo**:

1. En los cálculos, se conservan las cifras significativas y el resto se descartan. El último dígito que se conserva se aumenta en uno si el primer dígito descartado es mayor que 5. Si es 5 o es 5 seguido de ceros, entonces el último dígito retenido se incrementa en uno solo si este último es impar:

número	6 cifras significativas	8 cifras significativas
5.6170431500	5.61704	5.6170432
5.6170462500	5.61705	5.6170462

2. En la suma y en la resta, el redondeo se lleva a cabo de forma tal que el último dígito retenido en la respuesta corresponda al último dígito MÁS significativo de los números que se están sumando o restando. Nótese que un dígito en la columna de las centésimas es más significativo que una en la columna de las milésimas.

   $2.2 - 1.768 = 0.432 \longrightarrow 0.4$

   $4.68 \times 10^{-7} + 8.3 \times 10^{-4} - 228 \times 10^{-6} = 6.02468 \times 10^{-4} \longrightarrow 6.0 \times 10^{-4}$

3. En la multiplicación y división se conserva el número mínimo de cifras significativas que tenga los números que intervienen:

   $0.0642 \times 4.8 = 0.30816 \longrightarrow 0.31$
   $945/0.3185 = 2967.032967 \longrightarrow 297 \times 10$

4. Para combinaciones de las operaciones aritméticas, existen dos casos generales.

   (multiplicación o división) $\pm$ (multiplicación o división)

   (suma o resta) $\overset{\times}{\div}$ (suma o resta)

En ambos casos, se ejecutan primero las operaciones entre paréntesis y se redondea el resultado antes de proceder con otra operación.

• Otras causas de error: equivocación de usuario, mala formulación del modelo, incertidumbre en los datos físicos recogidos, etc.

1.2 Cálculos estables e inestables. Condicionamiento.

1.2.1 Inestabilidad

Decimos que un proceso numérico es **inestable** cuando los pequeños errores que se producen en alguna de sus etapas se agrandan en etapas posteriores y degradan seriamente la exactitud del cálculo en su conjunto.

Ejemplo: Consideremos la sucesión

$$x_0 = 1, \quad x_1 = \frac{1}{3}, \quad x_{n+1} = \frac{13}{3}x_n - \frac{4}{3}x_{n-1} \quad \forall n \geq 1$$

La sucesión anterior verifica que $x_n = (\frac{1}{3})^n$. Si calculamos el término x_{11} con el programa Maxima, obtenemos $-1.086162170293413 \times 10^{-6}$ con la definición inductiva y $2.867971990792441 \times 10^{-10}$ con la otra definición. El cálculo de los valores de la sucesión por el método inductivo es inestable, pues cualquier error que se presente en x_n se multiplica por 13/3 al calcular x_{n+1}.

1.2.2 Condicionamiento

Un problema está mal condicionado si pequeños cambios en los datos pueden dar lugar a grandes cambios en las respuestas. En ciertos problemas se puede calcular un número de condición. Si este es grande significa que el problema está mal condicionado.

Ejemplo: Dado un sistema de ecuaciones lineales en forma matricial $Ax = b$, se puede calcular el número de condición así:

$$k(A) = \|A\| \cdot \|A^{-1}\|,$$

donde $\|A\| = max_{i,j}|a_{ij}|$. Dado el sistema

$$\begin{aligned} x + 1.01y &= 1 \\ 0.99x + y &= 1 \end{aligned}$$

el número de condición de la matriz del sistema sería $k(A) = \|A\| \cdot \|A^{-1}\| = (1.01)(10100) = 10201$. Esto significa que pequeños cambios en b producirán grandes cambios en la solución. De hecho:

$$\left.\begin{array}{rcl} x + 1.01y & = & 1 \\ 0.99x + y & = & 1 \end{array}\right\} \Rightarrow [x = -100, y = 100]$$

y

$$\left.\begin{array}{rcl} x + 1.01y & = & 1.01 \\ 0.99x + y & = & 1 \end{array}\right\} \Rightarrow [x = 0, y = 1]$$

1.3 Aritmética de la computadora

1.3.1 Aritmética de punto flotante.

Las computadoras utilizan la llamada *aritmética del punto flotante*, es decir, almacenan cada número desplazando sus dígitos de modo que aparezca de la forma

$$0.\beta_1\beta_2...\beta_n \times 10^e.$$

Por ejemplo:

$$13.524 \rightarrow 0.13524 \times 10^2$$

$$-0.0442 \rightarrow -0.442 \times 10^{-1}$$

Cómo todos los números se traducen según la aritmética del punto flotante, todos empezarán con $0.\cdots$ y terminarán con 10^e. El ordenador, para almacenarlos ahorrando memoria, prescinde de lo que es común a todos los números, y así el 13.524 se almacena como $13524E2$ y el -0.0442 como $-442E-1$.

Por otra parte, la mayor parte de las computadoras trabajan con números reales en sistema binario, en contraste con el sistema decimal que normalmente se utiliza. Así, el número 9.90625 en sistema decimal, se convierte en binario en:

$$9 = 2*(2*(2*\mathbf{1}+\mathbf{0})+\mathbf{0})+\mathbf{1} = 1*2^3+0*2^2+0*2^1+1*2^0 = (1001)_2$$

$$\left.\begin{array}{rcc} 0.90625*2 = & \mathbf{1}.81250 & 1*2^{-1} \\ 0.81250*2 = & \mathbf{1}.625 & 1*2^{-2} \\ 0.625*2 = & \mathbf{1}.25 & 1*2^{-3} \\ .25*2 = & \mathbf{0}.5 & 1*2^{-4} \\ .5*2 = & \mathbf{1} & 1*2^{-5} \end{array}\right\} \Rightarrow$$

$$0.90625 = 1 * 2^{-1} + 1 * 2^{-2} + 1 * 2^{-3} + 0 * 2^{-4} + 1 * 2^{-5} = (0.11101)_2$$

Con lo que

$$\begin{aligned} 1.90625 &= 1 * 2^3 + 0 * 2^2 + 0 * 2^1 + 1 * 2^0 + 1 * 2^{-1} + \\ &\quad +1 * 2^{-2} + 1 * 2^{-3} + 0 * 2^{-4} + 1 * 2^{-5} = (1001.11101)_2 \end{aligned}$$

En general, la representación de un número en una computadora será de la forma:

$$\pm' d_1 d_2 ... d_p \times B^e$$

y se almacenará de la forma

$$\pm d_1 d_2 ... d_p E e$$

donde B es la base, $d_i \in \{0, 1, 2, ..., B-1\}$, $d_1 > 0$, p es el número máximo de bits significativos y e es el exponente de la potencia de B. El 0 es un caso especial que se almacena como $00...0E0$. La parte $d_1 d_2 ... d_p$ se llama parte significativa o mantisa. p es el número máximo de bits que puede tener esa parte significativa del número que se almacena y es, por tanto, finita. Cuanto más grande sea p, mayor será la precisión de la máquina. Como consecuencia de todo lo anterior, en cualquier máquina solo se pueden almacenar una cantidad finita de números, llamados *números máquina*. En los ordenadores de base 2, d_1 siempre es 1, con lo que se puede suprimir, aumentando la precisión. Se gana con ello un bit que se denomina *bit escondido*.

Ejemplo: Si en una computadora se tiene que $B = 2$, $p = 2$ y $-3 \leq e \leq 3$, entonces solo se podrán almacenar números de la forma

$$\pm .10_2 \times 2^e \quad \pm .11_2 \times 2^e \quad \text{con } -3 \leq e \leq 3.$$

Como

$$.10_2 = 1 \times 2^{-1} + 0 \times 2^{-2} = 1/2$$

los números $\pm .10_2 \times 2^e$ son

$$\pm 4, \pm 2, \pm 1, \pm 1/2, \pm 1/4, \pm 1/8,$$

y como

$$.11_2 = 1 \times 2^{-1} + 1 \times 2^{-2} = 1/2 + 1/4 = 3/4$$

los números $\pm .11_2 \times 2^e$ son

$$\pm 6, \pm 3, \pm 3/2, \pm 3/4, \pm 3/8, \pm 3/16.$$

En este sistema el número $2'9$ se almacenaría, según esté diseñada la computadora, bien como 2 si se hace por truncamiento, o bien como 3 si se hace por redondeo. También se almacenarían como 2 los números $2'3$ o $2'4$.

□

Los parámetros B, p y e de una máquina real pueden ser los siguientes: $B = 2$, $p = 23$ bits, $e = 8$ bits, con lo cual el número máximo que se podría almacenar sería $1'701E38$. Todo número mayor que el anterior estaría en el *desbordamiento positivo* de la máquina. El número mínimo en valor absoluto sería $1'755E-38$. Todo número positivo menor estaría en el *subdesbordamiento positivo* de la máquina. De forma similar se definen *desbordamiento negativo* y *subdesbordamiento negativo*.

1.3.2 Operaciones con computadoras

Supongamos una computadora cuyos parámetros sean $B = 10$, $p = 3$ y $-9 \leq e \leq 9$. Veamos como trabajan y como se pueden generar errores al manejar las operaciones básicas y el redondeo:

1. Para la suma, el de menor exponente se desplaza para alinear la coma decimal

$$\begin{aligned} 1.37 + 0.0269 &= 0.137 \times 10^1 + 0.269 \times 10^{-1} = \\ &= 0.137 \times 10^1 + 0.00269 \times 10^1 = 0.13969 \times 10^1 \end{aligned}$$

   que se almacenaría como 0.139×10^1 por truncamiento y 0.140×10^1 por redondeo.

2. En la resta:

$$4850 - 4820 = 0.485 \times 10^4 - 0.482 \times 10^4 = 0.003 \times 10^4 = 0.300 \times 10^2$$

   que se almacenaría como 0.300×10^2 tanto por redondeo como por truncamiento.

3. En el producto, se multiplican las cifras significativas y se suman los exponentes:

$$403000 \cdot 0.0197 = 0.403 \times 10^6 \cdot 0.197 \times 10^{-1} = 0.079391 \times 10^5$$

   que se almacena como 0.793×10^4 por truncamiento y 0.794×10^4 por redondeo.

4. En la división, se dividen las cifras significativas y se restan los exponentes:

$$0.0356/1560 = 0.356 \times 10^{-1}/0.156 \times 10^{4} = 2.28205 \times 10^{-5}$$

que se almacena como 0.228×10^{-4}.

Otra fuente de error puede ser el paso de sistema decimal a sistema binario. Así el número 1/10 en sistema binario es $(0.0001100110011001...)_2$.

1.3.3 Epsilon de la máquina

La diferencia más pequeña entre dos números que puede detectar una máquina se denomina *epsilon de la máquina*. También se puede definir como el menor número $\varepsilon > 0$ tal que

$$1.0 + \varepsilon \neq 1.0$$

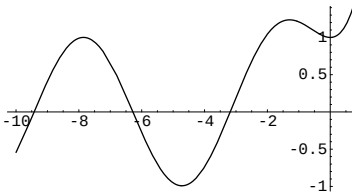
y se puede calcular programando este algoritmo:

input $s \leftarrow 1.0$
for $k = 1, 2, \ldots, 100$ do
$s \leftarrow 0.5s$
$t \leftarrow s + 1.0$
if $t \leq 1.0$ then
$s \leftarrow 2.0s$
output k-1,s
stop
end if
end

Capítulo 2

Resolución de ecuaciones no lineales

Dada una función $f : \mathbb{R} \to \mathbb{R}$, trataremos de encontrar métodos que permitan localizar valores aproximados de las soluciones de $f(x) = 0$.



$f(x) = e^x - \text{sen}(x)$

Ejemplo: *Dada la ecuación* $e^x - \text{sen}(x) = 0$*, encuentra la solución más cercana a* 0*.*

□

Para resolver ejercicios como el anterior es conveniente recordar los siguientes resultados:

Teorema [Bolzano]: *Sea* $f : \mathbb{R} \to \mathbb{R}$ *una función continua en* $[a, b]$ *tal que* $f(a)f(b) < 0$*. Entonces existe* $c \in (a, b)$ *tal que* $f(c) = 0$*.*

El teorema de Bolzano se utiliza en este contexto para localizar intervalos que contengan una solución de $f(x) = 0$ cuando la función $f(x)$ sea continua.

Ejercicio: *Dada la ecuación* $e^x - \text{sen}(x) = 0$*, localiza un intervalo donde exista una solución.*

Teorema [Rolle]: *Sea $f : \mathbb{R} \to \mathbb{R}$ una función continua en $[a, b]$, derivable en (a, b) y tal que $f(a) = f(b)$. Entonces existe $c \in (a, b)$ tal que $f'(c) = 0$.*

Del teorema de Rolle se deduce que, si $f(x)$ cumple las hipótesis, entre cada dos soluciones de la ecuación $f(x) = 0$ debe existir al menos una solución de $f'(x) = 0$. De este modo, el número máximo de soluciones de $f(x) = 0$ será el número de soluciones de $f'(x) = 0$ más uno. Por tanto, el teorema de Rolle se utilizará para determinar el número máximo de soluciones que puede tener una ecuación $f(x) = 0$.

Ejercicio: *Determina el número de soluciones de la ecuación $e^x - x = 0$.*

Teorema [Taylor]: *Sea $f : \mathbb{R} \to \mathbb{R}$ una función n veces derivable en $[a, b]$ y $n + 1$ veces derivable en (a, b). Entonces, para cada $x_0 \in [a, b]$ existe un polinomio de grado menor o igual que n que tiene un punto de contacto con $f(x)$ de grado n en x_0. Dicho polinomio es*

$$P_{n,x_0}(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{n)}(x_0)}{n!}(x - x_0)^n.$$

Además, para cada $x \in [a, b]$ existe un punto θ_x entre x y x_0 tal que

$$f(x) - P_{n,x_0}(x) = \frac{f^{n+1)}(\theta_x)}{(n + 1)!}(x - x_0)^{n+1}.$$

Teorema [de valor medio]: *Sea $f : \mathbb{R} \to \mathbb{R}$ una función continua en $[a, b]$ y derivable en (a, b). Entonces existe $c \in (a, b)$ tal que*

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

2.1 Método de la bisección

2.1.1 Descripción del método

Consiste en aplicar de forma reiterada el teorema de Bolzano en intervalos cada vez más pequeños. Supongamos que $f(a) < 0$ y $f(b) > 0$ (de forma similar se razonaría si $f(a) > 0$ y $f(b) < 0$).

Paso 0. Partimos del intervalo $[a, b]$ y calculamos $c_0 = \frac{a+b}{2}$.

Paso 1. Si $f(c_0) < 0$, definimos $a_1 = c_0$ y $b_1 = b$. Si $f(c_0) > 0$, definimos $a_1 = a$ y $b_1 = c_0$. En ambos casos, estudiamos el intervalo $[a_1, b_1]$ y calculamos $c_1 = \frac{a_1+b_1}{2}$.

Paso 2. Si $f(c_1) < 0$, definimos $a_2 = c_1$ y $b_2 = b_1$. Si $f(c_1) > 0$, definimos $a_2 = a_1$ y $b_2 = c_1$. En ambos casos, estudiamos el intervalo $[a_2, b_2]$ y calculamos $c_2 = \frac{a_2+b_2}{2}$.

Paso n. Si $f(c_{n-1}) < 0$, definimos $a_n = c_{n-1}$ y $b_n = b_{n-1}$. Si $f(c_{n-1}) > 0$, definimos $a_n = a_{n-1}$ y $b_n = c_{n-1}$. En ambos casos, estudiamos el intervalo $[a_n, b_n]$ y calculamos $c_n = \frac{a_n+b_n}{2}$.

2.1.2 Convergencia del método

Teorema: *Si* $[a, b], [a_1, b_1], \ldots, [a_n, b_n], \ldots$ *denotan los intervalos utilizados en el método de la bisección, entonces*

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \lim_{n\to\infty} c_n = r$$

y además $f(r) = 0$.

Demostración. Por la construcción de las sucesiones se tiene que

$$a \leq a_1 \leq a_2 \leq \cdots \leq b_2 \leq b_1 \leq b$$

y

$$(b_n - a_n) = \frac{1}{2}(b_{n-1} - a_{n-1}) = \cdots = \frac{1}{2^n}(b - a).$$

Por tanto, la sucesión $(a_n)_{n\in\mathbb{N}}$ converge por ser creciente y acotada superiormente y la sucesión $(b_n)_{n\in\mathbb{N}}$ converge por ser decreciente y acotada inferiormente. Veamos que sus límites son iguales

$$\lim_{n\to\infty} b_n - \lim_{n\to\infty} a_n = \lim(b_n - a_n) = \lim_{n\to\infty} \frac{1}{2^n}(b - a) = 0.$$

Dado lo anterior y puesto que $a_n \leq c_n \leq b_n$ se concluye que

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = \lim_{n\to\infty} b_n.$$

Llamemos a ese límite r. Veamos ahora que r es la solución de $f(x) = 0$ utilizando las hipótesis de continuidad de $f(x)$ y la condición $f(a_n)f(b_n) \leq 0$:

$$\begin{aligned} 0 \geq f(r)f(r) &= f(\lim_{n\to\infty} a_n)f(\lim_{n\to\infty} b_n) = \\ &\lim_{n\to\infty} f(a_n) \lim_{n\to\infty} f(b_n) = \lim_{n\to\infty} f(a_n)f(b_n) \leq 0. \end{aligned}$$

Por tanto $f(r) = 0$ que era lo que se tenía que probar.

□

2.1.3 Aproximación y error

Una aproximación a la solución de la ecuación $f(x) = 0$ en el paso n es c_n. Una cota del error cometido será:

$$|r - c_n| \leq \frac{1}{2}|b_n - a_n| = \frac{1}{2^2}|b_{n-1} - a_{n-1}| = \frac{1}{2^{n+1}}|b - a|.$$

Ejemplo: *Encuentra una aproximación de la solución de* $e^x = \text{sen}(x)$ *con un error menor que* 10^{-2}.

Aplicaremos el método de la bisección a la función $f(x) = e^x - \text{sen}(x)$. Por el teorema de Bolzano sabemos que hay una solución de $f(x) = 0$ en el intervalo $[-4, -3]$, puesto que $f(-4) < 0$ y $f(-3) > 0$. Si c_n es la aproximación y r es la solución exacta, como queremos que el error sea menor que 10^{-2}, tendremos que:

$$|r - c_n| \leq \frac{1}{2^{n+1}}|b - a| = \frac{1}{2^{n+1}}|-3-(-4)| = \frac{1}{2^{n+1}} < 10^{-2}$$

por lo que $n \geq 6$, es decir, tendremos que calcular c_6.

Los cálculos serían los siguientes:

$a_0 = -4$	$b_0 = -3$	$c_0 = -3.5$	$f(c_0) = -0.3206$
$a_1 = -3.5$	$b_1 = -3$	$c_1 = -3.25$	$f(c_1) = -0.0694$
$a_2 = -3.25$	$b_2 = -3$	$c_2 = -3.125$	$f(c_2) = 0.0554$
$a_3 = -3.25$	$b_3 = -3.125$	$c_3 = -3.1875$	$f(c_3) = -0.0046$
$a_4 = -3.1875$	$b_4 = -3.125$	$c_4 = -3.15625$	$f(c_4) = 0.0277$
$a_5 = -3.1875$	$b_5 = -3.15625$	$c_5 = -3.175$	$f(c_5) = 0.0084$
$a_6 = -3.1875$	$b_6 = -3.175$	$c_6 = -3.18125$	

La solución aproximada propuesta sería -3.18125. □

2.1.4 Variaciones del método: Regula Falsi

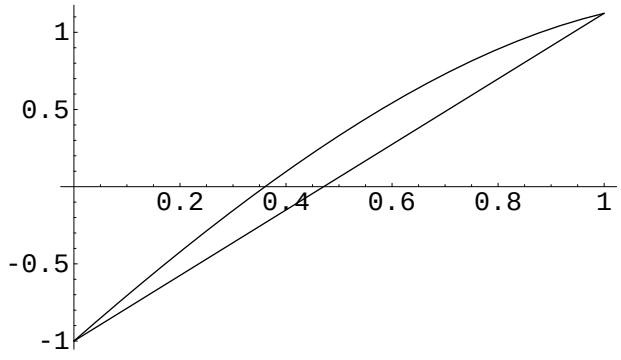
El método de la Regula-falsi difiere respecto de la bisección en el cálculo del punto c_n. Con este método se calcula así:

$$c_n = b_n - f(b_n)\frac{b_n - a_n}{f(b_n) - f(a_n)}.$$

El paso siguiente es elegir el intervalo formado por los puntos a_n y c_n o bien el formado por c_n y b_n asegurando que la función en los extremos sea de signo contrario.

Ejemplo: *Aplicar el método de la regula falsi para encontrar una solución de* $f(x) = 3x + \text{sen}(x) - e^x = 0$ *a partir del intervalo* $[0, 1]$.

$a_0 = 0$	$b_0 = 1$	$c_0 = 0.470990$	$f(c_0) = 0.265160$
$a_1 = 0$	$b_1 = 0.470990$	$c_1 = 0.372277$	$f(c_1) = 0.029533$
$a_2 = 0$	$b_2 = 0.372277$	$c_2 = 0.361598$	$f(c_2) = 2.94 * 10^-3$
$a_3 = 0$	$b_3 = 0.361598$	$c_3 = 0.360538$	$f(c_3) = 2.94 * 10^-4$
$a_4 = 0$	$b_4 = 0.360538$	$c_4 = 0.360433$	$f(c_4) = 2.94 * 10^-5$



regula falsi para $f(x) = 3x + \text{sen}(x) - e^x$

□

2.2 Método de Newton-Raphson

2.2.1 Descripción del método

Consiste en utilizar el polinomio de Taylor de grado 1 como aproximación $f(x)$.
Paso 1. Partimos de un punto inicial x_0. Calculamos el polinomio de Taylor de $f(x)$ de grado 1 en x_0

$$P_{1,x_0}(x) = f(x_0) + f'(x_0)(x - x_0).$$

Paso 2. Utilizamos $P_{1,x_0}(x)$ como aproximación de $f(x)$ y, en vez de resolver $f(x) = 0$, resolvemos $P_{1,x_0} = 0$, es decir

$$f(x_0) + f'(x_0)(x - x_0) = 0 \Rightarrow x = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Paso 3. Definimos

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

y repetimos los pasos 1, 2 y 3 sustituyendo el punto x_{n-1} por el punto x_n, es decir construimos la sucesión

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

Cuando la sucesión $x_0, x_1, x_2, \ldots$ es convergente y la función $f(x)$ tiene cierta regularidad, su límite r es la solución puesto que

$$r = \lim_{n\to} x_n = \lim_{n\to} \left[x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \right] = r - \frac{f(r)}{f'(r)} \Rightarrow \frac{f(r)}{f'(r)} = 0 \Rightarrow f(r) = 0.$$

Ejemplo *Aplica el método de Newton-Raphson para encontrar una aproximación de* $\sqrt{2}$.

Puesto que queremos una aproximación de $\sqrt{2}$, bastará encontrar una aproximación de $x^2 - 2 = 0$. Utilizamos el método de Newton-Raphson aplicado a la función $f(x) = x^2 - 2$, es decir, construimos la sucesión

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} = x_{n-1} - \frac{x_{n-1}^2 - 2}{2x_{n-1}}.$$

Por el teorema de Bolzano, en el intervalo $[1, 2]$ hay una solución. Comenzamos por el punto $x_0 = 1.5$:

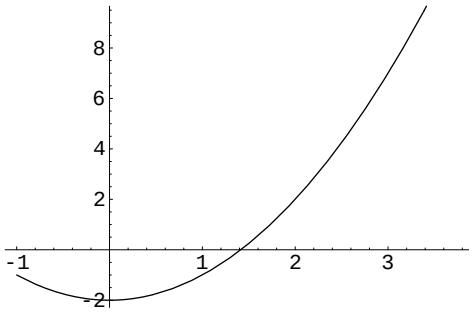
$$x_0 = 1.5, \quad x_1 = 1.4167, \quad x_2 = 1.4142, \quad x_3 = 1.4142, \quad \ldots$$

La aproximación sería 1.4142.

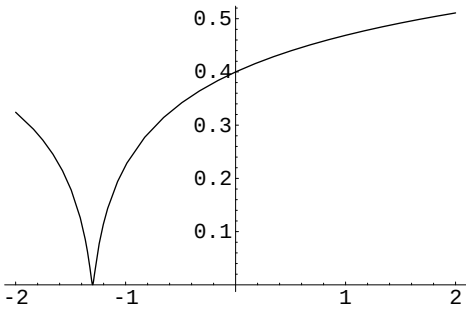
□

2.2.2 Convergencia del método

Teorema: *Si en un cierto intervalo se verifica que $f \in C^2(\mathbb{R})$, estrictamente creciente, $f''(x) \geq 0$ y f tiene un cero, entonces el cero es único y la iteración de Newton-Raphson convergerá a partir de cualquier punto inicial.*



Newton Raphson converge



Newton Raphson no converge

Demostración. Sea $x_0, x_1, \ldots, x_n$ la sucesión construida por el método de Newton-Raphson, r la solución de la ecuación $f(x) = 0$ y $e_n = x_n - r$ el error en el paso n. Por el teorema de Taylor sabemos que:

$$0 = f(r) = f(x_n) + f'(x_n)(-e_n) + \frac{f''(\theta_n)}{2}(-e_n)^2.$$

Por tanto

$$e_n f'(x_n) - f(x_n) = \frac{f''(\theta_n)}{2} e_n^2.$$

De modo que

$$e_{n+1} = x_{n+1} - r = x_n - \frac{f(x_n)}{f'(x_n)} - r = e_n - \frac{f(x_n)}{f'(x_n)} = \frac{e_n f'(x_n) - f(x_n)}{f'(x_n)} = \frac{f''(\theta_n) e_n^2}{2f'(x_n)} \geq 0$$

Por lo tanto

$$x_{n+1} \geq r \Rightarrow f(x_n) \geq f(r) = 0 \Rightarrow e_{n+1} = e_n - \frac{f(x_n)}{f'(x_n)} \leq e_n.$$

Puesto que $x_{n+1} \geq r$ y el error va decreciendo, se concluye que la sucesión $(x_n)_{n\in\mathbb{N}}$ es decreciente y acotada inferiormente por r, por lo tanto tiene límite. Así mismo, la sucesión $(e_n)_{n\in\mathbb{N}}$ es decreciente y acotada inferiormente por 0, por lo tanto también tiene límite. Se deduce que

$$\lim_{n\to\infty} e_{n+1} = \lim_{n\to\infty} \left(e_n - \frac{f(x_n)}{f'(x_n)}\right) = \lim_{n\to\infty} e_n - \frac{f(\lim_{n\to\infty} x_n)}{f'(\lim_{n\to\infty} x_n)} \Rightarrow f(\lim_{n\to\infty} x_n) = 0 \qquad (2.1)$$

y puesto que $f(x)$ es estrictamente creciente se obtiene que $\lim_{n\to\infty} x_n = r$ (en caso contrario $f(\lim_{n\to\infty} x_n) > f(r) = 0$), es decir, la sucesión de Newton-Raphson converge a la solución.

□

2.2.3 Aproximación y error

La aproximación a la solución en el paso n del método de Newton-Raphson es x_n. Por el razonamiento anterior sabemos que

$$e_{n+1} = \frac{f''(x_n)}{2f'(x_n)} e_n^2.$$

Si encontramos una constante C tal que

$$C \geq \frac{\max_{x\in I} |f''(x)|}{2 \min_{x\in I} |f'(x)|}$$

siendo I un intervalo que contenga a la sucesión $(x_n)_{n\in\mathbb{N}}$ y a la solución r, tendremos que

$$|e_{n+1}| \leq C e_n^2$$

lo que significa que la convergencia de la sucesión de errores es cuadrática. En estas condiciones

$$C|e_{n+1}| \leq C^2 e_n^2 \leq C^2 (Ce_{n-1}^2)^2 \leq \cdots \leq (Ce_0)^{2^{n+1}}$$

es decir

$$C|e_n| \leq (Ce_0)^{2^n} \Rightarrow |e_n| \leq \frac{1}{C}(Ce_0)^{2^n}$$

que es una cota del error que se puede obtener previamente a aplicar el método. Si $|Ce_0| < 1$, o lo que es lo mismo $|e_0| < \frac{1}{C}$, siendo $e_0 = x_0 - r$, tendremos que

$$|e_n| \leq \frac{1}{C}(Ce_0)^{2^n} \to 0$$

con lo que tenemos otro criterio de convergencia del método.

Ejemplo: *Calcula el número de pasos necesarios para encontrar la raíz cuadrada de* 3 *con un error menor que* 10^{-6}

1. *Con el método de Newton-Raphson*

2. *Con el método de la bisección.*

Puesto que la raíz cuadrada de 3 es solución de $x^2 - 3 = 0$, aplicaremos los métodos a la función $f(x) = x^2 - 3$. Puesto que $f(x)$ es continua en $[1, 2]$ y tiene distinto signo en los extremos del intervalo, por el teorema de Bolzano se sabe que en $[1, 2]$ hay una solución de $f(x) = 0$.

Apartado 1). Aplicamos el primer criterio para saber si el método de Newton-Raphson convergerá en el intervalo $[1, 2]$. $f(x)$ es un polinomio y por tanto es de clase C^2. Como $f'(x) = 2x > 0$ en $[1, 2]$, la función es estrictamente creciente en $[1, 2]$. Además $f''(x) = 2 > 0$ en $[1, 2]$. Por todo ello, podemos asegurar que el método de Newton-Raphson convergerá a partir de cualquier punto de $[1, 2]$.

Veamos el número de iteraciones necesarias para que el error sea menor que 10^{-6}. Sabemos que

$$|e_n| \leq \frac{1}{C}(Ce_0)^{2^n}$$

siendo

$$C \geq \frac{\max_{x\in[1,2]|f''(x)|}}{2\min_{x\in[1,2]}|f'(x)|} = \frac{2}{2\min_{x\in[1,2]}|2x|} = \frac{2}{2\cdot 2} = \frac{1}{2}.$$

Por tanto

$$|e_n| \leq \frac{1}{1/2}\left(\frac{1}{2}e_0\right)^{2^n} = 2\left(\frac{1}{2}e_0\right)^{2^n}$$

Si $n = 4$, entonces $|e_4| \leq 3.051 * 10^{-5}$
Si $n = 5$, entonces $|e_5| \leq 4.66 * 10^{-10}$. Luego basta con 5 iteraciones.
Aplicamos el método empezando en el punto $x_0 = 1$

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} = x_{n-1} - \frac{x_{n-1}^2 - 3}{2x_{n-1}} = \frac{x_{n-1}^2 + 3}{2x_{n-1}}.$$

$x_0 = 1$	$x_1 = 2$	$x_2 = 1.75$	$x_3 = 1.73414$	$x_4 = 1.73205$	$x_5 = 1.73205$

Apartado 2). Sabemos que

$$|e_n| \leq \frac{1}{2^{n+1}}(b-a) = \frac{1}{2^{n+1}} < 10^{-6}$$

y si $n = 19$, entonces $|e_{19}| \leq 9.54 * 10^{-7}$, es decir, necesitamos 19 iteraciones. □

Ejemplo: *Dada la ecuación* $e^{-x} - x = 0$,

1. *Aplica el método de Newton-Raphson con* 4 *iteraciones para encontrar la solución.*

2. *Calcula una cota del error cometido.*

Sea $f(x) = e^{-x} - x$. Aplicando el teorema de Bolzano al intervalo $[0, 1]$ se puede asegurar que en dicho intervalo hay una solución de la ecuación $f(x) = 0$. Como $f'(x) = e^{-x} - 1 < 0$ en $[0, 1]$ no podemos aplicar el criterio para asegurar que el método funcione empezando en cualquier punto del intervalo $[0, 1]$. No obstante, lo aplicamos empezando en el punto $x_0 = 0$

$$x_{n+1} = x_n - \frac{e^{-x_n} - x_n}{-e^{-x_n} - 1}$$

$x_0 = 0$	$x_1 = 0.5$	$x_2 = 0.566311$	$x_3 = 0.567143$	$x_4 = 0.567143$

Calculemos una cota del error

$$C|e_4| \leq (Ce_0)^{2^n}$$

siendo

$$C \geq \frac{\max_{x\in I}|f''(x)|}{2\min_{x\in I}|f'(x)|} = \frac{\max_{x\in[0,1]}|e^{-x}|}{2\min_{x\in[1,2]}|-e^{-x}-1|} = \frac{1}{2(1/e+1)} = \frac{e}{2(e+1)} \approx 0.365529.$$

Por tanto

$$|e_4| \leq C^{2^n-1}e_0^{2^n} \leq 0.37^{15}1^{16} \approx 0.33 * 10^{-7}$$

□

2.2.4 Variaciones del método: Método de la secante

Teniendo en cuenta que$f'(x_n) = \lim_{h\to 0}\frac{f(x_n+h)-f(x_n)}{h}$ este método se diferencia del de Newton-Raphson en que se sustituye $f'(x_n)$ por el valor aproximado $\frac{f(x_{n-1})-f(x_{n-2})}{x_{n-1}-x_{n-2}}$ de modo que la sucesión que queda es

$$x_n = x_{n-1} - f(x_{n-1})\frac{x_{n-1}-x_{n-2}}{f(x_{n-1})-f(x_{n-2})}.$$

Por tanto hay que utilizar dos puntos iniciales en lugar de uno. Es un método más rápido que el de la bisección pero más lento que el de Newton-Raphson.

Ejemplo: *Calcula una aproximación de* $\sqrt{2}$*, con* 4 *iteraciones del método de la secante.*

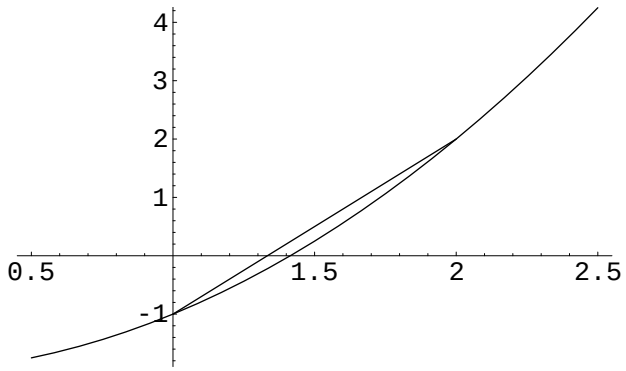
Aplicamos el método a la solución de la ecuación $f(x) = x^2 - 2 = 0$ en el intervalo $[1, 2]$.

$$x_n = x_{n-1} - (x_{n-1}^2-2)\frac{x_{n-1}-x_{n-2}}{x_{n-1}^2-2-[x_{n-2}^2-2]} = x_{n-1} - (x_{n-1}^2-2)\frac{x_{n-1}-x_{n-2}}{x_{n-1}^2-x_{n-2}^2}$$

y tomando tomando $x_0 = 1$ y $x_1 = 2$ nos queda

$x_0 = 1$	$x_1 = 2$	$x_2 = 1.3333$	$x_3 = 1.4$	$x_4 = 1.4146$

□



Método de la secante para $f(x) = x^2 - 2$

2.3 Método iterativo de punto fijo

2.3.1 Descripción del método

Ciertos problemas de aproximación se pueden resolver encontrando una solución de $F(x) = x$. Una solución r de $F(x) = x$ se denomina un *punto fijo* de la función $F(x)$, debido a que $F(r) = r$. A menudo, para resolver la ecuación $F(x) = x$ se construye la sucesión

$$x_n = F(x_{n-1}).$$

Cuando la sucesión x_n es convergente y $F(x)$ es continua, si $r = \lim_{n\to\infty} x_n$, entonces

$$F(r) = F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} x_{n+1} = r,$$

es decir, el límite de x_n es un punto fijo de $F(x)$. La búsqueda de soluciones de $F(x) = x$ empleando sucesiones de la forma $x_n = F(x_{n-1})$, se denomina método iterativo de punto fijo.

Ejemplo: *Encuentra una solución real de* $x^3 - x - 1 = 0$ *usando cinco pasos de un método iterativo de punto fijo.*

Como pretendemos resolver la ecuación $f(x) = x^3 - x - 1 = 0$, por el teorema de Bolzano sabemos que hay una solución en el intervalo $[1, 2]$.

Resolver $f(x) = x^3 - x - 1 = 0$ es equivalente a resolver una de estas ecuaciones

$$x^3 - 1 = x \qquad \text{o bien} \qquad \sqrt[3]{1 + x} = x,$$

con lo que podemos aplicar el método iterativo de punto fijo para resolver la ecuación $F(x) = x$ siendo $F(x) = x^3 - 1$ o bien $F(x) = \sqrt[3]{1+x}$. Construyamos la sucesión $x_n = F(x_{n-1})$ para ambas funciones, comenzando por un punto del intervalo $[1, 2]$, por ejemplo por $x_0 = 1$ resulta lo siguiente:

$x_n = F(x_{n-1})$	x_0	x_1	x_2	x_3	x_4	x_5
$F(x) = x^3 - 1$	1	0	-1	-2	-9	-730
$F(x) = \sqrt[3]{1+x}$	1	1.2600	1.3123	1.3224	1.3243	1.3246

Es claro que para la función $F(x) = x^3 - 1$, la sucesión no resulta convergente. Sin embargo, para la función $F(x) = \sqrt[3]{1+x}$ la sucesión que sale sí es convergente a un número r. Dicho número, puesto que $F(x)$ es continua, verificará que $F(r) = r$, es decir, $f(r) = 0$. La mejor aproximación de r obtenida sería 1.3246.

□

2.3.2 Convergencia del método

Definición: Sea $F : \mathbb{R} \to \mathbb{R}$ una función. Se dice que F es contractiva en un conjunto $C \subset \mathbb{R}$ si existe $0 \leq \lambda < 1$ tal que

$$|F(x) - F(y)| \leq \lambda |x - y| \quad \forall x, y \in C.$$

Ejemplo: *Comprueba que la función* $F(x) = \sqrt[3]{1+x}$ *es contractiva en* $[0, 1]$.

Por el teorema del valor medio

$$|F(y) - F(x)| = |F'(\theta)(y - x)| = \left| \frac{1}{3\sqrt[3]{(1+\theta)^2}}(y - x) \right| \leq$$
$$\leq \left| \frac{1}{3\sqrt[3]{1}}(y - x) \right| = \frac{1}{3}|y - x|$$

para todo $x, y \in [0, 1]$. Por tanto la función es contractiva con $\lambda = 1/3$.

□

Teorema: *Si* F *es una función contractiva que va de un cerrado* $C \subset \mathbb{R}$ *en* C *(*$F(C) \subset C$*), entonces* F *tiene un único punto fijo* $r \in C$*. Además* r *es el límite de cualquier sucesión que se obtenga a partir de la expresión* $x_{n+1} = F(x_n)$ *siendo* $x_0 \in C$*.*

Demostración. Sea $x_0, x_1, x_2, \ldots, x_n, \ldots$ la sucesión generada mediante la expresión $x_{n+1} = F(x_n)$. Veamos que $(x_n)_{n\in N}$ converge. Por ser contractiva se tiene que

$$|x_n - x_{n-1}| = |F(x_{n-1}) - F(x_{n-2})| \leq \lambda|x_{n-1} - x_{n-2}| \leq \\ \lambda^2|x_{n-2} - x_{n-3}| \leq \cdots \leq \lambda^{n-1}|x_1 - x_0|.$$

Por otra parte

$$x_n = x_0 + x_1 - x_0 + x_2 - x_1 + \cdots + x_n - x_{n-1} = x_0 + \sum_{k=1}^{n}(x_k - x_{k-1})$$

por lo que

$$\lim_{n\to\infty} x_n = x_0 + \lim_{n\to\infty}\sum_{k=1}^{n}(x_k - x_{k-1}) = x_0 + \sum_{k=1}^{\infty}(x_k - x_{k-1})$$

La serie $\sum_{k=1}^{\infty}(x_k - x_{k-1})$ es convergente porque

$$\left|\sum_{k=1}^{\infty}(x_k - x_{k-1})\right| \leq \sum_{k=1}^{\infty}|x_k - x_{k-1}| \leq \sum_{k=1}^{\infty}\lambda^{k-1}(x_1 - x_0) = |x_1 - x_0|\frac{1}{1-\lambda}$$

Como la serie es convergente, la sucesión $(x_n)_{n\in\mathbb{N}}$ también lo es. Sea $r = \lim_{n\to\infty} x_n$. Por la construcción de la sucesión, se tiene que $F(r) = r$, es decir, hay un punto fijo. $r \in C$ por ser C un conjunto cerrado y $x_n \in C$. Falta ver que es único. Supongamos que hubiese dos puntos fijos r_1 y r_2. Entonces

$$|r_1 - r_2| = |F(r_1) - F(r_2)| \leq \lambda|r_1 - r_2| < |r_1 - r_2|$$

llegando a contradicción. □

2.3.3 Aproximación y error

Si r es solución de $f(x) = 0$ obtenida por iteración de una función $F(x)$, entonces si $f(x)$ es continua y derivable, existe $\theta_n \in \mathbb{R}$ entre r y x_n tal que

$$\frac{f(r) - f(x_n)}{r - x_n} = f'(\theta_n).$$

Suponiendo que $x_n \leq r$ y puesto que $f(r) = 0$ y $-e_n = r - x_n$, se deduce que

$$\frac{f(x_n)}{e_n} = f'(\theta_n) \Rightarrow e_n = \frac{f(x_n)}{f'(\theta_n)} \Rightarrow |e_n| \leq \frac{|f(x_n)|}{\min_{\theta \in [x_n, r]} |f'(\theta)|}.$$

La fórmula anterior nos proporciona un método para encontrar una cota de error a posteriori y se puede aplicar a cualquier método iterativo de punto fijo.

Si $F(x)$ es contractiva de constante λ entonces

$$\begin{aligned} |x_{n+p} - x_n| &= |x_{n+p} - x_{n+p-1}| + |x_{n+p-1} - x_{n+p-2}| + \cdots + |x_{n+1} - x_n| \leq \\ &\leq \lambda^{n+p-1}|x_1 - x_0| + \lambda^{n+p-2}|x_1 - x_0| + \cdots + \lambda^n|x_1 - x_0| = \\ &= (\lambda^{n+p-1} + \lambda^{n+p-2} + \cdots + \lambda^n)|x_1 - x_0| = \\ &= \lambda^n(\lambda^{p-1} + \lambda^{p-2} + \cdots + 1)|x_1 - x_0| \end{aligned}$$

de modo que cuando p tiende a infinito

$$|e_n| = |r - x_n| \leq \lambda^n \sum_{k=0}^{\infty} \lambda^k = \lambda^n \left(\frac{1}{1-\lambda}\right) |x_1 - x_0|$$

que es una fórmula que permite calcular una cota de error a priori.

Ejemplo: *Encuentra una solución de la ecuación* $4 + \frac{1}{3}\sin(2x) - x = 0$ *con un método iterativo de punto fijo estimando una cota del error a priori y otra a posteriori.*

Sea $f(x) = 4 + \frac{1}{3}\sin(2x) - x$ y $F(x) = 4 + \frac{1}{3}\sin(2x)$. Aplicando el teorema de Bolzano a la función $f(x)$ en el intervalo $[-3.5, 4.5]$ se deduce que hay una solución de $f(x) = 0$ en dicho intervalo. Veamos si $F(x)$ es contractiva aplicando el teorema del valor medio

$$|F(y) - F(x)| = |F'(\theta)(y - x)| = |\frac{2}{3}\cos(2x)(y - x)| \leq \frac{2}{3}|y - x|.$$

Se deduce que $F(x)$ es contractiva con constante $\lambda = \frac{2}{3}$.

Por otra parte, puesto que $-3.5 \leq 4 + \frac{1}{3}\sin(2x) \leq 4.5$, se deduce que $F([-3.5, 4.5]) \subset [-3.5, 4.5]$.

Por tanto, la sucesión $x_{n+1} = F(x_n)$ convergerá a un punto fijo a partir de cualquier valor inicial x_0. Calculamos cinco iteraciones empezando en $x_0 = 0$:

x_0	x_1	x_2	x_3	x_4	x_5
0	4	4.329786082	4.230895147	4.273633799	4.256383406

Una cota del error para x_5 a priori sería

$$|e_5| \leq \left(\frac{2}{3}\right)^5 \left(\frac{1}{1-\frac{2}{3}}\right) |4-0| \approx 1.5802$$

que no es una cota muy buena.

Una cota del error para x_5 a posteriori sería

$$|e_5| \leq \frac{|f(x_5)|}{\min_{\theta\in[-3.5,4.5]} |\frac{2}{3}\cos(2\theta)-1|} \leq \frac{0.00719542}{1/3} = 0.00239847.$$

□

2.4 Raíces de polinomios

No existe ninguna fórmula algebraica para la resolución de raíces de polinomios de grado mayor que 4, por tanto, hay que aplicar métodos numéricos para el cálculo de dichas raíces.

Definición: Una solución r de una ecuación $f(x) = 0$ se dice que tiene multiplicidad n si

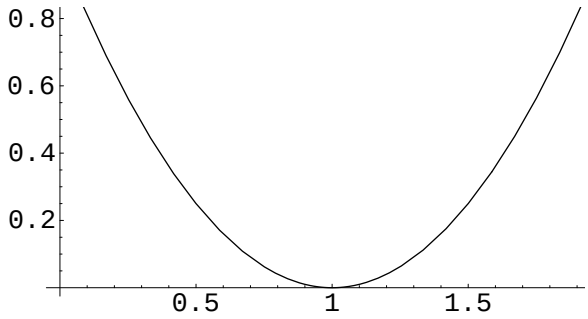
$$f(r) = f'(r) = \cdots = f^{n-1)}(r) = 0 \quad f^{n)}(r) \neq 0.$$

Ejemplo: $r = 1$ es una solución de multiplicidad 2 para la ecuación

$$x^2 - 2x + 1 = 0.$$

□

Para la localización de raíces de polinomios con multiplicidad par (como en el ejemplo anterior) no podemos emplear el teorema de Bolzano. Sin embargo podemos intentar calcular las raíces de otro polinomio con las mismas soluciones que el nuestro pero todas con multiplicidad 1.



$$f(x) = x^2 - 2x + 1$$

Teorema [Teorema fundamental del álgebra]: *Un polinomio de grado n con coeficientes reales tiene exactamente n ceros o raíces entre los números complejos* $\mathbb{C}$*, contando cada cero tantas veces como indique su multiplicidad.*

De modo que dado un polinomio $P_n(x)$, podemos descomponerlo así:

$$P_n(x) = a_n(x - r_k)^{m_k}(x - r_{k-1})^{m_{k-1}} \dots (x - r_1)^{m_1}$$

siendo m_i la multiplicidad de la raíz i-ésima. Si derivamos, obtenemos que

$$P_n'(x) = a_n(x - r_k)^{m_k - 1}(x - r_{k-1})^{m_{k-1} - 1} \dots (x - r_1)^{m_1 - 1} M(x)$$

siendo $M(x)$ un polinomio tal que $M(r_i) \neq 0$. El máximo común divisor de $P_n(x)$ y de $P_n'(x)$ es

$$D(x) = \text{M.C.D}(P_n(x), P_n'(x)) = a_n(x - r_k)^{m_k - 1}(x - r_{k-1})^{m_{k-1} - 1} \dots (x - r_1)^{m_1 - 1}$$

por lo que el polinomio

$$Q(x) = \frac{P(x)}{\text{M.C.D}(P_n(x), P_n'(x))} = a_n(x - r_k)(x - r_{k-1}) \dots (x - r_1)$$

tiene las mismas raíces que $P(x)$ pero todas con multiplicidad 1. Bastaría aplicar los métodos a $Q(x)$ para localizar las raíces de $P_n(x)$. El problema es el cálculo de $Q(x)$ y de $\text{M.C.D}(P_n(x), P_n'(x))$.

2.4.1 Separación de raíces. Sucesión de Sturm.

La sucesión de Sturm nos permite calcular el número de raíces de un polinomio en un intervalo $[a, b]$. Sea $P_n(x)$ un polinomio de grado n y $P'_n(x)$ su derivada. Para construir la sucesión de Sturm aplicamos la división de polinomios repetidas veces a $P_n(x)$, $P'_n(x)$ y a los restos obtenidos cambiados de signo:

$$\begin{aligned} P_n(x) &= c_1(x)P'_n(x) + r_1(x) \\ P'_n(x) &= c_2(x)(-r_1(x)) + r_2(x) \\ -r_1(x) &= c_3(x)(-r_2(x)) + r_3(x) \\ &\dots \\ -r_{k-2}(x) &= c_k(x)(-r_{k-1}(x)) + r_k(x) \end{aligned}$$

Puede ocurrir lo siguiente:

1. Hay un resto $r_k(x) = 0$. En este caso $r_{k-1}(x)$ es el máximo común divisor de $P_n(x)$ y de $P'_n(x)$. Construimos entonces la sucesión de Sturm correspondiente al polinomio

$$Q(x) = \frac{P_n(x)}{r_{k-1}(x)}$$

que es un polinomio que tiene las mismas raíces que $P_n(x)$ pero todas simples.

2. Llegamos a un resto $r_k(x)$ constante distinto de 0. En este caso la sucesión de Sturm es

$$P_n(x), P_{n-1}(x), -r_1(x), \dots, -r_k(x).$$

La diferencia entre el número de cambios de signos de las sucesiones siguientes

$$P_n(a), P_{n-1}(a), -r_1(a), \dots, -r_k(a) \qquad P_n(b), P_{n-1}(b), -r_1(b), \dots, -r_k(b)$$

es el número de raíces de $P_n(x)$ en el intervalo $[a, b]$.

Ejemplo: *Localiza en intervalos disjuntos las raíces del* $P(x) = 36x^4 - 12x^3 - 11x^2 + 2x + 1$.

La derivada de $P(x)$ es $P'(x) = 144x^3 - 36x^2 - 22x + 2$. Podemos multiplicar los polinomios implicados por numeros positivos para facilitar los cálculos sin que ello afecte al objetivo perseguido. Comenzamos los cálculos:

1. Multiplicamos $P_n(x)$ por 4 y efectuamos la primera división:

$$144x^4 - 48x^3 - 44x + 8x + 4 = (x-1)(144x^3 - 36x^2 - 22x + 2) + (-300x^2 + 50x + 50).$$

2. El tercer miembro de la sucesión de Sturm es $300x^2 - 50x - 50$. Efectuamos la siguiente división:

$$144x^3 - 36x^2 - 22x + 2 = 2/50(12x - 1)(300x^2 - 50x - 50) + 0$$

3. Como hay un resto que es 0 el máximo común divisor de $P(x)$ y $P'(x)$ es $300x^2 - 50x - 50$ o bien dividiendo por 50 $6x^2 - x - 1$. calculamos el polinomio

$$Q(x) = \frac{P(x)}{6x^2 - x - 1} = 6x^2 - x - 1.$$

4. Construimos la sucesión de Sturm para $Q(x)$.

$$6x^2 - x - 1 = (-\frac{1}{24} + \frac{x}{2})(12x - 1) + \frac{-25}{24}$$

Con la sucesión es

$$6x^2 - x - 1, 12x - 1, \frac{25}{24}$$

Estudiamos algunos cambios de signo:

	$-\infty$	-2	0	2	∞
$6x^2 - x - 1$	+	+	-	+	+
$12x - 1$	-	-	-	+	+
$\frac{25}{24}$	+	+	+	+	+
cambios de signo	2	2	1	0	0

En el intervalo $(\infty, -2)$ y en $(0, 2)$ no hay ninguna raíz porque en los extremos de los intervalos la sucesión de Sturm tiene los mismos cambios de signo.

En el intervalo $(-2, 0)$ hay una raíz y hay otra en el intervalo $(0, 2)$ porque la diferencia de cambios de signo en los extremos en ambos casos es 1.

□

Ejemplo: *Localiza en intervalos disjuntos las raíces del* $P(x) = x^4 + 2x^3 - 3x^2 - 4x - 1$.

La derivada es $P'(x) = 4x^3 + 6x^2 - 6x - 4 = 2(2x^3 + 3x^2 - 3x - 2)$. Para facilitar el cálculo multiplicamos $P(x)$ por 2 y calculamos es tercer miembro de la sucesión de Sturm:

$$2x^4 + 4x^3 - 6x^2 - 8x - 1 = (x + \frac{1}{2})(2x^3 + 3x^2 - 3x - 2) + (-\frac{9}{2}x^2 - \frac{9}{2}x - 1).$$

Podemos multiplicar por 2 el resto con lo que $-r_1(x) = 9x^2 - 9x + 2$. Para facilitar los cálculos multiplicamos $P'(x)$ por 9 y calculamos el siguiente miembro de la sucesión:

$$18x^3 + 27x^2 - 27x - 18 = (2x + 1)(9x^2 + 9x + 2) + (-40x - 20).$$

Dividiendo por 20 podemos suponer que $-r_2(x) = 2x + 1$. Multiplicamos $-r_1(x)$ por 2 y calculamos $-r_3(x)$:

$$18x^2 + 18x + 4 = (9x + \frac{9}{2})(2x + 1) + \frac{-1}{2}.$$

Con lo que la sucesión queda así:

$$x^4 + 2x^3 - 3x^2 - 4x - 1, \quad 2x^3 + 3x^2 - 3x - 2, \quad 9x^2 + 9x + 2, \quad 2x + 1, \quad \frac{1}{2}.$$

Separamos las raíces en intervalos disjuntos:

	$-\infty$	-3	-2	-1	-1/2	0	1	2	∞
$x^4 + 2x^3 - 3x^2 - 4x - 1$	+	+	-	-	-	-	-	+	+
$2x^3 + 3x^2 - 3x - 2$	-	-	0	+	-	-	0	+	+
$9x^2 + 9x + 2$	+	+	+	+	+	+	+	+	+
$2x + 1$	-	-	-	-	0	+	+	+	+
$\frac{1}{2}$	+	+	+	+	+	+	+	+	+
cambios de signo	4	4	3	3	2	1	1	0	0

En el intervalo $(\infty, -3)$, $(-2, -1)$, $(0, 1)$ y en $(2, \infty)$ no hay ninguna raíz porque en los extremos de los intervalos la sucesión de Sturm tiene los mismos cambios de signo.

En cada uno de los intervalos $(-3, -2)$, $(-1, -1/2)$, $-1/2, 0$ y $(1, 2)$ hay una raíz porque la diferencia de cambios de signo en los extremos en ambos casos es 1.

□

2.4.2 Acotación de raíces

Proposición: *Sea* $P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$. *Si* r *es una raíz de* $P(x)$, *entonces*

$$|r| < 1 + \frac{A}{|a_n|}$$

siendo $A = \max_{0\leq i\leq n-1} |a_i|$.

Demostración. Supongamos que $|r| > 1$.

$$\begin{aligned} 0 = |P(r)| = |a_n r^n + a_{n-1}r^{n-1} + \cdots + a_0| \geq \\ \geq |a_n r^n| - |a_{n-1}r^{n-1} + \cdots + a_0| \\ \geq |a_n r^n| - \big[|a_{n-1}||r^{n-1}| + \cdots + |a_0|\big] \geq \\ \geq |a_n r^n| - A\big[|r|^{n-1} + \cdots + |r| + 1\big] = \\ |a_n r^n| - A\frac{|r|^n - 1}{|r| - 1} > |a_n||r|^n - A\frac{|r|^n}{|r| - 1} \\ > |r|^n\big(|a^n| - \frac{A}{|r| - 1}\big) \end{aligned}$$

es decir,

$$0 > |a^n| - \frac{A}{|r| - 1} \Rightarrow |a^n| < \frac{A}{|r| - 1} \Rightarrow |r| - 1 < \frac{A}{|a_n|}$$

con lo que concluimos que

$$|r| < 1 + \frac{A}{|a_n|}.$$

Si $|r| \leq 1$, es trivial que $|r| < 1 + \frac{A}{|a_n|}$.

□

Ejemplo: *Acota las raíces del polinomio* $P(x) = x^4 + 2x^3 - 3x^2 - 4x - 1$.

Com $A = \max_{1\leq i\leq n-1} |a_i| = 4$, usando la proposición anterior obtenemos que cualquier raíz r del polinomio verifica lo siguiente

$$|r| < 1 + \frac{A}{|a_n|} = 1 + \frac{4}{1} = 5 \Rightarrow -5 < r < 5.$$

□

2.4.3 Raíces de polinomios con el algoritmo de Horner

Si usamos el método de Newton-Raphson para encontrar las raíces de un polinomio $P(x)$ hemos de calcular $P(x_n)$ y $P'(x_n)$. El algoritmo de Horner hace estos últimos cálculos de forma sencilla para polinomios.

Sea $x_0 \in \mathbb{R}$ y $P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Sea

$$P(x) = Q(x)(x - x_0) + R$$

(R será constante) el resultado de dividir $P(x)$ entre $(x - x_0)$. Supongamos que

$$Q(x) = b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \cdots + b_1 x + b_0.$$

Entonces $P(x_0) = R$ y $P(x) = Q(x)(x - x_0) + P(x_0)$, con lo que

$$P(x) - P(x_0) = Q(x)(x - x_0)$$

es decir

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 - P(x_0) = \\ = (b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \cdots + b_1 x + b_0)(x - x_0)$$

Desarrollando e igualando los coeficientes de ambos términos se deduce que

$$b_{n-1} = a_n, \quad b_{n-2} = a_{n-1} + b_{n-1}x_0, \quad \ldots, \quad b_0 = a_1 + b_1 x_0 \quad P(x_0) = a_0 + b_0 x_0.$$

La obtención de los coeficientes b_i se pueden representar gráficamente así

	a_n	a_{n-1}	a_{n-2}	$\ldots$	a_1	a_0
x_0		$b_{n-1}x_0$	$b_{n-2}x_0$	$\ldots$	$b_1 x_0$	$b_0 x_0$
	b_{n-1}	b_{n-2}	b_{n-3}	$\ldots$	b_0	$P(x_0)$

que también es una forma sencilla de calcular $P(x_0)$. Además,

$$P(x) = Q(x)(x - x_0) + R \Rightarrow P'(x) = Q'(x)(x - x_0) + Q(x) \Rightarrow P'(x_0) = Q(x_0).$$

De modo que para calcular $P'(x_0)$ podemos aplicar el algoritmo gráfico anterior a $Q(x)$:

	a_n	a_{n-1}	a_{n-2}	$\ldots$	a_1	a_0
x_0		$b_{n-1}x_0$	$b_{n-2}x_0$	$\ldots$	$b_1 x_0$	$b_0 x_0$
	b_{n-1}	b_{n-2}	b_{n-3}	$\ldots$	b_0	$P(x_0)$
x_0		$c_{n-2}x_0$	$c_{n-3}x_0$	$\ldots$	$c_0 x_0$	
	c_{n-2}	c_{n-3}	c_{n-4}	$\ldots$	$P'(x_0)$	

Ejemplo: *Siendo* $P(x) = x^4 + 2x^3 - 3x^2 - 4x - 1$*, calcula con el algoritmo de Horner* $P(2)$ *y* $P'(2)$*.*

	2	1	−3	4	−5
2		4	10	14	36
	2	5	7	18	31
2		4	18	50	
	2	9	25	68	

Se concluye que $P(2) = 31$ y $P'(2) = 68$.

□

2.4.4 Raíces múltiples

Si una ecuación $f(x) = 0$ tiene soluciones múltiples, el cálculo puede ser problemático. Si $f(x)$ es un polinomio, podemos intentar calcular las raíces de

$$Q(x) = \frac{P(x)}{MCD(P(x), P'(x))}$$

que será un polinomio con las mismas raíces que $f(x)$ pero todas con multiplicidad 1.

Sea o no $f(x)$ un polinomio, se puede intentar aplicar el método de Newton-Raphson. Si la sucesión generada converge muy lentamente, la causa puede encontrarse en esa multiplicidad de las raíces. Para detectar de qué multiplicidad es la raíz en cuestión, para cada termino de la sucesión x_n generada calculamos $f'(x_n), f''(x_n), f'''(x_n), \ldots, f^{k)}(x_n)$. Si las derivadas hasta orden $k-1$ son casi 0 en los valores x_n y $f^{k)}(x_n)$ no es cercano a 0, entonces puede ocurrir que la solución r verifique que

$$f^{n)}(r) = 0 \quad \forall\ 0 \leq n < k \quad \text{y} \quad f^{k)}(r) \neq 0,$$

es decir que r sea una solución de $f(x) = 0$ de multiplicidad k. En este caso, se puede acelerar la convergencia de la sucesión calculándola así

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

El error se obtendría así

$$|e_n| < \frac{|f(x_n)|}{\min_{\theta \in [a,b]} |f'(\theta)|}$$

Ejercicio: *Sea* $f(x) = (x-1)^3$. *Aplica Newton-Raphson para encontrar una solución de* $f(x) = 0$ *empezando en* $x_0 = 0$.

Capítulo 3

Sistemas Lineales

El objetivo de este tema es el desarrollo de técnicas de búsqueda de soluciones de sistemas de ecuaciones lineales como el siguiente:

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

Denotando $A = (a_{ij})_{1\leq i,j\leq n}$, $b = (b_i)_{1\leq i\leq n}$ y $x = (x_i)_{1\leq i\leq n}$, el sistema anterior se puede expresar

$$A \cdot x = b$$

3.1 Álgebra de matrices

Sea A una matriz.

A es de orden $m\times n$ si tiene tiene m filas y n columnas ($A = (a_{ij})_{1\leq i\leq m,1\leq j\leq n} \in \mathcal{M}_{m\times n}$). Por ejemplo

$$\begin{pmatrix} 3 & 5 & 2 \\ 2 & 1 & 3 \end{pmatrix}$$

tiene orden 2×3.

La matriz traspuesta de A es la matriz A^t que resulta de intercambiar las filas de A por las columnas de A.

A es simétrica si $A^t = A$.

Si $\alpha \in \mathbb{R}$ y B es una matriz, podemos realizar las siguientes operaciones:

- Multiplicar una matriz por un número:

$$\alpha \cdot A = (\alpha a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}.$$

- Sumar dos matrices A y B si $A, B \in \mathcal{M}_{m \times n}$:

$$A + B = (a_{i,j} + b_{ij})_{1 \leq i,j \leq m}.$$

- Multiplicar dos matrices $A \in \mathcal{M}_{m \times p}$ y $B \in \mathcal{M}_{p \times n}$:

$$A \cdot B = (\sum_{k=1}^{p} a_{ik} b_{kj})_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathcal{M}_{m \times n}.$$

La matriz identidad de orden $n \times n$ es

$$Id = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ & \dots & & & \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Se verifica que $A \cdot Id = Id \cdot A = A$, $\forall A \in \mathcal{M}_{n \times n}$.

B es una matriz inversa de A por la derecha si $B \cdot A = Id$

B es una matriz inversa de A por la izquierda si $A \cdot B = Id$.

Si $A \in \mathcal{M}_{n \times n}$, entonces la inversa por la izquierda es también inversa por la derecha. En este caso la inversa es única y se denota A^{-1}

$$A^{-1} \cdot A = A \cdot A^{-1} = Id.$$

Si $A \in \mathcal{M}_{n \times n}$ tiene inversa, entonces A es regular, invertible o no singular. La matriz inversa es

$$A^{-1} = \frac{1}{|A|} (A_{i,j})^t,$$

siendo $A_{i,j}$ el adjunto del elemento a_{ij}.

En este caso el sistema $A \cdot x = b$ tiene como solución $x = A^{-1} \cdot b$.

Se llaman operaciones elementales de una matriz al intercambio de dos filas o columnas, al producto de una fila o columna por un número distinto de 0, a la suma de una fila o columna otra fila o columna multiplicada por un número, o cualquier combinación finita de las operaciones anteriores. Cada

operación fundamental en una matriz A se puede expresar mediante producto de A por otra matriz, que se llama matriz fundamental. Por ejemplo, dada la matriz

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

el intercambio de la segunda y tercera filas se puede expresar así:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}.$$

El producto de la segunda fila por un número α se puede expresar así:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Y la suma a la tercera fila de la segunda fila multiplicada por un número α es:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \alpha a_{21} + a_{31} & \alpha a_{22} + a_{32} & \alpha a_{23} + a_{33} \end{pmatrix}$$

El producto de matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \alpha \\ 0 & 1 & \alpha\lambda \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

representa aplicar a la segunda matriz los siguientes cambios:

- Sustituir la fila segunda por la tercera multiplicada por α.
- Multiplicar la tercera fila por $\alpha\lambda$ y sumarle la segunda.

Si $A \in \mathcal{M}_{n\times n}$ tiene inversa, entonces existe un conjunto de matrices fundamentales E_k con $k = 1, \ldots, p$, tales que

$$E_p E_{p-1} \ldots E_1 A = Id,$$

con lo que multiplicando por A^{-1} por la derecha en ambos términos se obtiene que

$$E_p E_{p-1} \ldots E_1 Id = A^{-1},$$

es decir, aplicando las mismas operaciones fundamentales a la matriz Id obtendremos la matriz inversa A^{-1}.

Ejemplo: *Calcula la inversa de la matriz*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \\ 2 & 4 & 7 \end{pmatrix}.$$

Partimos de A y la matriz Id. El primer cambio consiste en restar a la segunda fila la primera (correspondiente al producto por una cierta matriz fundamental E_1):

$$E_1 \cdot A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 2 & 4 & 7 \end{pmatrix} \qquad E_1 \cdot Id = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

El segundo cambio (multiplicar por E_2) consiste en restar a la tercera dos veces la primera:

$$E_2 \cdot E_1 \cdot A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad E_2 \cdot E_1 \cdot Id = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

A la primera le resto dos veces la segunda (E_3):

$$E_3 \cdot E_2 \cdot E_1 \cdot A = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad E_3 \cdot E_2 \cdot E_1 \cdot Id = \begin{pmatrix} 3 & -2 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

A la primera le resto tres veces la tercera (E_4):

$$E_4 \cdot E_3 \cdot E_2 \cdot E_1 \cdot A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad E_4 \cdot E_3 \cdot E_2 \cdot E_1 \cdot Id = \begin{pmatrix} 9 & -2 & -3 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} = A^{-1}.$$

Esta última matriz es la inversa de A.

□

Para toda matriz $A \in \mathcal{M}_{n\times n}$ son equivalentes:

1. A^{-1} existe.
2. $|A| \neq 0$.
3. Los vectores fila de la matriz A son linealmente independientes.
4. Los vectores columna de la matriz A son linealmente independientes.
5. Para cada $b \in \mathbb{R}^n$, el sistema $A \cdot x = b$ tiene una única solución.
6. A es producto de matrices elementales.

3.1.1 Valores propios y vectores propios

Sea $A \in \mathcal{M}_{n\times n}$. Si existe un vector $x \in R^n$ no nulo y un número $\lambda \in R$ tal que $A \cdot x = \lambda x$, entonces, λ es un valor propio o autovalor de la matriz A, y x es un vector propio o autovector para el autovalor λ. Los valores propios de una matriz A son las raíces del llamado polinomio característico $|A - \lambda \cdot Id|$. Los vectores propios de A para el autovalor λ, son los elementos de núcleo de la aplicación lineal $A - \lambda \cdot Id$.

3.1.2 Matriz definida

Sea $A \in \mathcal{M}_{n\times n}$. Se dice que A es definida positiva (equivalentemente, negativa) si para todo $x \in \mathbb{R}^n$ no nulo, se verifica que

$$x^t \cdot A \cdot x > 0.$$

Ejemplo. *La matriz* $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ *es definida positiva.*

Sea (x_1, x_2) un vector no nulo.

$$(x_1, x_2) \cdot \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \cdot \begin{pmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{pmatrix} =$$
$$2x_1^2 + x_1x_2 + x_1x_2 + 2x_2^2 = (x_1 + x_2)^2 + x_1^2 + x_2^2 > 0.$$

□

Si A es una matriz definida positiva y simétrica, entonces los autovalores de A son todos números reales y positivos.

3.2 Resolución de sistemas de ecuaciones lineales: método de Gauss

Dado el sistema $A \cdot x = b$, si $|A| = 0$, puede no haber solución. Si $|A| \neq 0$, existe una única solución y se puede calcular así:

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & \dots & a_{1n} \\ b_2 & a_{22} & \dots & a_{2n} \\ & \dots & & \\ b_n & a_{n2} & \dots & a_{nn} \end{vmatrix}}{|A|}, \quad \cdots, \quad x_n = \frac{\begin{vmatrix} a_{11} & \dots & a_{1n-1} & b_1 \\ a_{21} & & a_{2n-1} & b_2 \\ & \dots & & \\ a_{2n} & \dots & a_{nn-1} & b_n \end{vmatrix}}{|A|}$$

Los cálculos anteriores son poco prácticos cuando el número de ecuaciones es grande. Por ejemplo, para $n = 50$ se necesitan 10^{64} operaciones. Para evitar este problema se utilizan métodos numéricos de resolución de ecuaciones lineales que pueden ser:

- Directos: proporcionan un resultado que será exacto salvo errores de redondeo tras un número determinado de operaciones.
- Iterativos: construyen una sucesión de soluciones aproximadas, de modo que en cada paso se mejora la aproximación anterior.

Se dice que dos sistemas de ecuaciones son equivalentes si tienen las mismas soluciones. En un sistema de ecuaciones se pueden efectuar operaciones elementales en la matriz de los coeficientes y en la de los términos independientes y resulta un sistema equivalente. Es decir, resulta un sistema equivalente si efectuamos alguna de las siguientes operaciones:

1. Multiplicamos una ecuación por un número distinto de cero.
2. Sumamos a una ecuación otra ecuación multiplicada por un número.
3. Intercambiamos dos ecuaciones.

3.2.1 Método de Gauss

El método de Gauss clásico es un método directo que transforma mediante operaciones elementales el sistema de ecuaciones que se pretende resolver en un sistema equivalente triangular de fácil resolución.

El método de Gauss puede resumirse en lo siguiente:

- Paso 1. Si $a_{11} \neq 0$, se elige a_{11} como elemento pivote. Si $a_{11} = 0$, se intercambian las filas para que esto no suceda. Sumando a cada una las filas restantes la fila del elemento pivote multiplicada por ciertos números, llamados multiplicadores, se convierten en 0 los elementos que se encuentran en la primera columna distintos del pivote. En las operaciones de sumas de filas hay que incluir los términos independientes.

- Paso i ($i \geq 2$). Si $a_{ii} \neq 0$, se elige a_{ii} como elemento pivote. Si $a_{ii} = 0$, se intercambian las filas de orden mayor o igual que i para que esto no suceda. Sumando a cada una de las filas que ocupan el lugar j ($j > i$) la fila del elemento pivote multiplicada por números, llamados multiplicadores, se convierten en 0 los elementos que se encuentran en la columna i de la fila j ($j > i$). En las operaciones de sumas de filas hay que incluir los términos independientes.

Se efectúan los pasos necesarios hasta que la matriz de los coeficientes sea triangular. Conseguido esto, el cálculo de la solución es sencillo siguiendo el siguiente orden: $x_n, x_{n-1}, \ldots, x_1$.

Ejemplo: *Resuelve el siguiente sistema de ecuaciones:*

$$\begin{array}{rcl} 6x_1 - 2x_2 + 2x_3 + 4x_4 & = & 12 \\ 12x_1 - 8x_2 + 6x_3 + 10x_4 & = & 34 \\ 3x_1 - 13x_2 + 9x_3 + 3x_4 & = & 27 \\ -6x_1 + 4x_2 + x_3 - 18x_4 & = & -38. \end{array}$$

En forma matricial es sistema sería

$$\begin{pmatrix} 6 & -2 & 2 & 4 \\ 12 & -8 & 6 & 10 \\ 3 & -13 & 9 & 3 \\ -6 & 4 & 1 & -18 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 12 \\ 34 \\ 27 \\ -38 \end{pmatrix}.$$

En el sistema anterior el método de Gauss podría aplicarse así:

1. Paso 1. Elegimos como elemento pivote el término 6 de la primera fila y efectuamos las siguientes operaciones:

$$(\text{fila } 2^a) - 2(\text{fila } 1^a) \qquad (\text{fila } 3^a) - 1/2(\text{fila } 1^a) \qquad (\text{fila } 4^a) - (-1)(\text{fila } 1^a)$$

obtenemos el sistema

$$\begin{pmatrix} 6 & -2 & 2 & 4 \\ 0 & -4 & 2 & 2 \\ 0 & -12 & 8 & 1 \\ 0 & 2 & 3 & -14 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 12 \\ 10 \\ 21 \\ -26 \end{pmatrix}.$$

Paso 2. Elegimos como elemento pivote el término -4 de la segunda fila y efectuamos las siguientes operaciones:

$$(\text{fila } 3^a) - 3(\text{fila } 2^a) \qquad (\text{fila } 4^a) - (-1/2)(\text{fila } 2^a)$$

obtenemos el sistema

$$\begin{pmatrix} 6 & -2 & 2 & 4 \\ 0 & -4 & 2 & 2 \\ 0 & 0 & 2 & -5 \\ 0 & 0 & 4 & -13 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 12 \\ 10 \\ -9 \\ -21 \end{pmatrix}.$$

Paso 3. Elegimos como elemento pivote el término 2 de la tercera fila y efectuamos la siguiente operación:

$$(\text{fila } 4^a) - 2(\text{fila } 3^a)$$

obtenemos el sistema

$$\begin{pmatrix} 6 & -2 & 2 & 4 \\ 0 & -4 & 2 & 2 \\ 0 & 0 & 2 & -5 \\ 0 & 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 12 \\ 10 \\ -9 \\ -3 \end{pmatrix}.$$

que es un sistema de ecuaciones triangular superior de fácil resolución:

$$\left.\begin{array}{rcl} 6x_1 - 2x_2 + 2x_3 + 4x_4 &=& 12 \\ -4x_2 + 2x_3 + 2x_4 &=& 10 \\ 2x_3 - 5x_4 &=& -9 \\ -3x_4 &=& -3 \end{array}\right\} \Rightarrow \begin{array}{l} x_1 = 1/6(12 - 4x_4 - 2x_3 + 2x_2) = 1 \\ x_2 = -1/4(10 - 2x_4 - 2x_3) = -3 \\ x_3 = 1/2(-9 + 5x_4) = -2 \\ x_4 = 1 \end{array}$$

Se verifica que $A = L \cdot U$, siendo A la matriz de los coeficientes inicial, U la matriz de los coeficientes final y L la matriz formada por los números que hemos utilizado para multiplicar las filas (multiplicadores):

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1/2 & 3 & 1 & 0 \\ -1 & -1/2 & 2 & 1 \end{pmatrix}$$

□

3.2.2 Método de Gauss con pivoteo

En un sistema como este

$$\begin{pmatrix} 0.001 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

el método de Gauss no funcionaría correctamente por errores de redondeo. En general, si $\epsilon \neq 0$ es pequeño, el sistema

$$\begin{pmatrix} \epsilon & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

aplicando el método de Gauss se transformaría en

$$\begin{pmatrix} \epsilon & 1 \\ 0 & 1 - \frac{1}{\epsilon} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 - \frac{1}{\epsilon} \end{pmatrix}$$

con lo que las soluciones, si consideramos el redondeo, serían:

$$x_2 = \frac{2 - \frac{1}{\epsilon}}{1 - \frac{1}{\epsilon}} \simeq 1 \qquad \Rightarrow \qquad x_1 = (1 - x_2)\frac{1}{\epsilon} \simeq 0.$$

Si $\epsilon = 0.001$, obtendríamos que $x_2 = 0.998998998999$, que, redondeando a dos decimales, sería $x_2 = 0$, por lo que $x_1 = 0$. Sin embargo las soluciones exactas son $x_2 = \frac{998}{999} \approx 1$ y $x_1 = 1000/999 \approx 1$. De hecho, al hacer los cálculos con le programa Maxima o Mathematica, con $\epsilon = 10^{-17}$ se obtiene $x_2 = 1$ y $x_1 = 0$.

Esto sucede cuando el elemento pivote en algún paso es muy pequeño comparado con el resto de los elementos a los que divide. En este caso el

multiplicador correspondiente será muy grande y la ecuación que origina es "casiün múltiplo de la ecuación del elemento pivote:

$$(\text{fila } 2^a) - \frac{1}{\epsilon}\ (\text{fila } 1^a) \approx -\frac{1}{\epsilon}\ (\text{fila } 1^a).$$

Este problema se puede resolver de dos formas:

1. Pivoteo parcial: si estamos en el paso i, se elige como elemento pivote el elemento a_{ki} que tenga mayor valor absoluto entre los de la columna i, con $k \geq i$, independientemente de si a_{ii} es cero o no, y se reordenan las filas.

2. Pivoteo total: si estamos en el paso i, se elige como elemento pivote el elemento a_{kj} que tenga mayor valor absoluto entre los de la columna i y los de la fila i, con $k, j \geq i$, independientemente de si a_{ii} es cero o no, y se reordenan las filas y columnas. Esta opción tiene un importante inconveniente pues hay que reordenar las incógnitas.

Ejemplo: *Resuelve el siguiente sistema de ecuaciones:*

$$\begin{aligned} 2x_2 + x_4 &= 0 \\ 2x_1 + 2x_2 + 3x_3 + 2x_4 &= -2 \\ 4x_1 - 3x_2 + x_4 &= -7 \\ 6x_1 + x_2 - 6x_3 - 5x_4 &= 6. \end{aligned}$$

aplicando el pivoteo parcial.

La matriz de los coeficientes y de los términos independientes sería

$$\left(\begin{array}{cccc|c} 0 & 2 & 0 & 1 & 0 \\ 2 & 2 & 3 & 2 & -2 \\ 4 & -3 & 0 & 1 & -7 \\ 6 & 1 & -6 & -5 & 6 \end{array}\right)$$

El pivoteo parcial sería este proceso:

Paso 1. Elegimos como elemento pivote el término 6 de la última fila por ser el de mayor valor absoluto de entre los elementos de la primera columna. Obtenemos la matriz

$$\left(\begin{array}{cccc|c} 6 & 1 & -6 & -5 & 6 \\ 2 & 2 & 3 & 2 & -2 \\ 4 & -3 & 0 & 1 & -7 \\ 0 & 2 & 0 & 1 & 0 \end{array}\right)$$

Efectuamos las operaciones indicadas obtenemos la matriz siguiente:

$$\begin{matrix}(\text{fila } 2^a) - \frac{2}{6}\ (\text{fila } 1^a) \\ (\text{fila } 3^a) - \frac{4}{6}\ (\text{fila } 1^a)\end{matrix} \rightsquigarrow \left(\begin{array}{cccc|c} 6 & 1 & -6 & -5 & 6 \\ 0 & 1.6667 & 5 & 3.6667 & -4 \\ 0 & -3.6667 & 4 & 4.3333 & -11 \\ 0 & 2 & 0 & 1 & 0 \end{array}\right)$$

Paso 2. Elegimos como elemento pivote el término -3.667 de la tercera fila por ser el de mayor valor absoluto de entre los elementos de la segunda columna. Obtenemos la matriz

$$\left(\begin{array}{cccc|c} 6 & 1 & -6 & -5 & 6 \\ 0 & -3.6667 & 4 & 4.3333 & -11 \\ 0 & 1.6667 & 5 & 3.6667 & -4 \\ 0 & 2 & 0 & 1 & 0 \end{array}\right)$$

Efectuamos las operaciones indicadas obtenemos la matriz siguiente:

$$\begin{matrix}(\text{fila } 3^a) - \frac{1.6667}{-3.6667}\ (\text{fila } 2^a) \\ (\text{fila } 4^a) - \frac{2}{-3.6667}\ (\text{fila } 2^a)\end{matrix} \rightsquigarrow \left(\begin{array}{cccc|c} 6 & 1 & -6 & -5 & 6 \\ 0 & -3.6667 & 4 & 4.3333 & -11 \\ 0 & 0 & 6.8182 & 5.6364 & -9.0001 \\ 0 & 0 & 2.1818 & 3.3636 & -5.9999 \end{array}\right)$$

Paso 3. El elemento pivote será 6.8182 de la tercera fila y no es necesario intercambiar filas. Efectuamos la operación indicada y obtenemos la matriz siguiente:

$$(\text{fila } 4^a) - \frac{2.1818}{6.8182}\ (\text{fila } 3^a) \rightsquigarrow U = \left(\begin{array}{cccc|c} 6 & 1 & -6 & -5 & 6 \\ 0 & -3.6667 & 4 & 4.3333 & -11 \\ 0 & 0 & 6.8182 & 5.6364 & -9.0001 \\ 0 & 0 & 0 & 1.5600 & -3.1199 \end{array}\right)$$

que se corresponde con el sistema de ecuaciones

$$\begin{array}{rcl} 6x_1 + x_2 + -6x_3 - 5x_4 & = & 6 \\ -3.6667x_2 + 4x_3 + 4.3333x_4 & = & -11 \\ 6.8182x_3 + 5.6364 & = & -9.0001 \\ 1.5600x_4 & = & -3.1199. \end{array} .$$

Este sistema se resuelve fácilmente por ser triangular despejando las incógnitas en orden inverso. Se obtiene la solución:

$$x_4 = -1.9999, \quad x_3 = 0.33325, \quad x_2 = 1.0000, \quad x_1 = -0.50000,$$

que es una respuesta aceptable teniendo en cuenta que la solución exacta es

$$x_4 = -2, \quad x_3 = 1/3, \quad x_2 = 1, \quad x_1 = -1/2.$$

□

La matriz L de los multiplicadores en el ejemplo anterior es la siguiente:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.6667 & 1 & 0 & 0 \\ 0.3333 & -0.45454 & 1 & 0 \\ 0 & -0.54545 & 0.32 & 1 \end{pmatrix}$$

y se verifica que

$$L \cdot U = \begin{pmatrix} 6 & 1 & -6 & 5 \\ 4 & -3 & 0 & 1 \\ 2 & 2 & 3 & 2 \\ 0 & 2 & 0 & 1 \end{pmatrix} := A'.$$

El resultado no es exactamente A porque al aplicar el pivoteo hemos permutado filas. En todo caso se verifica que existe una matriz P, que representa permutaciones de filas si empleamos pivoteo parcial, tal que

$$P \cdot A = A' = L \cdot U.$$

Esto último nos permite calcular el determinante de A de forma sencilla al ser $|L| = 1$, $|P| = (-1)^p$ (p es el número de permutaciones de filas) y U una matriz triangular:

$$|A| = |P^{-1} \cdot A'| = |P^{-1} \cdot L \cdot U| = |P^{-1}||L||U| = (-1)^p|U|.$$

El método de Gauss permite:

1. Encontrar las soluciones de un sistema de ecuaciones.
2. Calcular el determinante de la matriz de los coeficientes.
3. Descomponer la matriz de los coeficientes A en un producto de matrices $P \cdot A'$ donde P representa una permutación de filas de A y $A' = L \cdot U$, siendo L y U matrices triangulares inferior y superior, respectivamente. $A = A'$ si no se utiliza el pivoteo.

3.3 Factorización LU. Factorización de Cholesky

Dado el sistema de ecuaciones $Ax = b$, si aplicamos el método de Gauss sin pivoteo solo a la matriz de los coeficientes y no a los términos independientes b, obtendremos una descomposición de la matriz A en dos matrices $A = L \cdot U$, siendo L triangular inferior y U triangular superior. Esto permite, por ejemplo, calcular de forma sencilla el determinante de A:

$$|A| = |L||U| = (\prod_{i=1}^{n} l_{ii})(\prod_{i=1}^{n} u_{ii}).$$

Además, podemos resolver el sistema de ecuaciones siguiendo estos dos pasos:

Paso 1. Resolvemos el sistema $Ly = b$ y obtenemos una solución y.

Paso 2. Resolvemos el sistema $Ux = y$. La solución x obtenida será la solución de $Ax = b$ porque

$$Ax = LUx = Ly = b.$$

Los métodos de factorización LU consisten en descomponer la matriz en producto de dos matrices triangulares, para después aplicar los pasos 1 y 2 anteriores y así resolver el sistema.

Si $A = L \cdot U$ tendremos que

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \dots & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ & \dots & & & \\ l_{n1} & l_{n2} & \dots & & l_{nn} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ & \dots & & u_{3n} \\ 0 & 0 & \dots & u_{nn} \end{pmatrix}.$$

Haciendo el producto de matrices e igualando término a término se obtienen n^2 ecuaciones lineales con $n^2 + n$ incógnitas. Si imponemos la condición

$$l_{11} = l_{22} = \cdots = l_{nn} = 1$$

obtendremos un sistema de ecuaciones fácilmente resoluble. Las soluciones dan lugar a una descomposición $L \cdot U$ que es la misma que la obtenida por el método de Gauss sin intercambio de filas.

Si A es no singular, una condición necesaria y suficiente para que A admita una descomposición $L \cdot U$ de la forma anterior es que los menores fundamentales (todos los determinantes de las matrices formadas con las primeras k filas y columnas) sean no nulos. En este caso la factorización es única.

Ejemplo: *Resuelve el siguiente sistema mediante factorización $L \cdot U$ de Gauss*

$$\begin{pmatrix} 3 & 1 & 2 \\ 6 & 3 & 2 \\ -3 & 0 & -8 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}.$$

Como $|A_{11}| = 3$, $|A_{21}| = \begin{vmatrix} 3 & 1 \\ 6 & 3 \end{vmatrix} \neq 0$ y $A_{33} = \begin{vmatrix} 3 & 1 & 2 \\ 6 & 3 & 2 \\ -3 & 0 & -8 \end{vmatrix} \neq 0$, el sistema admite la factorización $L \cdot U$ buscada. Igualando término a término las matrices siguientes

$$\begin{pmatrix} 3 & 1 & 2 \\ 6 & 3 & 2 \\ -3 & 0 & -8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}.$$

obtenemos

$$\left.\begin{array}{r} u_{11} = 3 \\ u_{12} = 1 \\ u_{13} = 2 \\ l_{21}u_{12} = 6 \\ l_{21}u_{12} + u_{22} = 3 \\ l_{21}u_{13} + u_{23} = 2 \\ l_{31}u_{11} = -3 \\ l_{31}u_{12} + l_{32}u_{22} = 0 \\ l_{31}u_{13} + l_{32}u_{23} + u_{33} = -8 \end{array}\right\} \Rightarrow \begin{array}{l} u_{11} = 3 \\ u_{12} = 1 \\ u_{13} = 2 \\ l_{21} = 2 \\ u_{22} = 1 \\ u_{23} = -2 \\ l_{31} = -1 \\ l_{32} = 1 \\ u_{33} = -4 \end{array}$$

Paso 1.

$$Ly = b \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \Rightarrow \begin{array}{l} y_1 = 0 \\ y_2 = 1 \\ y_3 = 1 \end{array}.$$

Paso 2.

$$Ux = y \Rightarrow \begin{pmatrix} 3 & 1 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & -4 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \Rightarrow \begin{array}{l} x_1 = 0 \\ x_2 = \frac{1}{2} \\ x_3 = \frac{-1}{4} \end{array},$$

que es la solución del sistema.

□

3.3.1 Método de Crout

Si en la factorización LU fijamos

$$u_{11} = u_{22} = \cdots = u_{nn} = 1$$

se obtiene la descomposición por el método de Crout, que tiene el inconveniente que puede producir grandes errores en la resolución del sistema de ecuaciones.

3.3.2 Método de Cholesky

Si en la factorización LU exigimos que $U = L^t$ de modo que

$$l_{11} = u_{11}, l_{21} = u_{12}, l_{31} = u_{13}, \ldots, l_{ij} = u_{ji}, \ldots, l_{nn} = u_{nn},$$

se obtiene la descomposición por el método de Cholesky que funciona si y solo si la matriz inicial A es simétrica y definida positiva. Por tanto, si consigo la descomposición de una matriz por el método de Cholesky, puedo asegurar que dicha matriz es simétrica y definida positiva.

La resolución de un sistema de $n = 50$ ecuaciones con el método de Cholesky requiere $50^3/3$ operaciones.

3.3.3 Sistemas triangulares

Son sistemas de ecuaciones lineales en los que la matriz de los coeficientes es de la forma:

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & 0 & \ldots & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 0 & \ldots & 0 \\ 0 & a_{32} & a_{33} & a_{34} & 0 & \ldots & 0 \\ & \ldots & & & & & \\ 0 & 0 & \ldots & 0 & a_{nn-1} & a_{nn} & \end{pmatrix}.$$

Aparecen frecuentemente en la resolución numérica de ecuaciones diferenciales y en la aproximación por splines cúbicos. En la mayoría de los casos, la matriz admite una descomposición de la forma

$$\begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ l_{21} & 1 & 0 & \ldots & 0 \\ 0 & l_{32} & 1 & & 0 \\ & \ldots & & & \\ 0 & 0 & \ldots & l_{nn-1} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & 0 & \ldots & & 0 \\ 0 & u_{22} & u_{23} & 0 & & 0 \\ 0 & 0 & u_{32} & u_{33} & \ldots & 0 \\ & \ldots & & & & \\ 0 & 0 & \ldots & & u_{nn-1} & u_{nn} \end{pmatrix}.$$

Ejemplo: *Dada la matriz*

$$\begin{pmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{pmatrix}$$

a) Obtén la descomposición de Cholesky.
b) Obtén la factorización de Gauss sin intercambio de filas y, a partir de ella, las descomposiciones de Crout y de Cholesky.

a)

$$\begin{pmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{pmatrix} = LL^t = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix}$$

por tanto

$$\left.\begin{aligned} l_{11}^2 &= 60 \\ l_{11}l_{21} &= 30 \\ l_{11}l_{31} &= 20 \\ l_{21}^2 l_{22}^2 &= 20 \\ l_{21}l_{31} + l_{22}l_{32} &= 15 \\ l_{31}^2 + l_{32}^2 + l_{33}^2 &= 12 \end{aligned}\right\} \Rightarrow \begin{aligned} l_{11} &= \sqrt{60} \\ l_{21} &= \sqrt{60}/2 \\ l_{31} &= \sqrt{60}/3 \\ l_{22} &= \sqrt{5} \\ l_{32} &= \sqrt{5} \\ l_{33} &= \sqrt{3}/3 \end{aligned}$$

b) La factorización de Gauss se obtiene resolviendo el sistema de ecuaciones que se deduce de

$$\begin{pmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}.$$

La descomposición que resulta es

$$A = \begin{pmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/3 & 1 & 1 \end{pmatrix} \begin{pmatrix} 60 & 30 & 20 \\ 0 & 5 & 5 \\ 0 & 0 & 1/3 \end{pmatrix}.$$

Descomponiendo la segunda matriz de la parte derecha de la igualdad en un producto de una matriz diagonal por otra triangular superior con 1 en cada

elemento de la diagonal, se obtiene que

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/3 & 1 & 1 \end{pmatrix} \begin{pmatrix} 60 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1/3 \end{pmatrix} \begin{pmatrix} 1 & 1/2 & 1/3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = (*)$$

Multiplicando las dos primeras matrices anteriores

$$(*) = \begin{pmatrix} 60 & 0 & 0 \\ 30 & 5 & 0 \\ 20 & 4 & 1/3 \end{pmatrix} \begin{pmatrix} 1 & 1/2 & 1/3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

que es la descomposición de Crout.

Si en $(*)$ descomponemos la matriz diagonal

$$(*) = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/3 & 1 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{60} & 0 & 0 \\ 0 & \sqrt{5} & 0 \\ 0 & 0 & \sqrt{1/3} \end{pmatrix} \begin{pmatrix} \sqrt{60} & 0 & 0 \\ 0 & \sqrt{5} & 0 \\ 0 & 0 & \sqrt{1/3} \end{pmatrix} \begin{pmatrix} 1 & 1/2 & 1/3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} \sqrt{60} & 0 & 0 \\ \sqrt{60}/2 & \sqrt{5} & 0 \\ \sqrt{60}/3 & \sqrt{5} & \sqrt{3}/3 \end{pmatrix} \begin{pmatrix} \sqrt{60} & \sqrt{60}/2 & \sqrt{60}/3 \\ 0 & \sqrt{5} & \sqrt{5} \\ 0 & 0 & \sqrt{3}/3 \end{pmatrix}$$

que es la descomposición de Cholesky.

□

3.4 Normas y análisis del error

Definición: *Sea V un espacio vectorial. Una **norma** sobre V es una aplicación $\|\cdot\| : V \to \mathbb{R}$ tal que $\forall v, w \in V$ y $\forall \alpha \in \mathbb{R}$ se verifica que:*

1. $\|v\| \geq 0$ *($v = 0 \Leftrightarrow \|v\| = 0$)*
2. $\|v + w\| \leq \|v\| + \|w\|$
3. $\|\alpha v\| = |\alpha| \|v\|$.

Ejemplo: *las siguientes son normas definidas sobre el espacio vectorial* R^3*:*

1. *Norma euclídea o norma 2:* $\|(x,y,z)\|_2 = \sqrt{x^2+y^2+z^2}$
2. *Norma del máximo o norma infinito:* $\|(x,y,z)\|_\infty = max(|x|,|y|,|z|)$
3. *Norma 1:* $\|(x,y,z)\|_1 = |x|+|y|+|z|$

□

La norma de un vector mide la "distancia"de ese vector al origen. La distancia entre dos vectores se mide con la norma del vector diferencia. Las normas del ejemplo son diferentes formas de "medir"las distancias entre los vectores de $\mathbb{R}^3$.

Puesto que el conjunto $\mathcal{M}_{n\times n}$ de las matrices cuadradas de orden n, con las operaciones suma y producto por un número real es un espacio vectorial, para "medir"matrices también se utilizarán normas que, por tanto, tendrán que verificar las condiciones de la definición de norma. Si además de esas condiciones se verifica que

$$\|A\cdot B\| \leq \|A\|\|B\| \qquad A,B \in \mathcal{M}_{n\times n}$$

entones dicha norma es una **norma matricial**.

Una norma matricial $\|\cdot\|_M$ sobre $M_{n\times n}$ y una norma vectorial $\|\cdot\|_V$ sobre $\mathbb{R}^n$ se dice que son compatibles si se verifica que

$$\|Av\|_V \leq \|A\|_M \|v\|_V \qquad \forall A \in \mathcal{M}_{n\times n} \forall v \in \mathbb{R}^n.$$

Dada cualquier norma vectorial $\|\cdot\|_V$ sobre $\mathbb{R}^n$, es posible definir una norma matricial sobre $\mathcal{M}_{n\times n}$ de la siguiente forma:

$$\|A\|_M = \max\{\|Av\|_V : \|v\|_V = 1\}$$

Esta nueva norma se llama **norma matricial inducida o subordinada** a la norma vectorial correspondiente. La norma vectorial y su norma matricial subordinada son siempre compatibles.

Ejemplos:

1. *La norma matricial inducida por la norma vectorial 1 en* $\mathbb{R}^3$ *es*

$$\|A\|_1 = \left\| \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \right\|_1 = max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|.$$

De modo que $\left\| \begin{pmatrix} 3 & 1 & 2 \\ 6 & 3 & 2 \\ -3 & 0 & -8 \end{pmatrix} \right\|_1 = max(12, 4, 12) = 12.$

2. *La norma matricial inducida por la norma vectorial 2 en* $\mathbb{R}^3$ *es*

$$\|A\|_2 = \sqrt{\rho(A^t \cdot A)}.$$

$\rho(A^t \cdot A)$ es el radio espectral de $A^t \cdot A$ que es el máximo de los valores absolutos de los autovalores de la matriz $A^t \cdot A$.

Para calcular $\|A\|_2 = \left\| \begin{pmatrix} 2 & 3 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\|_2$ debemos calcular primero los autovalores de $\|A^t \cdot A\|$ que son las raíces de

$$|A^t A - \alpha Id| = \left| \begin{pmatrix} 13 - \alpha & -3 & 0 \\ -3 & 1 - \alpha & 0 \\ 0 & 0 & 1 - \alpha \end{pmatrix} \right| = (1 - \alpha)(\alpha^2 - 14\alpha + 4).$$

Los autovalores de $A^t A$ son $7 \pm \sqrt{5}$ y 1 y como consecuencia

$$\|A\|_2 = \left\| \begin{pmatrix} 2 & 3 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\|_2 = \sqrt{7 + \sqrt{5}}$$

3. *La norma matricial inducida por la norma infinito en* $\mathbb{R}^3$ *es*

$$\|A\|_\infty = \left\| \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \right\|_\infty = max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|.$$

De modo que $\left\| \begin{pmatrix} 3 & 1 & 2 \\ 6 & 3 & 2 \\ -3 & 0 & -8 \end{pmatrix} \right\|_\infty = \max(6, 11, 11) = 11.$

4. *Se define la norma matricial de Fröbenius como* $\|A\| = \sqrt{\sum_{1 \le i,j \le n} |a_{i,j}|^2}$.

□

3.4.1 Número condición de una matriz

Se define el *número de condición* de una matriz A como

$$k(A) = \|A\|\|A^{-1}\|.$$

El número de condición siempre es mayor o igual que 1 y se utiliza para estimar si un sistema de ecuaciones $Ax = b$ está bien o mal condicionado, es decir, si pequeños cambios en los datos (A o b) pueden producir grandes cambios en la solución del sistema.

Supongamos que en vez de b utilizamos $\widetilde{b}$ de modo que la solución del sistema es $\widetilde{x}$ y $A\widetilde{x} = \widetilde{b}$. Esto puede ser debido a que hay un pequeño error cometido al hacer alguna medida en b o bien porque en vez de la solución exacta x hemos obtenido una solución aproximada $\widetilde{x}$. Si denotamos e al error en la solución y r al error en b o error residual, se verifica que $Ae = Ax - A\widetilde{x} = b - \widetilde{b} = r$.

Teorema:

$$\frac{1}{k(a)} \frac{\|r\|}{\|b\|} \le \frac{\|e\|}{\|x\|} \le k(a) \frac{\|r\|}{\|b\|}.$$

Demostración. Veamos que $\frac{1}{k(a)} \frac{\|r\|}{\|b\|} \le \frac{\|e\|}{\|x\|}$. Se verifica que

$$\|r\|\|x\| = \|Ae\|\|A^{-1}b\| \le \|A\|\|e\|\|A^{-1}\|\|b\|$$

Por tanto

$$\frac{1}{\|A\|\|A^{-1}\|} \frac{\|r\|}{\|b\|} \le \frac{\|e\|}{\|x\|}$$

Vemos que $\frac{\|e\|}{\|x\|} \le k(a) \frac{\|r\|}{\|b\|}$. Se verifica que

$$\|e\|\|b\| \le \|A^{-1}r\|\|Ax\| \le \|A^{-1}\|\|r\|\|A\|\|x\|.$$

Por tanto

$$\frac{\|e\|}{\|x\|} \leq \|A\|\|A^{-1}\|\frac{\|r\|}{\|b\|}.$$

□

Si $k(A)$ están cerca de 1 el sistema estará bien condicionado porque el error relativo en la solución $\frac{\|e\|}{\|x\|}$ será similar al error relativo $\frac{\|r\|}{\|b\|}$ en los datos. Si por el contrario $k(A)$ es muy grande, el sistema puede estar mal condicionado y pequeños cambios en los datos podrán producir grandes cambios en la solución.

Ejemplo: *Dado el sistema* $\begin{pmatrix} 3.02 & -1.05 & 2.53 \\ 4.33 & 0.56 & -1.78 \\ -0.83 & -0.54 & 1.47 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$

1. *Encuentra el número de condición de la matriz A de los coeficientes con* $\|\cdot\|_\infty$.
2. *Acota el error relativo de las posibles soluciones en función del error en los datos.*

a) La matriz inversa redondeada a dos decimales es

$$A^{-1} = \begin{pmatrix} 5.66 & -7.27 & -18.55 \\ 200.51 & -268.26 & -669.91 \\ -76.85 & -102.65 & -255.88 \end{pmatrix}$$

con lo que

$$k(a) = \|A\|\|A^{-1}\| = 6.67 \cdot 1138.68 = 7595.$$

b) Se verifica que

$$\frac{1}{7595}\frac{\|r\|}{5} \leq \frac{\|e\|}{\|x\|} \leq 7595\frac{\|r\|}{5}$$

es decir

$$2.63 \cdot 10^{-5}\|r\| \leq \frac{\|e\|}{\|x\|} \leq 1591\|r\|$$

□

3.5 Mejora de soluciones

3.5.1 Refinamiento iterativo

Es un método para mejorar una solución aproximada ya obtenida. Sea $\widetilde{x}$ una solución aproximada de $Ax = b$ tal que $A\widetilde{x} = \widetilde{b}$. Se verifica que

$$x = A^{-1}b = \widetilde{x} + x - \widetilde{x} = \widetilde{x} + A^{-1}b + A^{-1}\widetilde{b} = \widetilde{x} + A^{-1}(b - \widetilde{b}) = \widetilde{x} + A^{-1}r = \widetilde{x} + e.$$

El método consiste en aplicar los siguientes pasos:
1. Calculo $r = b - A\widetilde{x}$ con doble precisión.
2. Calculo e resolviendo el sistema $Ae = r$.
3. La nueva solución aproximada será $\widetilde{\widetilde{x}} = \widetilde{x} + e$.

   Los pasos anteriores se repiten las veces necesarias para mejorar la solución.

Ejemplo: *Si como solución aproximada del sistema*

$$\begin{pmatrix} 420 & 210 & 140 & 105 \\ 210 & 140 & 105 & 84 \\ 140 & 105 & 84 & 70 \\ 105 & 84 & 70 & 60 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 875 \\ 539 \\ 399 \\ 319 \end{pmatrix}$$

se obtiene $\widetilde{x} = (0.999988, 1.000137, 0.999670, 1.000215)$*, aplica el método de refinamiento iterativo para mejorarla.*

1. Calculo $r = b - A\widetilde{x}$ con doble precisión.

$$r = \begin{pmatrix} 875 \\ 539 \\ 399 \\ 319 \end{pmatrix} - \begin{pmatrix} 420 & 210 & 140 & 105 \\ 210 & 140 & 105 & 84 \\ 140 & 105 & 84 & 70 \\ 105 & 84 & 70 & 60 \end{pmatrix} \begin{pmatrix} 0.999988 \\ 1.000137 \\ 0.999670 \\ 1.000215 \end{pmatrix} = \begin{pmatrix} -1.05000000 \cdot 10^{-4} \\ -7.0000000 \cdot 10^{-5} \\ -3.5000000 \cdot 10^{-5} \\ -4.8000000 \cdot 10^{-5} \end{pmatrix}.$$

2. Resuelvo

$$\begin{pmatrix} 420 & 210 & 140 & 105 \\ 210 & 140 & 105 & 84 \\ 140 & 105 & 84 & 70 \\ 105 & 84 & 70 & 60 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix} = \begin{pmatrix} -1.05 \cdot 10^{-4} \\ -7.0 \cdot 10^{-5} \\ -3.5 \cdot 10^{-5} \\ -4.8 \cdot 10^{-5} \end{pmatrix}$$

y obtengo $e = (1.1 \cdot 10^{-5}, -1.4 \cdot 10^{-5}, 3.30 \cdot 10^{-4}, -2.15 \cdot 10^{-4})$

3. La nueva solución es $\widetilde{\widetilde{x}} = (0.999999, 0.999997, 1, 1)$.

□

3.5.2 Escalamiento

Si en un sistema de ecuaciones los coeficientes de las incógnitas son de muy diferente magnitud, la solución numérica del sistema puede contener errores.

Ejemplo: Resolvamos el siguiente sistema por eliminación gaussiana:

$$\begin{aligned} 2x_1 + 100000x_2 &= 100000 \\ x_1 + x_2 &= 2 \end{aligned}$$

$$\begin{pmatrix} 2 & 100000 & 100000 \\ 1 & 1 & 2 \end{pmatrix} \begin{matrix} \text{(fila 2)} - 1/2 \text{ (fila 1)} \\ \rightsquigarrow \end{matrix} \begin{pmatrix} 2 & 100000 & 100000 \\ 0 & -999991 & -99998 \end{pmatrix}$$

De modo que las soluciones son

$$\begin{aligned} x_1 &= \tfrac{100000 - 100000x_2}{2} \approx 0 \\ x_2 &= \tfrac{-99998}{-99999} \approx 1 (\text{ redondeado a 4 cifras}) \end{aligned}$$

Sin embargo el resultado correcto es $x_1 = 1.00002$ y $x_2 = 0.99998$.

□

Para solucionar este problema se puede recurrir al escalamiento que consiste en multiplicar cada ecuación por un número para que el coeficiente más grande de las incógnitas en valor absoluto sea 1. Al aplicar el escalamiento podemos cometer a su vez errores de redondeo, pero puede ayudar en casos extremos cuando hay mucha diferencia entre los coeficientes.

Ejemplo: En el caso anterior

$$\begin{aligned} 0.00002x_1 + x_2 &= 1 \\ x_1 + x_2 &= 2 \end{aligned}$$

redondeando a 4 decimales

$$\begin{aligned} x_2 &= 1 \\ x_1 + x_2 &= 2 \end{aligned}$$

aplicando el pivoteo

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_2 &= 1 \end{aligned}$$

y las soluciones son $x_1 =$ y $x_2 = 1$.

□

3.6 Métodos iterativos

Los métodos directos estudiados anteriormente requieren aproximadamente $n^3/3$ operaciones. Son sensibles a errores de redondeo que se acrecientan al aumentar n. De modo que, aunque teóricamente conducen a una solución exacta, en la práctica la solución obtenida puede ser peor que la obtenida aproximadamente por un método iterativo.

Los métodos iterativos están especialmente indicados para la resolución de sistemas con una matriz de gran dimensión pero dispersa, es decir, con muchos ceros, que suelen aparecer por ejemplo en la resolución de ecuaciones diferenciales en derivadas parciales.

Un método iterativo de resolución de un sistema de n ecuaciones $Ax = b$ es aquel a partir de un $x^0 \in \mathbb{R}^n$ genera una sucesión de posibles soluciones aproximadas $x^1, x^2, \ldots$. El método es convergente si la sucesión generada converge a la solución del sistema a partir de cualquier vector inicial x^0. Es consistente si, en caso de converger, el límite es la solución.

Todos las métodos iterativos de resolución de un sistema $Ax = b$ se basan en la descomposición de la matriz A en diferencia de dos matrices $M - N$, siendo M una matriz regular. De modo que

$$Ax = b \Leftrightarrow (M - N)x = b \Leftrightarrow Mx = Nx + b \Leftrightarrow x = M^{-1}(Nx + b)$$

Si definimos $G(x) = M^{-1}(Nx+b)$, veremos que para encontrar la solución del sistema basta encontrar un vector fijo de la función $G(x)$. Para ello se genera una sucesión $x^k = G(x^{k-1})$ partiendo de una determinada solución aproximada x^0. La sucesión convergerá a la solución en determinadas condiciones.

En la práctica, dadas las matrices M y N, puesto que

$$x^k = G(x^{k-1}) = M^{-1}(Nx^{k-1}) + b) \Rightarrow Mx^k = Nx^{k-1} + b,$$

para calcular x^k a partir de x^{k-1} se resuelve el sistema

$$Mx^k = Nx^{k-1} + b.$$

Es esencial que M sea una matriz regular y es conveniente que sea una matriz sencilla para facilitar el cálculo.

La condición que asegura la convergencia es la siguiente.

Teorema: *En las condiciones anteriores, si existe una norma matricial tal que* $\|M^{-1}N\| < 1$*, entonces la sucesión* $x^{k+1} = G(x^k)$ *converge a un punto fijo de la función* $G(x)$*.*

Demostración. Teniendo en cuenta que

$$\begin{aligned}\|G(y) - G(y')\| &= \|M^{-1}(Ny + b) - M^{-1}(Ny' + b)\| \\ &= \|M^{-1}Ny - M^{-1}Ny'\| = \|M^{-1}N(y - y')\| \\ &\leq \|M^{-1}N\|\|y - y'\|.\end{aligned}$$

Si $\|M^{-1}N\| < \lambda < 1$ para un cierto $\lambda \in \mathbb{R}$, entonces $\|G(y)-G(y')\| < \lambda\|y-y'\|$. A partir de un vector x^0 vamos generando la sucesión $x^{k+1} = G(x^k)$ y se verifica que

$$\|x^{k+1} - x^k\| = \|G(x^k) - G(x)^{k-1}\| = \|G(x^k - x^{k-1})\| < \lambda\|x^k - x^{k-1}\| < \ldots < \lambda^k\|x^1 - x^0\|.$$

Por tanto

$$\begin{aligned}\|x^{k+p} - x^k\| &< \|x^{k+p} - x^{k+p-1} + x^{k+p-1} - \ldots - x^k\| \\ &\leq \sum_{i=1}^{p} \|x^{k+i} - x^{k+i-1}\| < \sum_{i=0}^{p-1} \lambda^{k+i-1}\|x^1 - x^0\| \\ &= \|x_1 - x_0\| \sum_{i=0}^{p-1} \lambda^{k+i-1} = \|x_1 - x_0\| \frac{\lambda^{k-1} - \lambda^{k+p-1}}{1 - \lambda} \overset{k\to\infty}{\to} 0.\end{aligned}$$

De modo que la sucesión $x^{k+1} = G(x^k)$ es una sucesión convergente a un vector $x \in \mathbb{R}^n$ si $\|M^{-1}N\| < \lambda < 1$ para un cierto $\lambda \in \mathbb{R}$. Dicho vector x verifica que

$$\begin{aligned}x - G(x) &= \lim_{k\to\infty} x^k - G(\lim_{k\to\infty} x^k) = \lim_{k\to\infty} G(x^{k-1}) - G(\lim_{k\to\infty} x^k) \\ &= G(\lim_{k\to\infty} (x^{k-1} - x^k)) = G(0) = 0 \Rightarrow x = G(x).\end{aligned}$$

□

En resumen, un método iterativo de resolución de un sistema $Ax = b$ consiste en generar una sucesión $x^k = G(x^{k-1})$ siendo $G(x) = M^{-1}(Nx + b)$ y $A = M - N$. Hay que tener en cuenta lo siguiente:

1. Si $|M| \neq 0$, puedo calcular x^k a partir de x^{k-1}.

2. Si el sistema $Mx = c$ se resuelve fácilmente, el método iterativo es más rápido.

3. Si $\|M^{-1}N\| < 1$ para alguna norma matricial, la sucesión x^k converge a la solución del sistema $Ax = b$.

3.6.1 Método de Jacobi

Dada la matriz $A = (a_{ij})_{1 \leq i,j \leq n}$, se definen:

$$D = \begin{pmatrix} a_{11} & 0 & \cdots & & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & & \cdots & 0 & a_{nn} \end{pmatrix}, \; L = \begin{pmatrix} 0 & \cdots & & & 0 \\ a_{21} & 0 & \cdots & & 0 \\ a_{31} & a_{32} & 0 & \cdots & 0 \\ & & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn-1} & 0 \end{pmatrix}$$

$$U = \begin{pmatrix} 0 & u_{12} & \cdots & & u_{1n} \\ 0 & 0 & u_{23} & \cdots & u_{2n} \\ & \cdots & & & \\ 0 & \cdots & & 0 & u_{n-1n} \\ 0 & \cdots & & 0 & 0 \end{pmatrix}$$

En el método de Jacobi $M = D$ y $N = -(L + U)$. Para calcula x^k a partir de x^{k-1} se resuelve el sistema

$$Dx^k = -(L + U)x^{k-1} + b.$$

Para aplicar el método se requiere que $a_{ii} \neq 0$ para todo i. Si $|A| \neq 0$, esto se puede conseguir intercambiando filas si fuese necesario.

El método converge si $\|M^{-1}N\| < 1$ para alguna norma matricial. Puesto que

$$M^{-1}N = \begin{pmatrix} 0 & \frac{-a_{12}}{a_{11}} & \frac{-a_{13}}{a_{11}} & \cdots & \frac{-a_{1n}}{a_{11}} \\ \frac{-a_{21}}{a_{22}} & 0 & \frac{-a_{23}}{a_{22}} & \cdots & \frac{-a_{2n}}{a_{22}} \\ & \cdots & & & \\ \frac{-a_{n1}}{a_{nn}} & \cdots & & \frac{-a_{nn-1}}{a_{nn}} & 0 \end{pmatrix}$$

y

$$\|M^{-1}N\|_\infty = \max_{1\le i\le n}\left(\sum_{j=1,j\neq i}^{n}|a_{ij}|\right)\frac{1}{a_{ii}},$$

una condición suficiente para que el método de Jacobi converja es que la matriz A sea diagonal dominante, es decir que $|a_{ii}| > \sum_{j=1,j\neq i}^{n}|a_{ij}| \quad \forall i = 1, 2, \ldots, n$.

Como

$$\|M^{-1}N\|_1 = \max_{1\le j\le n}\sum_{i=1,i\neq j}^{n}\left(|a_{ij}|\frac{1}{a_{ii}}\right),$$

otra condición suficiente es que lo anterior sea menor que 1.

Por último, una tercera condición suficiente es que la matriz $D - L - U$ sea simétrica y definida positiva.

Ejemplo: *Aplica una iteración del método de Jacobi al siguiente sistema*

$$\begin{pmatrix} 6 & 5 & 0 \\ 0 & 2 & 1 \\ 2 & 0 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

La matriz es diagonal dominante de modo que la sucesión x^k de vectores generados mediante el método de Jacobi convergerá a la solución. Se tiene que

$$D = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \qquad -(L+U) = \begin{pmatrix} 0 & -5 & 0 \\ 0 & 0 & -1 \\ -2 & 0 & 0 \end{pmatrix}$$

Si elegimos como vector inicial el $x^0 = (0, 0, 0)$, habrá que resolver el sistema

$$\begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} x_1^1 \\ x_2^1 \\ x_3^1 \end{pmatrix} = \begin{pmatrix} 0 & -5 & 0 \\ 0 & 0 & -1 \\ -2 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Por tanto

$$\begin{array}{ll} 6x_1^1 = 0 & x_1^1 = 0 \\ 2x_2^1 = 1 \;\Rightarrow & x_2^1 = \frac{1}{2} \\ 3x_3^1 = 0 & x_3^1 = 0 \end{array}$$

El siguiente vector sería $x^1 = (0, 1/2, 0)$.

□

3.6.2 Método de Gauss-Seidel

En este método $M = D + L$ y $N = -U$. El sistema a resolver para calcular x^k a partir de x^{k-1} será $(D+L)x^k = -Ux^{k-1}+b$. Distintas condiciones suficientes para que el método funcione son:

1. que A sea diagonal dominante
2. que A^t sea diagonal dominante
3. que A sea simétrica y definida positiva.

El método de Gauss-Seidel es dos veces más rápido que el de Jacobi.

Ejemplo: *Aplica el método de Gauss-Seidel al siguiente sistema*

$$\begin{pmatrix} 6 & 2 & 0 \\ 0 & 5 & 1 \\ 2 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Tendremos que

$$D + L = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 5 & 0 \\ 2 & 0 & 3 \end{pmatrix} \qquad U = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Si partimos de $x^0 = (0, 0, 0)$, habrá que resolver el sistema

$$\begin{pmatrix} 6 & 0 & 0 \\ 0 & 5 & 0 \\ 2 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1^1 \\ x_2^1 \\ x_3^1 \end{pmatrix} = \begin{pmatrix} 0 & -2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

y se obtiene que $x^1 = (0, 1/5, 0)$

□

3.6.3 Métodos de relajación

En el método de Jacobi, para calcular cada componente de x^k se utilizan todos los valores de x^{k-1}. Sin embargo, en el método de Gauss-Seidel, para calcular la componente i-ésima de x^k (x_i^k) se utilizan las componentes $x_1^k, x_2^k, \ldots, x_{i-1}^k$ del vector x^k y las componentes $x_{i+1}^{k-1}, \ldots, x_n^{k-1}$ del vector x^{k-1}. De este modo, el cálculo de las componentes de x^k se divide en etapas, utilizando en cada una de ellas un vector diferente:

Etapa 1: calculo x_1^k usando $(x_2^{k-1}, x_3^{k-1}, \ldots, x_n^{k-1})$.

Etapa 2: calculo x_2^k usando $(x_1^k, x_3^{k-1}, \ldots, x_n^{k-1})$.

Etapa 3: calculo x_3^k usando $(x_1^k, x_2^k, x_4^{k-1}, \ldots, x_n^{k-1})$.

Etapa n: calculo x_n^k usando $(x_1^k, x_2^k, \ldots, x_{n-1}^k)$.

En cada etapa se manejan tan solo vectores de $n-1$ variables en lugar de n con el consiguiente ahorro en tiempo. Además, la información obtenida en cada etapa se incorpora a la siguiente. Los métodos que utilizan esto último se llaman "métodos de relajación sucesiva"y se desarrollaron para la resolución de sistemas con matrices de dimensión grande pero con casi todos los elementos nulos.

En un método de relajación se considera

$$M = \frac{1}{w}D + L \qquad N = \frac{1-w}{w}D - U$$

con lo que $M - N = A$. w se llama factor de relajación. En función de su valor se tiene

- Subrelajación si $0 < w < 1$.
- Super-relajación si $w > 1$
- Gauss-Seidel si $w = 1$.

Si $w \notin (0, 2)$ el método no converge. Una condición suficiente para la convergencia es que A sea simétrica y definida positiva.

Capítulo 4

Aproximación de funciones

Dada una tabla de datos como esta

x	-1	0	1	2
y	-2	-2	0	4

que representa unos valores y en función de otros x, pretendemos encontrar una función $y = f(x)$ que pase por esos puntos.

4.1 Construcción del polinomio interpolador

Teorema: *Dada una tabla de $n+1$ puntos (x_0, y_0), (x_1, y_1), $\ldots$, (x_n, y_n) con $x_i \neq x_j$, existe un único polinomio $P_n(x)$ de grado menor o igual que n de modo que*

$$P_n(x_i) = y_i \qquad \forall i = 0, 1, \ldots, n.$$

Dicho polinomio se llama polinomio interpolador de los puntos.

Demostración. Consideremos un polinomio cualquiera de grado menor o igual que n :

$$P_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n.$$

Puesto que queremos que $P_n(x_i) = y_i \ \forall i = 0, 1, \ldots, n$ tendremos que

$$\begin{array}{c} a_0 + a_1x_0 + a_2x_0^2 + \cdots + a_nx_0^n = y_0 \\ a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n = y_1 \\ \cdots \\ a_0 + a_1x_n + a_2x_n^2 + \cdots + a_nx_n^n = y_n \end{array}$$

que es un sistema de $n+1$ ecuaciones lineales con $n+1$ incógnitas en el que el determinante de la matriz de los coeficientes es distinto de 0 :

$$\begin{vmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ & \cdots & & & \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{vmatrix} \neq 0.$$

Por lo tanto, el sistema es compatible determinado y tiene una única solución. □

Ejemplo: *Demuestra que existe un polinomio que pasa por los puntos de la tabla*

x	-1	0	1	2
y	-2	-2	0	4

Como son 4 puntos buscaremos un polinomio de grado menor o igual que 3 de la forma $P_n(x) = a_0 + a_1x + a_2x^2 + a_3x^3$. El sistema que resulta es

$$\begin{array}{rcl} a_0 + a_1(-1) + a_2(-1)^2 + a_3(-1)^3 & = & -2 \\ a_0 + a_1(0) + a_2(0)^2 + a_3(0)^3 & = & -2 \\ a_0 + a_1(1) + a_2(1)^2 + a_3(1)^3 & = & -2 \\ a_0 + a_1(2) + a_2(2)^2 + a_3(2)^3 & = & 4 \end{array}$$

Que es un sistema compatible determinado y que, por tanto, tiene una única solución, es decir, existe un único polinomio de grado menor o igual que 3 que pasa por los puntos de la tabla. Dicho polinomio es $P(x) = -2 + x + x^2$.

□

4.1.1 Método de Lagrange

Dados los puntos (x_0, y_0), (x_1, y_1), ..., (x_n, y_n), el polinomio interpolador $P_n(x)$ se puede calcular así:

$$P_n(x) = \sum_{i=0}^{n} L_i(x) y_i$$

donde

$$L_i(x) = \frac{\Pi_{j \neq i}(x - x_j)}{\Pi_{j \neq i}(x_i - x_j)} = \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}.$$

Se verifica que

$$L_i(x_i) = 1 \qquad L_i(x_j) = 0 \quad \forall i \neq j$$

de modo que

$$P_n(x_j) = \sum_{i=0}^{n} L_i(x_j) y_i = y_j.$$

Ejemplo: En el ejemplo anterior

$$L_0(x) = \frac{x(x-1)(x-2)}{(-1)(-2)(-3)} \quad L_1(x) = \frac{(x+1)(x-1)(x-2)}{1(-1)(-2)}$$
$$L_2(x) = \frac{(x+1)x(x-2)}{2 \cdot 1(-1)} \quad L_3(x) = \frac{(x+1)x(x-1)}{3 \cdot 2 \cdot 1}$$

□

4.1.2 Método de Newton

Dados los puntos (x_0, y_0), (x_1, y_1), ..., (x_n, y_n), se pueden calcular las *diferencias divididas*

- de primer orden: $[x_i, x_{i+1}] = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}$
- de segundo orden: $[x_i, x_{i+1}, x_{i+2}] = \frac{[x_{i+1}, x_{i+2}] - [x_i, x_{i+1}]}{x_{i+2} - x_i}$

...

- de orden k $[x_i, x_{i+1}, \dots, x_{i+k}] = \frac{[x_{i+1}, \cdots, x_{i+k}] - [x_i, \cdots, x_{i+k-1}]}{x_{i+k} - x_i}$

Las diferencias divididas son invariantes frente a permutaciones. Por ejemplo:

$$[x_0, x_1, x_2] = [x_1, x_0, x_2].$$

La tabla siguiente permite construir de forma sencilla las diferencias divididas:

x_0	y_0	$[x_0, x_1]$	$[x_0, x_1, x_2]$	$[x_0, x_1, x_2, x_3]$
x_1	y_1	$[x_1, x_2]$	$[x_1, x_2, x_3]$	
x_2	y_2	$[x_2, x_3]$		
x_3	y_3			
$\dots$				

El polinomio interpolador se construye a partir de las diferencias divididas así:

$$P_n(x) = y_0 + [x_0, x_1](x - x_0) + [x_0, x_1, x_2](x - x_0)(x - x_1) + \dots$$
$$\dots + [x_0, x_1, x_2, \dots, x_n](x - x_0)(x - x_1)\dots(x - x_{n-1}).$$

Ejemplo: En el ejemplo anterior la tabla sería la siguiente

-1	-2	0	1	0
0	-2	2	1	
1	0	4		
2	4			

y el polinomio sería $P(x) = -2 + 1(x + 1)x$.

□

4.2 Error del polinomio interpolador

Dados los puntos (x_0, y_0), (x_1, y_1), $\dots$, (x_n, y_n) que provienen de una cierta función $y = f(x)$, si conocemos cierta información sobre las derivadas de la función $f(x)$, podemos estimar el error que se comete al aproximar $f(x)$ por su polinomio interpolador $P_n(x)$ de grado menor o igual que n.

Teorema: *Sea $f(x)$ una función $n + 1$ veces diferenciable en un intervalo (a, b) y sea $P_n(x)$ su polinomio interpolador en los nodos $x_0, x_1, \dots, x_n$ contenidos en $[a, b]$. Para cada $x \in [a, b]$ existe un punto $\theta_x \in (a, b)$ tal que*

$$f(x) - P_n(x) = \frac{f^{n+1)}(\theta_x)}{(n+1)!}(x - x_0)(x - x_1)\dots(x - x_n).$$

Demostración. Sea $w(t) = (t - x_0)(t - x_1)\dots(t - x_n)$. Para un valor fijo $x \in [a, b]$ existe un $\lambda \in R$ tal que $f(x) - P_n(x) = \lambda w(x)$. De modo que tenemos la función

$$\Phi(t) = f(t) - P_n(t) - \lambda w(t)$$

que vale 0 en $x, x_0, x_1, \ldots, x_n$. Por tanto, $\Phi^{1)}(t)$ se anulará en $n+1$ puntos, $\Phi^{2)}(t)$ se anulará en n puntos y $\Phi^{n+1)}(t)$ se anulará en 1 punto $\theta_x \in (a,b)$:

$$\Phi^{n+1)}(\theta_x) = f^{n+1)}(\theta_x) - P^{n+1)}(\theta_x) - \lambda w^{n+1)}(\theta_x) = 0$$

es decir

$$f^{n+1)}(\theta_x) - \lambda(n+1)! = 0 \Rightarrow \lambda = \frac{f^{n+1)}(\theta_x)}{(n+1)!}.$$

Como consecuencia

$$f(x) - P_n(x) = \frac{f^{n+1)}(\theta_x)}{(n+1)!} w(x).$$

□

Ejemplo: El error al aproximar la función $f(x) = \text{sen}(x)$ en un punto $x \in [0,1]$ mediante un polinomio interpolador en 9 nodos contenidos en el intervalo $[0,1]$ se puede acotar así:

$$|f(x) - P_9(x)| \leq \frac{|f^{10)}(\theta_x)|}{10!} |\Pi_{i=0}^{9}(x - x_i)| \leq \frac{1}{10!} < 2.8 \cdot 10^{-7}$$

□

4.2.1 Elección de nodos. Polinomios de Chebyshev

Si pretendemos encontrar un polinomio que interpole a una determinada función y los nodos no están previamente fijados, ¿qué nodos podemos elegir? Elegir los nodos proporcionados por los polinomios de Chebyshev tienen alguna ventaja.

Definición: *Los polinomios de Chebyshev se definen así:*

$$T_0(x) = 1 \qquad T_1(x) = x \qquad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \text{ si } n \geq 1$$

Proposición: *Las raíces del polinomio de Chebyshev $T_n(x)$ si $n \geq 1$ son los valores*

$$\cos\left(\frac{2j-1}{2n}\pi\right) \qquad \forall 1 \leq j \leq n.$$

Como consecuencia, todas las raíces de cualquier polinomio de Chebyshev están contenidas en el intervalo $[-1, 1]$.

Teorema: *Si los nodos* (x_i)*, con* $i = 0, 1, \ldots, n$*, de interpolación de una función* $f(x)$ *en* $x \in [-1, 1]$ *son las raíces del polinomio de Chebyshev* $T_{n+1}(x)$*, entonces*

$$|f(x) - P_n(x)| \leq \frac{1}{2^n(n+1)!} \max_{|t|\leq 1} |f^{n+1)}(t)|.$$

La justificación del teorema se basa en que la expresión

$$\max_{|x|\leq 1} \Pi_{i=0}^{n} |(x - x_i)|$$

que proviene de la fórmula general del error en la interpolación, se hace mínima cuando x_i son las raíces del polinomio de Chebyshev correspondiente. El resultado del teorema permite encontrar una cota del error del polinomio interpolador más pequeña que la establecida con carácter general.

Si en lugar del intervalo $[-1, 1]$ tenemos un intervalo $[a, b]$, si aplicamos a los nodos de Chebyshev la transformación afín que lleva el intervalo $[-1, 1]$ en el intervalo $[a, b]$, obtenemos unos nuevos nodos tales que:

$$|f(x) - P_n(x)| \leq \frac{\left(\frac{b-a}{2}\right)^{n+1}}{2^n(n+1)!} \max_{t\in[a,b]} |f^{n+1)}(t)|.$$

Ejemplo: *Calcula los cuatro nodos para la interpolación de una función en el intervalo* $[-1, 1]$ *y en el intervalo* $[2, 8]$ *haciendo uso de los polinomio de Chebyshev. Establece una cota del error en cada caso para la función* $f(x) = \sin(x)$.

Calculamos las raíces de $T_4(x) = 8x^4 - 8x^2 + 1$:

$$x_0 = -0.9239 \qquad x_1 = -0.3827 \qquad x_2 = 0.3827 \qquad x_3 = 0.9239.$$

Estos sería los nodos en el intervalo $[-1, 1]$. Para calcular los nodos en el intervalo $[2, 8]$ necesitamos una transformación afín $\alpha x+\beta$ que lleve el intervalo

$[-1,1]$ en el intervalo $[2,8]$:

$$\begin{array}{rcl} [-1,1] & \to & [2,8] \\ -1 & \mapsto & \alpha(-1)+\beta = 2 \\ 1 & \mapsto & \alpha(1)+\beta = 8 \end{array}$$

Resolviendo el sistema formado por las dos últimas ecuaciones, obtenemos que la transformación es $3x+5$. Aplicamos esta transformación a los nodos de Chebyshev y obtenemos los nuevos nodos en el intervalo $[2,8]$:

$$x'_0 = 2.2284 \qquad x'_1 = 3.8519 \qquad x'_2 = 6.1481 \qquad x'_3 = 7.7716.$$

Veamos las cotas de los errores para la función $f(x) = \sin(x)$ en cada caso. Para el intervalo $[-1,1]$:

$$|f(x) - P(x)| \leq \frac{1}{2^n(n+1)!} \max_{|t|\leq 1} |f^{n+1)}(t)| \leq \frac{1}{2^3(4)!}$$

Para el intervalo $[2,8]$:

$$|f(x) - P_n(x)| \leq \frac{\left(\frac{b-a}{2}\right)^{n+1}}{2^n(n+1)!} \max_{t\in[a,b]} f^{n+1)}(t) \leq \frac{3^4}{2^3 4!}.$$

□

¿Es cierto que si n aumenta el polinomio $P_n(x)$ se "parece"más a la función $f(x)$? En general la respuesta es que no. Por ejemplo, dada función $f(x) = \frac{1}{1+x^2}$ en $[-5,5]$, si elegimos los nodos x_i igualmente espaciados, se puede demostrar que los polinomios no convergen a la función fuera del intervalo $[-3.63, 3.63]$. (Ver figura 4.1)

Otro ejemplo es la función $f(x) = |x|$ en $[-1,1]$, en la cual la convergencia se produce solo en los puntos $-1, 0, 1$.

Sin embargo, si elegimos los nodos de Chebyshev y la función $f(x)$ a interpolar es continua y con derivada continua, entonces $P_n(x)$ converge uniformemente a $f(x)$ en todo el intervalo. No obstante, para cualquier elección de nodos es posible encontrar una función continua tal que no se produce esa convergencia uniforme.

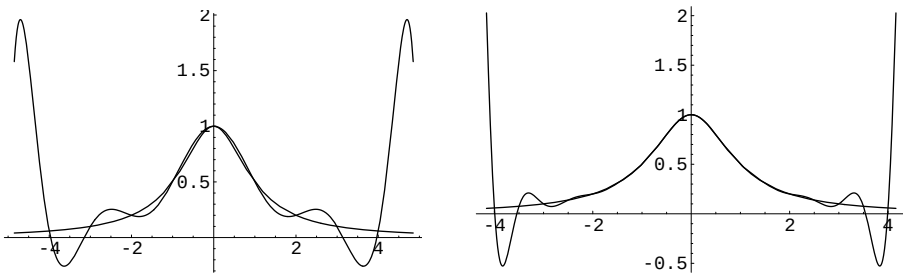
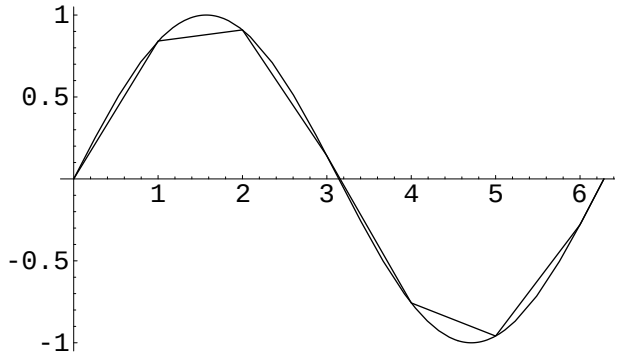


Figura 4.1: $\frac{1}{1+x^2}$ y sus polinomios interpoladores de grado 10 y 20.

4.3 Interpolación a trozos y con condiciones sobre la derivada

4.3.1 Interpolación a trozos

Si disponemos de ciertos valores de una función $f(x)$ en un intervalo $[a, b]$, se puede obtener una aproximación a la función $f(x)$ dividiendo el intervalo en varios intervalos más pequeños (subintervalos) y utilizando polinomios de pequeño grado (1,2,3, etc.) distintos para interpolar los valores de cada subintervalo. La siguiente gráfica muestra la función seno y una función que la interpola con trozos de recta en varios puntos:



Ejercicio: *Dada la tabla*

x	-1	0	1	2
y	-2	-2	0	4

encuentra una función formada
a) por trozos de recta
b) por trozos de parábola
que pase por los puntos anteriores. □

4.3.2 Interpolación con condiciones sobre la derivada

El siguiente es un ejemplo de problema de interpolación con condiciones sobre la derivada.

Ejemplo 1: *Encuentra un polinomio $P(x)$ tal que $P(0) = 0$, $P(1) = 1$ y $P'(0.5) = 2$.*

Puesto que son tres las condiciones que tenemos, cabe pensar que basta buscar entre los polinomios con tres parámetros, es decir, los polinomios de grado 3 de la forma $P(x) = a + bx + cx^2$:

$$\begin{array}{rcl} P(0) & = & 0 \\ P(1) & = & 1 \\ P'(0.5) & = & 2 \end{array} \Rightarrow \begin{array}{rcl} a & = & 0 \\ b+c & = & 1 \\ b+2\cdot c\cdot 0.5 & = & 2 \end{array} \Rightarrow \begin{array}{rcl} a & = & 0 \\ b+c & = & 1 \\ b+c & = & 2 \end{array}$$

El sistema anterior no tiene solución, con lo que habría que buscar entre los polinomios de grado 4. Al plantear las ecuaciones correspondientes resulta un sistema con solución no única.

□

Ejemplo 2 (Interpolación de Hermite): *Encuentra un polinomio $P(x)$ que interpole a una función $y = f(x)$ sabiendo que*

x	-1	0	1
y	1	1	-1
y'	2	0	0

Puesto que hay seis condiciones para la función y sus derivadas buscamos entre los polinomios de la forma $P(x) = a_0 + a_1x + \cdots + a_5x^5$ que tienen seis

coeficientes. Planteamos el sistema correspondiente según las condiciones y obtenemos lo siguiente:

$$\begin{array}{rclrcl} P(-1) & = & 1 & P'(-1) & = & 2 \\ P(0) & = & 1 & P'(0) & = & 0 \\ P(1) & = & 0 & P'(1) & = & 0 \end{array} \Rightarrow \begin{array}{ll} a_0 = 1 & a_3 = -3 \\ a_1 = 0 & a_4 = \frac{1}{2} \\ a_2 = \frac{-3}{2} & a_5 = 2 \end{array}$$

Con lo que el polinomio buscado es $P(x) = 1 - \frac{3}{2}x^2 - 3x^3 + \frac{1}{2}x^4 + 2x^5$.

□

Ejemplo 3 (Interpolación de Hermite cúbica a trozos): *Encuentra una función formada por trozos de polinomios de grado 3 que verifique las siguientes condiciones*

x	-1	0	1
y	1	1	-1
y'	2	0	0

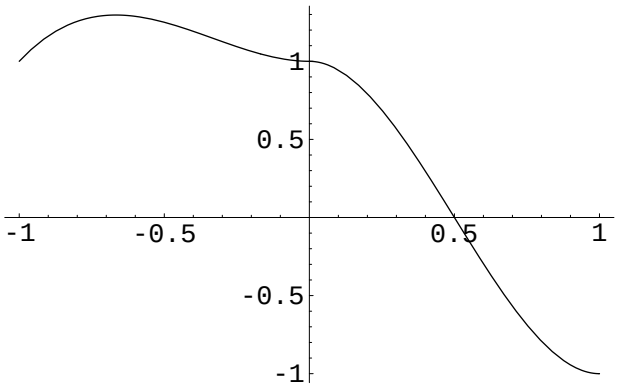


Figura 4.2: La función $g(x)$

Buscamos una función de la forma

$$g(x) = \begin{cases} P(x) & \text{si } x \in [-1, 0] \\ Q(x) & \text{si } x \in [0, 1] \end{cases}$$

donde $P(x) = a + bx + cx^2 + dx^3$ y $Q(x) = a' + b'x + c'x^2 + d'x^3$. A partir de las condiciones planteamos el sistema

$$\begin{array}{cccc} P(-1) = 1 & P'(-1) = 2 & Q(0) = 1 & Q'(0) = 0 \\ P(0) = 1 & P'(0) = 0 & Q(1) = -1 & Q'(1) = 0 \end{array}$$

con lo que

$$g(x) = \begin{cases} 1 + 2x^2 + 2x^3 & \text{si } x \in [-1, 0] \\ 1 - 6x^2 + 4x^3 & \text{si } x \in [0, 1] \end{cases}$$

(Ver Figura 4.2)

□

Ejemplo 5 (Splines cúbicos): *Encuentra una función polinómica cúbica a trozos que sea de clase $\mathcal{C}^2$ (continua, derivable y con derivada segunda continua) que verifique las siguientes condiciones*

x	-1	0	1
y	1	1	-1

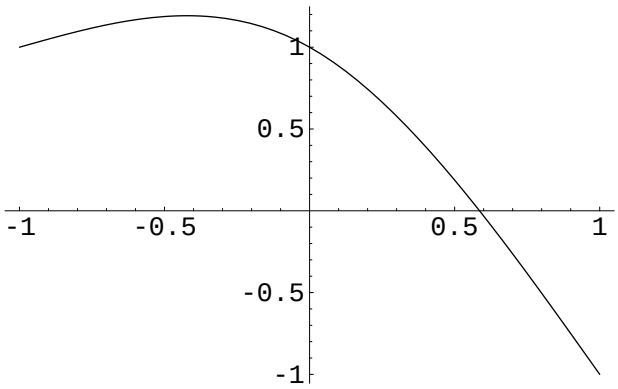


Figura 4.3: La función $h(x)$

Buscamos una función de la forma

$$h(x) = \begin{cases} P(x) & \text{si } x \in [-1, 0] \\ Q(x) & \text{si } x \in [0, 1] \end{cases}$$

donde $P(x) = a + bx + cx^2 + dx^3$ y $Q(x) = a' + b'x + c'x^2 + d'x^3$ de modo que

$$P(-1) = 1 \quad P(0) = 1 \quad Q(0) = 1 \quad Q(1) = -1 \quad P'(0) = Q'(0) \quad P''(0) = Q''(0)$$

Puesto que tenemos 6 ecuaciones y 8 incógnitas necesitamos 2 condiciones más. Por ejemplo, añadimos las siguientes

$$P''(-1) = 0 \qquad Q''(1) = 0$$

y obtenemos un sistema compatible determinado de modo que

$$h(x) = \begin{cases} 1 - x - \frac{3}{2}x^2 - \frac{1}{2}x^3 & \text{si } x \in [-1, 0] \\ 1 - x - \frac{3}{2}x^2 + \frac{1}{2}x^3 & \text{si } x \in [0, 1] \end{cases}$$

(Ver Figura 4.3)

□

Interpolación de Hermite

Un problema de interpolación de Hermite consiste en encontrar un polinomio $P(x)$ (Polinomio de Hermite) tal que

$$\begin{array}{lllll} P(x_0) = y_0 & P'(x_0) = y_0' & P''(x_0) = y_0'' & \cdots & P^{k_0-1)}(x_0) = y_0^{k_0-1} \\ P(x_1) = y_1 & P'(x_1) = y_1' & P''(x_1) = y_1'' & \cdots & P^{k_1-1)}(x_1) = y_1^{k_1-1} \\ & \cdots & & & \\ P(x_n) = y_n & P'(x_n) = y_n' & P''(x_n) = y_n'' & \cdots & P^{k_n-1)}(x_n) = y_n^{k_n-1} \end{array}$$

siendo $y_i, y_i', \cdots, y_i^{k)}$ los valores de una cierta función y sus derivadas en $n+1$ puntos. Si el número de condiciones impuestas es $k_1 + k_2 + \cdots + k_n = m + 1$ entonces

Teorema: *Existe un único polinomio de grado menor o igual que m que satisface las condiciones de interpolación de Hermite anteriores.*

Demostración. Sea $P(x)$ un polinomio de grado menor o igual que m. Las condiciones $P^{j)}(x_i) = y_i^j$ dan lugar a un sistema de $m + 1$ ecuaciones de la forma $Au = y$. Este sistema tiene solución única si el sistema $Au = 0$ tiene únicamente la solución 0, es decir, si el sistema formado por las condiciones $P^{j)}(x_i) = 0$ tiene como solución única el polinomio 0. Veamos que esto último

es cierto. Un polinomio $P(x)$ que sea solución del sistema $P^{j)}(x_i) = 0$ verifica que tiene la raíz x_i repetida k_i veces, con lo que si $P(x)$ no es cero, sería de grado mayor o igual que $m+1$, en contradicción con la hipótesis inicial.

□

Ejercicio: *¿Qué otro nombre recibe el polinomio de Hermite cuando solo hay un nodo?*

□

Teorema. *Sean $x_0, x_1, \ldots, x_n$ nodos distintos de $[a,b]$ y $f \in \mathcal{C}^2[a,b]$. Si $P(x)$ es el polinomio de grado menor o igual que $2n+1$ tal que $P(x_i) = f(x_i)$ y $P'(x_i) = f'(x_i)$, con $0 \leq i \leq n$, entonces $\forall x \in [a,b] \quad \exists \theta \in [a,b]$ tal que*

$$f(x) - P(x) = \frac{f^{2n+2)}(\theta)}{(2n+2)!}\Pi_{i=0}^{n}(x - x_i)^2$$

□

El polinomio de Hermite se puede calcular usando las diferencias divididas. Para ello definimos

$$[x_0, x_0] = lim_{x \to x_0}[x_0, x] = lim_{x \to x_0}\frac{y - y_0}{x - x_0} = lim_{x \to x_0}\frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$

$$[x_0, x_0, x_1] = \frac{[x_0, x_1] - [x_0, x_0]}{x_1 - x_0}$$

$$[x_0, x_0, x_0] = \frac{1}{2!}f^{2)}(x_0) \qquad [x_0, x_0, x_0, x_0] = \frac{1}{3!}f^{3)}(x_0)$$

Y de forma similar el resto de diferencias. En estas condiciones el polinomio de interpolación de Hermite se calcula de forma parecida a como se calculaba el polinomio de interpolación sin condiciones sobre la derivada.

El cálculo es más sencillo haciendo uso de una tabla. Por ejemplo, si el problema de Hermite incluye datos sobre la función y su derivada en tres puntos la tabla quedaría así:

x	y	1^0	2^0	3^0	4^0	5^0
x_0	y_0	$[x_0, x_0]$	$[x_0, x_0, x_1]$	$[x_0, x_0, x_1, x_1]$	$[x_0, x_0, x_1, x_1, x_2]$	$[x_0, \ldots, x_2]$
x_0	y_0	$[x_0, x_1]$	$[x_0, x_1, x_1]$	$[x_0, x_1, x_1, x_2]$	$[x_0, x_1, x_1, x_2, x_2]$	
x_1	y_1	$[x_1, x_1]$	$[x_1, x_1, x_2]$	$[x_1, x_1, x_2, x_2]$		
x_1	y_1	$[x_1, x_2]$	$[x_1, x_2, x_2]$			
x_2	y_2	$[x_2, x_2]$				
x_2	y_2					

siendo

$$\underset{k \text{ veces}}{[x_0, x_0, \ldots, x_0]} = \tfrac{1}{k!} f^{k)}(x_0) \quad .$$

El polinomio sería

$$\begin{aligned} P(x) = \; & y_0 + [x_0, x_0](x - x_0) + [x_0, x_0, x_1](x - x_0)^2 + \\ & + [x_0, x_0, x_1, x_1](x - x_0)^2 (x - x_1) + \\ & + [x_0, x_0, x_1, x_1, x_2](x - x_0)^2 (x - x_1)^2 \\ & + [x_0, x_0, x_1, x_1, x_2, x_2](x - x_0)^2 (x - x_1)^2 (x - x_2). \end{aligned}$$

Ejemplo: *Calcula el polinomio que verifica estas condiciones:*

x	-1	0	1
y	1	1	-1
y'	*2*	*0*	*0*

En este caso la tabla quedaría así

x	y	1^0	2^0	3^0	4^0	5^0
-1	**1**	**2**	-2	2	$-3/2$	2
-1	**1**	0	0	-1	$5/2$	
0	**1**	**0**	-2	4		
0	**1**	-2	2			
1	**-1**	**0**				
1	**-1**					

El polinomio sería

$P(x) = 1+2(x+1)-2(x+1)^2+2(x+1)^2x-3/2(x+1)^2x^2+2(x+1)^2x^2(x-1)$

□

Ejemplo: *Calcula el polinomio que verifica las siguientes condiciones:*

x	1	2
y	2	6
y'	3	7
y''		8

En este caso la tabla quedaría así

x	y	1^0	2^0	3^0	4^0
1	**2**	**3**	1	2	-1
1	**2**	4	3	1	
2	**6**	**7**	**8/2!=4**		
2	**6**	**7**			
2	**6**				

El polinomio sería

$P(x) = 2 + 3(x-1) + 1(x-1)^2 + 2(x-1)^2(x-2) - 1(x-1)^2(x-2)^2.$

□

Interpolación cúbica de Hermite a trozos

Dada la tabla de valores

x	x_0	x_1	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n
y'	y'_1	y'_2	$\cdots$	y'_n

correspondientes a una cierta función y sus derivadas, el método de interpolación cúbica de Hermite a trozos consiste en encontrar una única función derivable definida a trozos, tal que en cada intervalo $[x_i, x_{i+1}]$ $(i = 0, \ldots, n-1)$, la función es un polinomio $P_i(x)$ de grado menor o igual a 3 que verifica lo siguiente

$$P_i(x_i) = y_i \quad P_i(x_{i+1}) = y_{i+1} \quad P'(x_i) = y'_i \quad P'_i(x_{i+1}) = y'_{i+1}.$$

Planteando los correspondientes sistemas de ecuaciones, se puede encontrar la única función que interpola en esas condiciones.

Splines

Una función spline de grado k con nodos $x_0 \leq x_1 \leq \cdots \leq x_n$ es una función $S(x)$ definida a trozos tal que:

a) $S(x)$ es un polinomio $S_i(x)$ de grado menor o igual que k en cada intervalo $[x_i, x_{i+1}]$, con $i = 0, 1, \ldots, n-1$.

b) $S(x)$ admite derivada segunda continua de orden $k-1$ en $[x_0, x_n]$.

El spline cúbico ($k = 3$) es el más usado en interpolación. Un spline cúbico de interpolación en $\{x_0, x_1, \ldots, x_n\}$ es una función que cumple las propiedades $a), b)$ y $S(x_i) = f(x_i)$, con $i = 0, 1, \ldots, n$. El número de condiciones que se han de cumplir es el siguiente:

condición	ecuaciones	num. ec.
continuidad	$S_{i-1}(x_i) = S_i(x_i)$ con $i = 1, \ldots, n-1$	$n-1$
derivada 1^a continua	$S'_{i-1}(x_i) = S'_i(x_i)$ con $i = 1, \ldots, n-1$	$n-1$
derivada 2^a continua	$S''_{i-1}(x_i) = S''_i(x_i)$ con $i = 1, \ldots, n-1$	$n-1$
interpolación	$S_i(x_i) = y_i$ con $i = 0, \ldots, n$	$n+1$
	total	$4n-2$

Cada polinomio cúbico $S_i(x)$, con $i = 0, \ldots, n-1$, tendrá 4 parámetros, con lo que el número de incógnitas será $4n-4$. Por lo tanto, para poder resolver el correspondiente sistema de ecuaciones hemos de añadir dos condiciones más y podemos hacerlo de distintas formas:

- Si exigimos $S''(x_0) = 0$ y $S''(x_n) = 0$ obtendremos el spline cúbico natural.
- Si exigimos $S'(x_0) = S'(x_n)$ y $S''(x_0) = S''(x_n)$ obtendremos el spline periódico.

Se pueden elegir otras condiciones distintas del spline natural o del periódico. Sin embargo, los splines natural y periódico existen y son únicos para cualquier conjunto de nodos y de datos. Además, si la función a interpolar $f(x)$ tiene derivada segunda continua en $[x_0, x_n]$, entonces

$$\int_a^b [S''(x)]^2 dx \leq \int_a^b [f''(x)]^2 dx,$$

siendo $S(x)$ el spline natural. Es decir, de entre todas las funciones con derivada segunda continua que pasan por ciertos puntos $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$ el spline natural es el que tiene menor energía elástica (menor tensión[1]).

Por otra parte, si aumentamos el número de puntos a interpolar en el intervalo $[a, b]$ ($n \to \infty$) disminuyendo simultáneamente el tamaño máximo de los subintervalos $[x_i, x_{i+1}]$ ($\max_{0 \leq i \leq n}[x_i, x_{i+1}] \to 0$), entonces el spline natural y su derivada convergen uniformemente a $f(x)$ y $f'(x)$ respectivamente.

[1]La tensión o energía elástica en $[a, b]$ es proporcional a $\int_a^b [f''(x)]^2 dx$.

Capítulo 5

Diferenciación e integración numérica

5.1 Diferenciación numérica y extrapolación de Richarson

¿Puede calcularse la derivada de una función $f(x)$ en un punto si se conocen solo los valores de la función $f(x_0)$, $f(x_1)$,...,$f(x_n)$ en $n+1$ puntos? Si la función es un polinomio de grado n, entonces seguro que se puede. Si de la función solo sabemos que es derivable, entonces no será posible, pues hay muchas funciones derivables que pasan por los mismos puntos. El objetivo de esta sección es estimar el valor de $f'(x)$ a partir de los valores de la función en ciertos puntos $f(x_0)$, $f(x_1)$,...,$f(x_n)$ dando una cota del error cometido. Dado que

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

una primera estimación es

$$\boxed{f'(x) \approx \frac{f(x+h) - f(x)}{h}}$$

que es aplicable cuando conocemos los valores de la función en dos puntos: $x_0 = x$ y $x_1 = x+h$. La fórmula anterior es exacta para funciones lineales (por ejemplo, para $f(x) = 3x+2$) y en otros casos de manera fortuita. Calculemos

una cota del error. Por el teorema de Taylor:

$$f(x+h) = f(x) + f'(x)h + \frac{f''(\theta)}{2!}h^2$$

es decir

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \frac{h}{2}f''(\theta).$$

El error al aproximar $f'(x)$ por $f'(x) = \frac{f(x+h)-f(x)}{h}$, llamado error de truncamiento, es

$$\boxed{\text{error} = -\frac{h}{2}f''(\theta) \quad \text{(fórmula de orden h)}}$$

siendo θ un número entre x y $x+h$. Cuanto menor sea h, mejor será la aproximación de $f'(x)$, hasta que el error de redondeo impida la mejora.

Ejemplo: *Estima el valor de la derivada de* $\cos(x)$ *en el punto* $\frac{\pi}{4}$ *y calcula una cota del error cometido.*

Tomando como $h = 0.01$

$$f'(\frac{\pi}{4}) \approx \frac{f(\frac{\pi}{4}+h) - f(\frac{\pi}{4})}{h} = \frac{\cos(\frac{\pi}{4}+0.01) - \cos(\frac{\pi}{4})}{0.01} = -0.71063050.$$

La cota del error sería

$$|\text{error}| = |-\frac{h}{2}f''(\theta)| = |\frac{0.01}{2}\cos(\theta)| \leq \frac{0.01}{2} = 0.005.$$

□

Otra estimación de $f'(x)$ es

$$\boxed{f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}.}$$

Veamos la estimación del error. Del teorema de Taylor se deduce que

$$f(x+h) = f(x) + f'(x)h + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(\theta_1)$$
$$f(x-h) = f(x) - f'(x)h + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(\theta_2).$$

Si restamos las ecuaciones anteriores

$$f(x+h) - f(x-h) = 2f'(x)h + \frac{h^3}{6}[f'''(\theta_1) + f'''(\theta_2)].$$

Despejando

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{12}[f'''(\theta_1) + f'''(\theta_2)]$$

con lo que el error de truncamiento sería

$$-\frac{h^2}{12}[f'''(\theta_1) + f'''(\theta_2)]$$

siendo θ_1 un número entre x y $x+h$ y θ_2 un número entre x y $x-h$. Si $f'''(x)$ existe y es continua en $[x-h, x+h]$, entonces existe un número $\theta \in [x-h, x+h]$ tal que $f'''(\theta) = \frac{f'''(\theta_1)+f'''(\theta_2)}{2}$, con lo que error de truncamiento quedaría

$$\boxed{\text{error} = -\frac{h^2}{6}f'''(\theta) \quad \text{(fórmula de orden } h^2\text{)}.}$$

Para estimar el valor de una derivada segunda podemos usar la fórmula

$$\boxed{f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}.}$$

Veamos la estimación del error. Por el teorema de Taylor

$$\begin{array}{l} f(x+h) = f(x) + f'(x)h + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) + \frac{h^4}{24}f^{4)}(\theta_1) \\ f(x-h) = f(x) - f'(x)h + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(x) + \frac{h^4}{24}f^{4)}(\theta_2) \end{array}$$

.

Si sumamos las ecuaciones anteriores y despejamos

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} - \frac{h^2}{24}[f^{4)}(\theta_1) + f^{4)}(\theta_2)].$$

Siendo θ_1 y θ_2 números que están entre x y $x+h$ y entre x y $x-h$. Si $f^{4)}(x)$ existe y es continua, entonces existe un número $\theta \in [x-h, x+h]$ tal que

$$\boxed{\text{error} = \frac{h^2}{12}f^{4)}(\theta) \quad \text{(fórmula de orden } h^2\text{)}}$$

5.1.1 Diferenciación mediante interpolación

Dada una función $f(x)$ y recordando cómo se calcula el polinomio de interpolación en los puntos x_0, x_1, ...,x_n, se tiene que :

$$f(x) = \sum_{i=0}^{n} f(x_i)L_{x_i}(x) + \frac{1}{(n+1)!}f^{n+1)}(\theta_x)w(x)$$

donde $w(x) = \Pi_{i=0}^{n}(x - x_i)$. Si derivamos

$$f'(x) = \sum_{i=0}^{n}\left[f(x_i)L'_{x_i}(x) + \frac{1}{(n+1)!}f^{n+1)}(\theta_x)w'(x) + \frac{1}{(n+1)!}w(x)\frac{d}{dx}(f^{n+1)}(\theta_x))\right]$$

Si $x = x_k$

$$f'(x_k) = \sum_{i=0}^{n}\left[f(x_i)L'_{x_i}(x_k) + \frac{1}{(n+1)!}f^{n+1)}(\theta_{x_k})w'(x_k)\right].$$

Y como

$$w'(x) = \sum_{i=0}^{n}\Pi_{j=0,j\neq i}^{n}(x - x_j) \Rightarrow w'(x_k) = \Pi_{j=0,j\neq k}^{n}(x_k - x_j)$$

nos queda

$$\boxed{f'(x_k) = \sum_{i=0}^{n}\left[f(x_i)L'_{x_i}(x_k)\right] + \frac{1}{(n+1)!}f^{n+1)}(\theta_{x_k})\Pi_{j=0,j\neq k}^{n}(x_k - x_j)}$$

que es una forma de calcular un valor aproximado de $f'(x_k)$ a partir de los valores x_0, x_1, ...,x_n.

Ejemplo: *a) Aplica la fórmula anterior para calcular* $f'(x_1)$ *con* $n = 2$. *b) Calcula lo anterior con* $x_1 - x_0 = x_2 - x_1 = h$

Apartado a). Puesto que

$$L_{x_0} = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} \quad L_{x_1} = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}$$
$$L_{x_2} = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

se tiene que

$$L'_{x_0} = \frac{2x - x_1 - x_2}{(x_0 - x_1)(x_0 - x_2)} \; L'_{x_1} = \frac{2x - x_0 - x_2}{(x_1 - x_0)(x_1 - x_2)}$$
$$L'_{x_2} = \frac{2x - x_0 - x_1}{(x_2 - x_0)(x_2 - x_1)}.$$

Por tanto

$$f'(x_1) \approx f(x_0)\frac{x_1 - x_2}{(x_0 - x_1)(x_0 - x_2)} + f(x_1)\frac{2x_1 - x_0 - x_2}{(x_1 - x_0)(x_1 - x_2)}$$
$$+ f(x_2)\frac{x_1 - x_0}{(x_2 - x_0)(x_2 - x_1)}$$

siendo el error de esta aproximación

$$\frac{1}{3!}f^{3)}(\theta_{x_1})(x_1 - x_0)(x_1 - x_2).$$

Apartado b). Siendo $x_0 = x_1 - h$ y $x_2 = x_1 + h$ se obtiene la fórmula ya conocida

$$f'(x_1) \approx f(x_1 - h)\frac{-h}{(-h)(-2h)} + f(x_1)\frac{2x_1 - (x_1 - h) - (x_1 + h)}{(h)(-h)}$$
$$+ f(x_1 + h)\frac{h}{(2h)(h)}\frac{f(x_1 + h) - f(x_1 - h)}{2h},$$

siendo el error de esta aproximación

$$\frac{1}{3!}f^{3)}(\theta_{x_1})(h)(-h) = -\frac{1}{3!}f^{3)}(\theta_{x_1})h^2.$$

□

5.1.2 Extrapolación de Richarson

La extrapolación de Richarson sirve para generar resultados de gran exactitud cuando se usan fórmulas de bajo orden. La extrapolación puede aplicarse siempre a cualquier método de aproximación en el que sepamos el término de error de una forma previsible y se basa en un parámetro que generalmente es el tamaño de paso h.

Por el teorema de Taylor sabemos que

$$f(x+h) = \sum_{k=0}^{\infty} \frac{1}{k!} h^k f^{k)}(x) \quad \text{y} \quad f(x+h) = \sum_{k=0}^{\infty} \frac{1}{k!} (-1)^k h^k f^{k)}(x).$$

Restando las fórmulas anteriores

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{2}{3!} h^3 f^{3)}(x) + \frac{2}{5!} f^{5)}(x) + \dots$$

y despejando $f'(x)$

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \left[\frac{f^{3)}(x)}{3!} h^2 + \frac{f^{5)}(x)}{5!} h^4 + \frac{f^{7)}(x)}{7!} h^6 + \dots \right].$$

Si definimos $\Phi(h) = \frac{f(x+h)-f(x-h)}{2h}$ y denotamos a_i al coeficiente de h^i en la fórmula anterior, tenemos

$$f'(x) = \Phi(h) + a_2 h^2 + a_4 h^4 + a_6 h^6 + \dots \tag{5.1}$$

que es la fórmula ya conocida de aproximación de $f'(x)$

$$\boxed{f'(x) \approx \Phi(h) \quad \text{con error} = h^2 \left[a_2 + a_4 h^2 + a_6 h^4 + \dots \right]}$$

Si sustituimos h por $h/2$

$$f'(x) = \Phi(\frac{h}{2}) + \frac{a_2}{2^2} h^2 + \frac{a_4}{2^4} h^4 + \frac{a_6}{2^6} h^6 + \dots$$

y multiplicando por 4

$$4f'(x) = 4\Phi(\frac{h}{2}) + a_2 h^2 + \frac{a_4}{4} h^4 + \frac{a_6}{2^4} h^6 + \dots \tag{5.2}$$

con lo que (5.2)-(5.1) es

$$3f'(x) = 4\Phi(\frac{h}{2}) - \Phi(h) - \frac{3a_4}{4}h^4 - \frac{15a_6}{16}h^6$$

que, despejando $f'(x)$, da origen a la fórmula de aproximación

$$\boxed{f'(x) \approx \frac{4}{3}\Phi(\frac{h}{2}) - \frac{1}{3}\Phi(h) \quad \text{con error} = h^4\left[-\frac{a_4}{4} - \frac{5a_6}{16}h^2 - \dots\right]}$$

que también puede expresarse

$$f'(x) \approx \Phi(\frac{h}{2}) + \frac{\Phi(\frac{h}{2}) - \Phi(h)}{3} \quad \text{con error} = h^4\left[-\frac{a_4}{4} - \frac{5a_6}{16}h^2 - \dots\right]$$

Ejemplo: *Siendo* $f(x) = \arctan(x)$, *calcula* $f'(\sqrt{2})$ *haciendo uso de la extrapolación de Richarson.*

Tomamos $h = 0.1$.

$$\begin{aligned} f'(x) \approx \Phi(h) &= \frac{f(x+h) - f(x-h)}{2h} \\ &= \frac{f(\sqrt{2}+0.1) - f(\sqrt{2}-0.1)}{2 \cdot 0.1} = 0.3339506968 \end{aligned}$$

y

$$f'(x) \approx \Phi(h/2) = 0.3334876594$$

Usando la extrapolación

$$f'(x) \approx \Phi(h/2) + \frac{\Phi(h/2) - \Phi(h)}{3} = 0.3333333136$$

(El valor exacto es 1/3).

□

Si en la fórmula (5.1.2) denotamos $\Psi(h) = \frac{4}{3}\Phi(\frac{h}{2}) - \frac{1}{3}\Phi(h)$, $b_4 = \frac{-a_4}{4}$, $b_6 = \frac{-5a_6}{16}$, ... obtenemos la expresión

$$f'(x) = \Psi(h) + b_4h^4 + b_6h^6 + \dots \tag{5.3}$$

Sustituyendo h por $h/2$

$$f'(x) = \Psi(h/2) + b_4\left(\frac{h}{2}\right)^4 + b_6\left(\frac{h}{2}\right)^6 + \dots$$

y multiplicando por 2^4

$$16f'(x) = 16\Psi(h/2) + b_4h^4 + b_6\frac{h^6}{4} + \dots \tag{5.4}$$

Restando (5.4)-(5.3) nos queda

$$15f'(x) = 16\Psi(h/2) - \Psi(h) - 3b_6\frac{h^6}{4} + \dots$$

y al despejar $f'(x)$

$$f'(x) = \frac{16}{15}\Psi(h/2) - \frac{1}{15}\Psi(h) - b_6\frac{h^6}{20} + \dots$$

da origen a la fórmula

$$\boxed{f'(x) \approx \frac{16}{15}\Psi(h/2) - \frac{1}{15}\Psi(h) \quad \text{con error } = h^6\left[-b_6\frac{1}{20} + \dots\right]}$$

que también puede expresarse

$$f'(x) \approx \Psi(h/2) + \frac{\Psi(h/2) - \Psi(h)}{15} \quad \text{con error } = h^6\left[-b_6\frac{1}{20} + \dots\right]$$

El proceso puede continuar de forma similar tomando ahora otra función $\Omega(h) = \frac{16}{15}\Psi(h/2) - \frac{1}{15}\Psi(h)$. Se puede demostrar el siguiente resultado:

Teorema: *El algoritmo de Richarson genera aproximaciones a $f'(x)$ cada vez de un orden superior.*

Ejemplo: *Aplica dos pasos del algoritmo de Richarson para calcular* $f'(\sqrt{2})$ *siendo* $f(x) = \arctan(x)$ *y tomando* $h = 0.1$.

Aplicamos la fórmula (5.1.2) con h y $h/2$ para obtener las dos primeras aproximaciones de orden 2

$$\Phi(0.1) = \frac{\arctan(\sqrt{2}+0.1) - \arctan(\sqrt{2}-0.1)}{2 \cdot 0.1} = 0.33395069677432$$

$$\Phi(0.05) = \frac{\arctan(\sqrt{2}+0.05) - \arctan(\sqrt{2}-0.05)}{2 \cdot 0.05} = 0.33348765942111$$

A partir de esos dos valores y aplicando la fórmula (5.1.2) calculamos la aproximación $\Psi(h)$ que es de orden 4

$$\Psi(0.1) = \Phi(0.05) + \frac{\Phi(0.05) - \Phi(0.1))}{3} = 0.33333331363671$$

Calculando $\Phi(h/4)$ ($\Phi(0.025) = 0.33337191390106$) podremos también calcular $\Psi(h/2)$

$$\Psi(0.05) = \Phi(0.025) + \frac{\Phi(0.025) - \Phi(0.05)}{3} = 0.33333333206104$$

y a partir de estos dos valores obtenemos la aproximación $\Omega(h)$

$$\Omega(0.1) = \Psi(0.05) + \frac{\Psi(0.05) - \Psi(0.1)}{15} = 0.33333333328933.$$

□

En general, el proceso de extrapolación de Richarson puede resumirse así:

Etapa 0. Calculo distintas aproximaciones $R(0,h)$ en función de un parámetro h (en nuestro caso $R(0,h) = \Phi(h) = \frac{f(x+h)-f(x-h)}{2h}$).

Etapa 1. A partir de $R(0,h)$ y de $R(0,h/2)$ calculo las aproximaciones

$$R(1,h) = R(0,h/2) + \frac{R(0,h/2) - R(0,h)}{4^1 - 1}.$$

Etapa 2. A partir de $R(1,h)$ y de $R(1,h/2)$ calculo las aproximaciones

$$R(2,h) = R(1,h/2) + \frac{R(1,h/2) - R(1,h)}{4^2 - 1}.$$

Etapa n. A partir de $R(n-1,h)$ y de $R(n-1,h/2)$ calculo las aproximaciones

$$R(n,h) = R(n-1,h/2) + \frac{R(n-1,h/2) - R(n-1,h)}{4^n - 1}.$$

Etapa 0	Etapa 1	Etapa 2	Etapa 3	...
$R(0,h)$				
$R(0,h/2)$	$R(1,h)$			
$R(0,h/4)$	$R(1,h/2)$	$R(2,h)$		
$R(0,h/8)$	$R(1,h/4)$	$R(2,h/2)$	$R(3,h)$	
...	...	...	...	...

5.2 Integración numérica mediante interpolación

El objetivo es obtener un valor aproximado de $\int_a^b f(x)dx$ sin calcular la primitiva de $f(x)$. La idea es sustituir $\int_a^b f(x)dx$ por $\int_a^b p(x)dx$, siendo $p(x)$ una función que se parezca a $f(x)$, por ejemplo un polinomio interpolador en unos nodos $x_0, x_1, \ldots, x_n$ del intervalo $[a,b]$. Si utilizamos el método de Lagrange para calcular el polinomio interpolador

$$p(x) = \sum_{i=0}^{n} f(x_i) l_{x_i}(x) \quad \text{con} \quad l_{x_i}(x) = \prod_{j=0}^{n} \frac{x - x_j}{x_i - x_j},$$

tendríamos que

$$\int_a^b f(x)dx \approx \int_a^b p(x)dx = \sum_{i=0}^{n} f(x_i) \int_a^b l_{x_i}(x)$$

es decir, obtenemos una fórmula como esta

$$\boxed{\int_a^b f(x)dx \approx \sum_{i=0}^{n} A_i f(x_i).} \tag{5.5}$$

En el ejemplo anterior $A_i = \int_a^b l_{x_i}(x)dx$. Una expresión como (5.5) se denomina *fórmula de cuadratura* y los puntos x_i son los *nodos de cuadratura*. Las fórmulas así construidas son exactas para los polinomios de grado menor o igual que n (*¿por qué?*). Si en el ejemplo utilizado los nodos están igualmente espaciados, estaremos hablando del "*método de Newton-Côtes*".

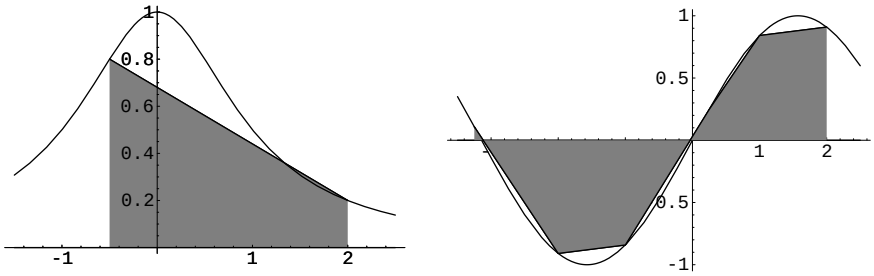


Figura 5.1: fórmulas de cuadratura gráficamente

5.2.1 Regla del trapecio

El caso particular del método de Newton-Côtes con $n = 1$ se conoce como el *método del trapecio*. Puesto que $l_{x_0}(x) = \frac{b-x}{a-x}$, $l_{x_1}(x) = \frac{x-a}{b-a}$, $A_0 = \int_a^b l_{x_0}(x)dx = \frac{1}{2}(b-a)$ y $A_0 = \int_a^b l_{x_1}(x)dx = \frac{1}{2}(b-a)$, la fórmula de cuadratura queda así

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n} A_i f(x_i) = \frac{b-a}{2}\left[f(a) + f(b)\right]$$

La expresión del error del polinomio de interpolación $p(x)$ para $n = 1$ es $f(x) - p(x) = \frac{f''(\xi_x)}{2}(x-a)(x-b)$, de modo que

$$\int_a^b f(x)dx - \int_a^b p(x)dx = \int_a^b f(x) - p(x)dx = \int_a^b \frac{f''(\xi_x)}{2}(x-a)(x-b)dx$$

Puesto que $(x-a)(x-b)$ no cambia de signo en todo $[a,b]$ y suponiendo que $f''(x)/2$ es una función continua en $[a,b]$, aplicando la generalización del teorema de valor medio del cálculo integral [1] obtenemos que el error es

$$\frac{f''(\xi)}{2}\int_a^b (x-a)(x-b)dx = \frac{f''(\xi)}{2}\frac{(a-b)^3}{6}$$

[1] **Teorema**: *Sean $h(x)$ y $g(x)$ dos funciones reales e integrables en un intervalo $[a,b]$. Si $g(x)$ no cambia de signo en todo $x \in [a,b]$ y $h(x)$ es continua en $[a,b]$, entonces existe un punto $\xi \in [a,b]$ tal que:*

$$\int_a^b h(x)g(x)dx = h(\xi)\int_a^b g(x)dx,$$

.

para un cierto valor $\xi \in (a,b)$. En resumen, si $h = b - a$, la **Regla del Trapecio** quedaría así:

$$\boxed{\int_a^b f(x)dx \approx \frac{h}{2}\left[f(a)+f(b)\right] \quad \text{con error}= -f''(\xi)\frac{h^3}{12}} \tag{5.6}$$

para un cierto $\xi \in [a,b]$.

Ejemplo: *Calcula* $\int_0^{0.2} e^{x^2}dx$ *con la fórmula de Newton-Côtes con* $n = 1$ *(regla del trapecio).*

$$\int_0^{0.2} e^{x^2}dx \approx \frac{0.2-0}{2}[e^0 + e^{0.2}] = 0.20408107741924$$

Sabiendo que si $x \in [0, 0.2]$ entonces

$$|f''(x)| = |2e^{x^2}+4x^2e^{x^2}| \leq |f''(0.2)| = |2e^{0.2^2}+4\cdot 0.2e^{0.2^2}| = 2.248151272255559,$$

una cota del error será

$$|-2.248151272255559\frac{0.2^3}{12}| = 0.00149876751484.$$

□

5.2.2 Regla de Simpson

La fórmula de Newton-Côtes para $n = 2$ o $n = 3$ se llama Regla de Simpson. El caso $n = 2$, $h = \frac{b-a}{2}$ y $x_i = x_0 + ih$, se conoce como la **regla de Simpson 1/3**:

$$\boxed{\int_a^b f(x)dx \approx \frac{h}{3}\left[f(x_0)+4f(x_1)+f(x_2)\right] \quad \text{con error} - f^{4)}(\xi)\frac{h^5}{90}}$$

para un cierto $\xi \in [a,b]$. Por ser una fórmula de Newton-Côtes con tres puntos es exacta para polinomios de grado menor o igual que 2; pero además es exacta para los polinomios de grado menor o igual que 3, ya que el término del error contiene el factor $f^{4)}(\xi)$.

El caso $n = 3$, $h = \frac{b-a}{3}$ y $x_i = x_0 + ih$, se conoce como la **regla de Simpson 3/8**:

$$\boxed{\int_a^b f(x)dx \approx \frac{3h}{8}\left[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)\right] \quad \text{con error} - f^{4)}(\xi)\frac{3h^5}{80}}$$

para un cierto $\xi \in [a, b]$. Es mejor usar la regla de Simpson 1/3 que la de 3/8, puesto que ambas son de orden 5 y $|-1/90| < |-3/80|$. En general es preferible usar una regla con n par (número impar de puntos) que con n impar.

Ejemplo: *Calcula* $\int_0^{0.2} e^{x^2}dx$ *con la fórmula de Newton-Côtes con* $n = 2$ *(regla de Simpson 1/3).*

$$\int_0^{0.2} e^{x^2}dx \approx \frac{0.2/2}{3}[e^0 + e^{0.1^2} + e^{0.2^2}] = 0.2027$$

Sabiendo que si $x \in [0, 0.2]$ entonces

$$\begin{aligned}|f^{4)}(x)| = |12e^{x^2} + 48e^{x^2}x^2 + 16e^{x^2}x^4| &\leq |f^{4)}(0.2)| \\ &= |12e^{0.2^2} + 48e^{0.2^2}0.2^2 + 16e^{0.2^2}0.2^4| = 14.5147,\end{aligned}$$

una cota del error será

$$|-14.5147\frac{(0.2/2)^5}{90}| = 1.6 * 10^{-6}$$

□

5.2.3 Reglas compuestas

Este método consiste en dividir el intervalo $[a, b]$ en subintervalos y aplicar en cada uno de ellos una regla simple. Por ejemplo, la regla del trapecio compuesta consiste en fijar unos puntos $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ y aplicar la regla del trapecio a cada intervalo $[x_{i-1}, x_i]$:

$$\int_a^b f(x)dx = \sum_{i=1}^{n}\int_{x_{i-1}}^{x_i} f(x)dx \approx \frac{1}{2}\sum_{i=1}^{n}(x_i - x_{i-1})\left[f(x_{i-1}) + f(x_i)\right].$$

Es equivalente a sustituir $f(x)$ por una función que interpole a $f(x)$ y que esté formada por trozos de línea. Si los puntos están igualmente espaciados

($h = x_i - x_{i-1} = \frac{b-a}{n}$, $x_i = a + ih$ con $i = 1, 2, \ldots, n$), la **regla del trapecio compuesta** nos queda

$$\int_a^b f(x)dx \approx \frac{h}{2}[f(x_0) + 2f(x_1) + \cdots + 2f(x_{n-1}) + f(x_n)]$$
$$\text{con error } -\frac{h^2}{12}(b-a)f''(\xi)$$

para un cierto $\xi \in [a, b]$. La fórmula es exacta para polinomios de grado menor o igual que 1.

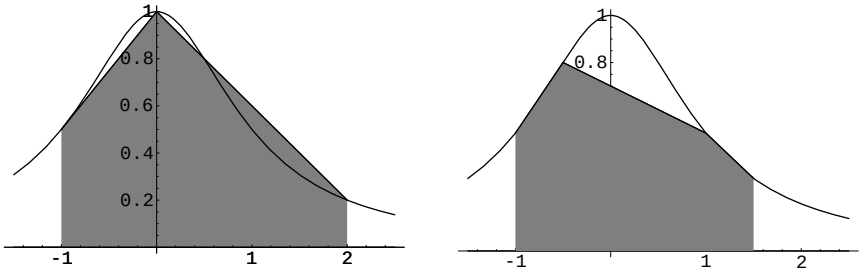


Figura 5.2: Regla del trapecio compuesta con 3 y 4 puntos

Ejemplo: *Calcula un valor aproximado de* $\int_0^{0.2} e^{x^2} dx$ *mediante la regla del trapecio compuesta.*

Si tomamos tres puntos $x_0 = 0$, $x_1 = 0.1$ y $x_2 = 0.2$

$$\int_0^{0.2} e^{x^2} dx \approx \frac{0.1}{2}\left[e^0 + 2e^{0.1^2} + e^{0.2^2}\right] = 0.201085$$

Una cota del error sería

$$\left|-\frac{0.1^2}{12}(0.2-0)(e^{x^2}(2+4e^{x^2}))\right| \leq \left|-\frac{0.1^2}{12}0.2 \cdot f''(0.2)\right| = 0.00187346$$

□

Si el número n de intervalos en los que dividimos $[a, b]$ es par podemos aplicar de forma sencilla la regla de Simpson compuesta con $h = \frac{b-a}{n}$, $x_i = a + ih$ $(0 \leq i \leq n)$, aplicando a cada dos subintervalos la regla de Simpson 1/3

$$\int_a^b f(x)dx = \int_{x_0}^{x_2} f(x)dx + \int_{x_2}^{x_4} f(x)dx + \cdots + \int_{x_{n-2}}^{x_n} f(x)dx$$

con lo que obtenemos la fórmula

$$\int_a^b f(x) \approx \frac{h}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 4f(x_{n-1}) + f(x_n)]$$

o lo que es lo mismo

$$\int_a^b f(x)dx \approx f(x_0) + \sum_{i=1}^{\frac{n}{2}}[2f(x_{2i}) + 4f(x_{2i-1})] + f(x_n)$$

$$\text{con error} - \frac{b-a}{180}h^4 f^{4)}(\xi)$$

con $\xi \in [a, b]$.

5.2.4 Método de los coeficientes indeterminados

Las reglas de Newton-Côtes con $n+1$ puntos son exactas para los polinomios de grado n, es decir, si en la fórmula de cuadratura $\int_a^b f(x)dx \approx \sum_{i=0}^n A_i f(x_i)$ tomamos los coeficientes $A_i = \int_a^b l_{x_i}(x)dx$, el resultado que obtenemos al calcular con esa regla la integral de un polinomio de grado n es el valor exacto. El objetivo de esta sección es, dados un conjunto de $n+1$ puntos x_0, x_1, …, x_n, encontrar el valor de los coeficientes A_0, A_1, …, A_n que aseguren que la fórmula de cuadratura $\int_a^b f(x)dx \approx \sum_{i=0}^n A_i f(x_i)$ es exacta para los polinomios de grado n. En cada caso, el problema se puede resolver de forma similar al siguiente ejemplo.

Ejemplo: *Encuentra los valores A_0, A_1 y A_2 para que la fórmula*

$$\int_0^1 f(x)dx \approx A_0 f(0) + A_1 f(\frac{1}{2}) + A_2 f(1)$$

sea exacta para los polinomios de grado menor o igual que 2. En particular

la fórmula tendrá que ser exacta para los polinomios 1, x y x^2 que forma una base del conjunto de los polinomios de grado menor o igual que 2:

$$\begin{array}{ccccc} 1 & = & \int_0^1 1dx & = & A_0 + 1A_1 + A_2 \\ \frac{1}{2} & = & \int_0^1 xdx & = & 0A_0 + \frac{1}{2}A_1 + A_2 \\ \frac{1}{3} & = & \int_0^1 x^2dx & = & 0A_0 + \frac{1}{4}A_1 + A_2 \end{array}$$

y obtenemos que $A_0 = \frac{1}{6}$, $A_1 = \frac{2}{3}$ y $A_2 = \frac{1}{6}$. Además, si $P(x) = a + bx + cx^2$ es cualquier otro polinomio de grado menor o igual que 2, la fórmula sería también exacta porque

$$\begin{aligned} \int_0^1 P(x)dx &= a\int_0^1 1dx + b\int_0^1 xdx + c\int_0^1 x^2dx = \\ &= a(A_0 + 1A_1 + A_2) + b(\frac{1}{2}A_1 + A_2) + c(\frac{1}{4}A_1 + A_2) = \\ &= A_0(a) + A_1(a + \frac{1}{2}b + (\frac{1}{2})^2c) + A_2(a + 1b + 1^2c) = \\ &= A_0P(0) + A_1P(\frac{1}{2}) + A_2P(1) \end{aligned}$$

que sería el resultado obtenido al aplicar la fórmula de cuadratura con los coeficientes calculados. Además, es la misma fórmula que se obtendría con el método de Simpson.

□

5.3 Cuadratura gaussiana

Supongamos que queremos obtener una fórmula de cuadratura

$$\int_{-1}^1 f(x)dx = A_1 f(t_1) + A_2 f(t_2)$$

pero no fijamos previamente los valores de t_1, t_2, A_1 y A_2. Para calcular estos valores y puesto que tenemos 4 incógnitas, le imponemos 4 condiciones: que

la fórmula sea exacta para los polinomios de grado menor o igual que 3, es decir, para 1, x, x^2 y x^3 :

$$\begin{array}{l} \int_{-1}^{1} x^3 dx = 0 = A_1 t_1^3 + A_2 t_2^3 \\ \int_{-1}^{1} x^2 dx = \frac{2}{3} = A_1 t_1^2 + A_2 t_2^2 \\ \int_{-1}^{1} x dx = 0 = A_1 t_1 + A_2 t_2 \\ \int_{-1}^{1} 1 dx = 2 = A_1 + A_2 \end{array}$$

Restando a la ecuación primera la tercera multiplicada por t_1^2 obtenemos

$$0 = A_2(t_2^3 - t_1^2 t_2) = A_2 t_2 (t_2 - t_1)(t_2 + t_1)$$

con lo que se obtiene alguno de los siguientes valores: $A_2 = 0$, $t_2 = 0$, $t_2 = t_1$, $t_2 = -t_1$. Sustituyendo en el resto de las ecuaciones podemos comprobar que solo es aceptable $t_2 = -t_1$ y que en ese caso los parámetros son $A_1 = A_2 = 1$ y $t_2 = -t_1 = \sqrt{\frac{1}{3}}$. Por tanto la fórmula

$$\int_{-1}^{1} f(x)dx = f(-\sqrt{\frac{1}{3}}) + f(\sqrt{\frac{1}{3}})$$

es exacta para integrar cualquier polinomio de grado menor o igual que 3 en el intervalo $[-1, 1]$. Para integrarlos en $[a, b]$ se utiliza una transformación afín que lleve un intervalo en el otro.

Los nodos t_i $(i = 1, 2, \ldots, n)$ en la cuadratura gaussiana coinciden con las raíces del polinomio de Legendre de grado n. Los polinomios de Legendre se construyen así:

$$L_0(x) = 1 \qquad L_1(x) = x \qquad L_{n+1} = \frac{(2n+1)xL_n(x) - nL_{n-1}(x)}{n+1}.$$

Ejemplo: *Calcula un valor aproximado de* $\int_0^{\frac{\pi}{2}} \text{sen}(x)dx$ *mediante cuadratura gaussiana de dos términos.*

Como el intervalo de integración es $[0, \frac{\pi}{2}]$, hemos de hacer un cambio de variable afín que lleve $[-1, 1]$ en $[0, \frac{\pi}{2}]$. Dicho cambio es $x = \frac{\pi}{4}t + \frac{\pi}{4}$ con lo que $dx = \frac{\pi}{4}dt$. Ya hemos deducido cuales son los nodos t_i y los coeficientes A_i $(i = 1, 2)$. Si no supiéramos cuales son, podríamos averiguar primero los nodos

calculando las raíces del polinomio de Legendre de grado 2 ($L_2 = \frac{3x^2-x}{2}$). Posteriormente calcularíamos los coeficientes A_i planteando un sistema de ecuaciones. En todo caso se obtendría

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \text{sen}(x)dx &= \int_{-1}^{1} \text{sen}\left(\frac{\pi}{4}t + \frac{\pi}{4}\right)\frac{\pi}{4}dt = \\ &= \frac{\pi}{4}\left[\text{sen}\left(\frac{\pi}{4}(-\sqrt{\frac{1}{3}}) + \frac{\pi}{4}\right) + \text{sen}\left(\frac{\pi}{4}(\sqrt{\frac{1}{3}}) + \frac{\pi}{4}\right)\right] \approx 0.99847.\end{aligned}$$

□

5.4 Integración de Romberg

Para aplicar el método de integración de Romberg construimos una tabla como esta

$R(0,0)$					
$R(1,0)$	$R(1,1)$				
$R(2,0)$	$R(2,1)$	**R(2,2)**			
$R(3,0)$	$R(3,1)$	**R(3,2)**	**R(3,3)**		
...					
$R(M,0)$	$R(M,1)$	$R(M,2)$	$R(M,3)$	...	$R(M,M)$

donde, si queremos aproximar $\int_a^b f(x)dx$

$$\begin{aligned}R(0,0) &= \frac{1}{2}(b-a)\left[f(a)+f(b)\right] \\ &\quad \text{(Regla del trapecio para } 2^0 \text{ intervalos)} \\ R(n,0) &= \frac{1}{2}R(n-1,0) + \frac{b-a}{2^n}\sum_{i=1}^{2^{n-1}} f(a+(2i-1)\frac{b-a}{2^n}) \\ &\quad \text{(Regla del trapecio para } 2^n \text{ intervalos)} \\ R(n,m) &= R(n,m-1) + \frac{1}{4^m-1}\left[R(n,m-1) - R(n-1,m-1)\right] \\ &\quad \text{(Extrapolación de Richarson)}\end{aligned} \tag{5.7}$$

La última fórmula se obtiene aplicando la extrapolación de Richarson que es un proceso que ya hemos utilizado en otra ocasión. El factor $\frac{1}{4^m-1}$ depende

del orden de convergencia del método empleado, en este caso la regla del trapecio.

Cada valor $R(i,j)$ se considera una aproximación a $\int_a^b f(x)dx$. A partir de la aproximación $R(n,m-1)$ que se considera que mejora a la anterior $R(n-1,m-1)$, se construye otra mejor aproximación $R(n,m)$:

$$R(n,m) = (\text{+exacta}) + \frac{1}{4^m - 1}\left[(\text{+exacta}) - (\text{-exacta})\right].$$

Teorema: *Si $f(x)$ es una función continua en $[a,b]$ entonces la sucesión que se obtiene en cada columna del método de Romberg converge a $\int_a^b f(x)dx$.*

Ejemplo: *Aplica el método de Romberg para obtener una aproximación de $\int_0^2 e^{x^2}dx$.*

$R(0,0)$ es la regla del trapecio con $2^0 = 1$ intervalo:

$$R(0,0) = \frac{2-0}{2}(e^0 + e^{2^2}) \approx 55.5981$$

$R(1,0)$ es la regla del trapecio compuesta con 2^1 intervalos:

$$R(1,0) = \frac{1}{2}\left[e^0 + 2e^1 + e^2\right] = 30.5174$$

$R(2,0)$ es la regla del trapecio compuesta con 2^2 intervalos:

$$R(2,0) = \frac{2/4}{2}\left[e^0 + 2e^{0.5} + 2e^1 + 2e^{1.5} + e^2\right] = 20.6446$$

y de modo similar los términos $R(n,0)$ de la primera columna. El resto de términos se calculan según la fórmula 5.7 La mejor aproximación obtenida sería $\int_0^2 e^{x^2}dx \approx 16.4529$.

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	55.5981				
$n=1$	30.5174	22.1572			
$n=2$	20.6446	17.3537	17.0334		
$n=3$	17.5651	16.5386	16.4843	16.4756	
$n=4$	16.7354	16.4588	16.4535	16.4530	16.4529

□

Ejemplo: *Aplica el método de Romberg para obtener una aproximación de* $\int_0^2 x^2 dx$.

Obtenemos la tabla siguiente en donde a mejor aproximación obtenida sería $\int_0^2 x^2 dx \approx 2.6667$.

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	4				
$n=1$	3	2.6667			
$n=2$	2.75	2.6667	2.6667		
$n=3$	2.6875	2.6667	2.6667	2.6667	
$n=4$	2.6719	2.6667	2.6667	2.6667	2.6667

□

Ejemplo: *Aplica el método de Romberg para obtener una aproximación de* $\int_{0.2}^{0.5} e^{-x^2} dx$.

Obtenemos la tabla siguiente

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	0.69302				
$n=1$	0.66211	0.65181			
$n=2$	0.65947	0.65859	0.65904		
$n=3$	0.65898	0.65882	0.65883	0.65883	
$n=4$	0.65886	0.65882	0.65882	0.65882	0.65882

La mejor aproximación obtenida sería $\int_{0.2}^{0.5} e^{-x^2} dx \approx 0.65882$.

□

5.5 Cuadratura adaptativa

A grandes rasgos, el método consiste en calcular un valor aproximado de una integral $\int_a^b f(x)dx$ dividiendo el intervalo $[a, b]$ en intervalos disjuntos $[x_i, x_{i+1}]$ con las siguientes condiciones:

- Se fija un parámetro llamado tolerancia T.
- Se obtienen dos estimaciones de $\int_{x_i}^{x_{i+1}} f(x)dx$ mediante una fórmula de cuadratura.
- Se fija la tolerancia en el intervalo $[x_i, x_{i+1}]$:

$$\mathrm{T}_{[}\, x_i, x_{i+1}] = \frac{\mathrm{T}}{(b-a)}(x_{i+1} - x_i).$$

- Si la diferencia entre las dos estimaciones de $\int_{x_i}^{x_{i+1}} f(x)dx$ es menor que $\mathrm{T}_{[x_i,x_{i+1}]}$ se estima definitivamente $\int_{x_i}^{x_{i+1}} f(x)dx$ por extrapolación. En caso contrario se divide el intervalo en dos subintervalos disjuntos.
- Se estima $\int_a^b f(x)dx$:

$$\int_a^b f(x)dx = \sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} f(x)dx$$

Ejemplo: *Fijado un valor de tolerancia* $\mathrm{T} = 0.02$, *calcula* $\int_{0.2}^{1} \frac{1}{x^2}dx$ *usando la regla de Simpson.*

Paso 1. Aplicamos la regla de Simpson para calcular la integral:

$$h_1 = \frac{1 - 0.2}{2} = 0.4 \Rightarrow S_1[0.2, 1] = 4.94814815.$$

Paso 2. Aplicamos la regla de Simpson compuesta en dos subintervalos de $[0.2, 1]$, en concreto en $[0.2, 0.6]$ y en $[0.6, 1]$:

$$h_2 = \frac{0.6 - 0.2}{2} = 0.2 \quad \Rightarrow \quad S_2[0.2, 0.6] = 3.51851852.$$
$$h_2 = \frac{1 - 0.6}{2} = 0.2 \quad \Rightarrow \quad S_2[0.6, 1] = 0.66851852.$$

Paso 3. Hasta ahora tenemos dos aproximaciones a $\int_{0.2}^{1} \frac{1}{x^2} dx$, a saber: $S_1[0.2, 1]$ y $S_2[0.2, 0.6] + S_2[0.6, 1]$. Calculo la diferencia entre ambas estimaciones:

$$S_1[0.2, 1] - (S_2[0.2, 0.6] + S_2[0.6, 1]) = 0.7611111 > \mathrm{T} = 0.02.$$

Si la diferencia hubiera sido menor que T, el proceso finalizaría efectuando una extrapolación del tipo:

$$(+\text{exacta}) + \frac{1}{15}\left[(+\text{exacta}) - (\text{-exacta})\right]$$

entendiendo que la primera estimación es menos exacta que la segunda. Puesto que la diferencia es mayor que T, el proceso continúa fijando la tolerancia de forma proporcional al tamaño de cada subintervalo y de modo que sume la tolerancia total T = 0.02. Con estas condiciones la tolerancia será 0.01 en $[0.2, 0.6]$ y 0.01 en $[0.6, 1]$.

Paso 4. Estudiamos el intervalo $[0.6, 1]$. Tenemos la estimación $S_2[0.6, 1]$ de $\int_{0.6}^{1} \frac{1}{x^2} dx$ y obtenemos otra sumando los resultados obtenidos al aplicar la regla de Simpson en los intervalos $[0.6, 0.8]$ ($S_3[0.6, 0.8] = 0.41678477$) y en $[0.8, 1]$ ($S_3[0.6, 1] = 0.25002572$). Comparamos ambas estimaciones:

$$S_2[0.6, 1] - (S_3[0.6, 0.8] + S_3[0.8, 1]) = 0.66851852 - 0.66681049 = 0.001708 < \mathrm{T}_{[0.6,1]} = 0.01$$

Como la diferencia es menor que la tolerancia extrapolando obtenemos la estimación definitiva de $\int_{0.6}^{1} \frac{1}{x^2} dx$:

$$\begin{aligned} \int_{0.6}^{1} \frac{1}{x^2} dx &\approx (+\text{exacta}) + \frac{1}{15}\left[(+\text{exacta}) - (\text{-exacta})\right] \\ & 0.66681049 + \frac{1}{15}(0.66681049 - 0.66851852) = \mathbf{0.66669662} \end{aligned}$$

Paso 5. Actuamos ahora sobre el intervalo $[0.2, 0.6]$ donde la tolerancia también es 0.01. Contamos con una estimación $S_2[0.2, 0.6]$ de la integral en el intervalo y aplicamos la regla de Simpson en $[0.2, 0.4]$ ($S_3[0.2, 0.4] = 2.52314815$)y en $[0.4, 0.6]$ ($S_3[0.4, 0.6] = 0.83435926$) para obtener una nueva ($S_3[0.2, 0.4] + S_3[0.4, 0.6]$). Comparamos ambas estimaciones:

$$S_2[0.2, 0.6] - (S_3[0.2, 0.4] + S_3[0.4, 0.6]) = 0.161111 > \mathrm{T}_{[0.2,0.6]} = 0.01.$$

Puesto que la diferencia es mayor que la tolerancia seguimos el proceso y dividimos el intervalo $[0.2, 0.6]$ en los intervalos $[0.2, 0.4]$ y $[0.4, 0.6]$. Fijamos la tolerancia de cada intervalo de forma proporcional al tamaño del intervalo correspondiente y de modo que la suma sea igual a la tolerancia en el intervalo [0.2,0.6] ($\mathrm{T}_{[0.2,0.6]} = 0.01$). Con estas condiciones $\mathrm{T}_{[0.2,0.4]} = 0.005$ y $\mathrm{T}_{[0.4,0.6]} = 0.005$.

Paso 6. Estudiamos el intervalo $[0.4, 0.6]$. Tenemos la estimación $S_3[0.4, 0.6]$ de $\int_{0.4}^{0.6} \frac{1}{x^2} dx$ y obtenemos otra sumando los resultados obtenidos al aplicar la regla de Simpson en los intervalos $[0.5, 0.6]$ ($S_4[0.4, 0.5] = 0.50005144$) y en $[0.5, 0.6]$ ($S_4[0.5, 0.6] = 0.33334864$). Comparamos ambas estimaciones:

$$S_3[0.4, 0.6] - (S_4[0.4, 0.5] + S_4[0.5, 0.6]) = 0.000859 < \mathrm{T}_{[0.4,0.6]} = 0.005$$

Como la diferencia es menor que la tolerancia extrapolando obtenemos la estimación definitiva de $\int_{0.4}^{0.6} \frac{1}{x^2} dx$:

$$\int_{0.4}^{0.6} \frac{1}{x^2} dx \approx (+\text{exacta}) + \frac{1}{15} [(+\text{exacta}) - (\text{-exacta})] = \mathbf{0.8333428}.$$

Paso 7. Estudiamos el intervalo $[0.2, 0.4]$. Tenemos la estimación $S_3[0.2, 0.4]$ de $\int_{0.2}^{0.4} \frac{1}{x^2} dx$ y obtenemos otra sumando los resultados obtenidos al aplicar la regla de Simpson en los intervalos $[0.2, 0.3]$ y $[0.3, 0.4]$. Comparamos ambas estimaciones:

$$S_3[0.2, 0.4] - (S_4[0.2, 0.3] + S_4[0.3, 0.4]) > \mathrm{T}_{[0.2,0.4]} = 0.005.$$

Puesto que la diferencia es mayor que la tolerancia seguimos el proceso y dividimos el intervalo $[0.2, 0.4]$ en los intervalos $[0.2, 0.3]$ y $[0.3, 0.4]$. Fijamos la tolerancia de cada intervalo de forma proporcional al tamaño del intervalo correspondiente y de modo que la suma sea igual a la tolerancia en el intervalo [0.2,0.4] ($\mathrm{T}_{[0.2,0.4]} = 0.005$). Con estas condiciones $\mathrm{T}_{[0.2,0.3]} = 0.0025$ y $\mathrm{T}_{[0.3,0.4]} = 0.0025$.

Paso 8. Estudiamos el intervalo $[0.3, 0.4]$. Tenemos la estimación $S_4[0.3, 0.4]$ de $\int_{0.3}^{0.4} \frac{1}{x^2} dx$ y obtenemos otra sumando los resultados obtenidos al aplicar la regla de Simpson en los intervalos $[0.3, 0.35]$ y $[0.35, 0.4]$. Comparamos ambas estimaciones:

$$S_4[0.3, 0.4] - (S_5[0.3, 0.35] + S_5[0.35, 0.4]) = 0.0002220 < \mathrm{T}_{[0.3,0.4]} = 0.0025.$$

Como la diferencia es menor que la tolerancia extrapolando obtenemos la estimación definitiva de $\int_{0.3}^{0.4} \frac{1}{x^2} dx$:

$$\int_{0.3}^{0.4} \frac{1}{x^2} dx \approx (+\text{exacta}) + \frac{1}{15} [(+\text{exacta}) - (\text{-exacta})] = \mathbf{0.83333492}.$$

Paso 9. Estudiamos el intervalo $[0.2, 0.3]$. Tenemos la estimación $S_4[0.2, 0.3]$ de $\int_{0.2}^{0.3} \frac{1}{x^2} dx$ y obtenemos otra sumando los resultados obtenidos al aplicar la regla de Simpson en los intervalos $[0.2, 0.25]$ y $[0.25, 0.3]$. Comparamos ambas estimaciones:

$$S_4[0.2, 0.3] - (S_5[0.2, 0.25] + S_5[0.25, 0.3]) < \mathrm{T}_{[0.2,0.3]} = 0.0025.$$

Como la diferencia es menor que la tolerancia extrapolando obtenemos la estimación definitiva de $\int_{0.2}^{0.3} \frac{1}{x^2} dx$:

$$\int_{0.2}^{0.3} \frac{1}{x^2} dx \approx (+\text{exacta}) + \frac{1}{15} [(+\text{exacta}) - (\text{-exacta})] = \mathbf{1.666686}.$$

Por lo tanto, la estimación mediante este método de $\int_{0.2}^{1} \frac{1}{x^2} dx$ sería:

$$\begin{aligned} \int_{0.2}^{1} \frac{1}{x^2} dx &= \int_{0.2}^{0.3} \frac{1}{x^2} dx + \int_{0.3}^{0.4} \frac{1}{x^2} dx + \int_{0.4}^{0.6} \frac{1}{x^2} dx + \int_{0.6}^{1} \frac{1}{x^2} dx \\ &= 1.666686 + 0.83333492 + 0.8333428 + 0.66669662 = 4.00005957. \end{aligned}$$

□

Tabla diferenciación e integración numérica

Derivación	
$f'(x) = \frac{f(x+h)-f(x)}{h}$	$-\frac{h}{2}f''(\xi)$
$f'(x) = \frac{f(x+h)-f(x-h)}{2h}$	$-\frac{h^2}{6}f'''(\xi)$
$f''(x) = \frac{f(x+h)-2f(x)+f(x-h)}{h^2}$	$-\frac{h^2}{12}f^{4)}(\xi)$
$f'(x_k) = \sum_{i=0}^{n} f(x_i)L_i'(x_k)$	$\frac{1}{(n+1)!}f^{n+1)}(\xi_{x_k})\Pi_{j=0,j\neq k}^{n}(x_k - x_j)$
$R(0,h) = \frac{f(x+h)-f(x-h)}{2h}$ $R(n,h) = R(n-1,h/2) + \frac{1}{4^n-1}[R(n-1,h/2) - R(n-1,h)]$	
Integración	
$\sum_{i=0}^{n}(\int_a^b l_i(x)dx)f(x_i)$	
$\frac{h}{2}[f(x_0)+f(x_1)]$	$-\frac{1}{12}h^3f''(\xi)$
$\frac{h}{3}[f(x_0)+4f(x_1)+f(x_2)]$	$-\frac{1}{90}h^5f^{4)}(\xi)$
$\frac{3h}{8}[f(x_0)+3f(x_1)+3f(x_2)+f(x_3)]$	$-\frac{3}{80}h^5f^{4)}(\xi)$
$\frac{h}{2}[f(x_0)+2f(x_1)+2f(x_2)+\cdots+2f(x_{n-1})+f(x_n)]$	$-\frac{(b-a)}{12}h^2f''(\xi)$
$\frac{h}{3}[f(x_0)+4f(x_1)+2f(x_2)+4f(x_3)+\cdots+4f(x_{n-1})+f(x_n)]$	$-\frac{(b-a)}{180}h^4f^{4)}(\xi)$
$f(-\sqrt{\frac{1}{3}}) + f(\sqrt{\frac{1}{3}})$	
$R(n,m) = R(n,m-1) + \frac{1}{4^m-1}[R(n,m-1) - R(n-1,m-1)]$	
$S_{i+1}[a,b] + \frac{1}{15}\big[S_{i+1}[a,b] - S_i[a,b]\big]$	

Capítulo 6

Resolución numérica de ecuaciones diferenciales

6.1 Existencia y unicidad de soluciones

Una ecuación diferencial ordinaria (EDO) es una expresión de la forma

$$F(x, y(x), y'(x), y''(x), \dots) = 0,$$

es decir, una ecuación que contiene una variable independiente x, otra dependiente de la primera $y(x)$ y derivadas de distinto orden de la segunda respecto de la primera ($y'(x)$, $y''(x)$, ...).

El *orden* de una EDO es el de la mayor derivada que aparece en la ecuación. El *grado* de una EDO es el mayor exponente de la derivada de mayor orden que aparece en la ecuación.

Ejemplo:

ecuación	orden	grado
$y' = y\tan(x+3)$	1	1
$(y')^2 = y\tan(x+3)$	1	2
$y'' = y\tan(x+3)$	2	1

□

La *solución general* de una EDO es una familia de funciones que verifican la EDO. En general, la solución general de una EDO de orden k es una familia de funciones que depende de una parámetro k.

Ejemplo: La solución general de $y' = \text{sen}(x)$ es $y = -\cos(x) + k$ con $k \in \mathbb{R}$. La solución general de $y'' = -y$ es $y = a\,\text{sen}(x) + b\cos(x)$, con $a \in \mathbb{R}$ y $b \in \mathbb{R}$.

□

Una *solución particular* de una EDO es un elemento particular de la solución general. Una *solución singular* es una función que verifica la EDO pero que no está recogida en la solución general.

Ejemplo: La ecuación $y' = y^{1/3}$ tiene como solución general $y = (\frac{2}{3}x)^{\frac{3}{2}} + C$ con $C \in \mathbb{R}$. Una solución particular es $y = (\frac{2}{3}x)^{\frac{3}{2}} + 4$ y una solución singular es $y = 0$.

□

Si una EDO de primer orden $F(x, y(x), y'(x)) = 0$ se puede expresar de la forma $y' = f(x, y(x))$ entonces

Teorema: *Si $f(x, y)$ y $\frac{\partial f(x,y)}{\partial y}$ son continuas en un rectángulo D del plano $\mathbb{R}^2$ y $(x_0, y_0) \in D$, entonces existe una única solución de la ecuación $y' = f(x, y)$ tal que $y(x_0) = y_0$.*

Ejemplo: *Encuentra dónde tiene solución única la ecuación $y' = \frac{-y}{x}$.*

Sea $f(x, y) = \frac{-y}{x}$. La gráfica de dicha función aparece en la figura 6.1. $f(x, y)$ es continua en $\mathbb{R}^2 - \{x = 0\}$. $\frac{\partial f(x,y)}{\partial y} = \frac{-1}{x}$ es continua también en $\mathbb{R}^2 - \{x = 0\}$.

Por el teorema anterior, si existe un rectángulo $D \subset \mathbb{R}^2 - \{x = 0\}$ y $(x_0, y_0) \in D$, entonces el problema $y' = f(x, y)$ tal que $y(x_0) = y_0$ tiene solución única. Por ejemplo, el problema $y' = \frac{-y}{x}$ con $y(2) = 1$, tiene solución única porque $(2, 1)$ están dentro de un rectángulo que a su vez están contenido en $\mathbb{R}^2 - \{x = 0\}$. Dicha solución es $y = \frac{2}{x}$. Sin embargo, si la condición inicial es $y(0) = 2$ el problema no tiene solución.

□

En general, la búsqueda de soluciones exactas de una ecuación diferencial es un problema difícil. Por ello se usan métodos numéricos para encontrar soluciones aproximadas. El planteamiento típico de una ecuación diferencial ordinaria de grado 1 con valor inicial es

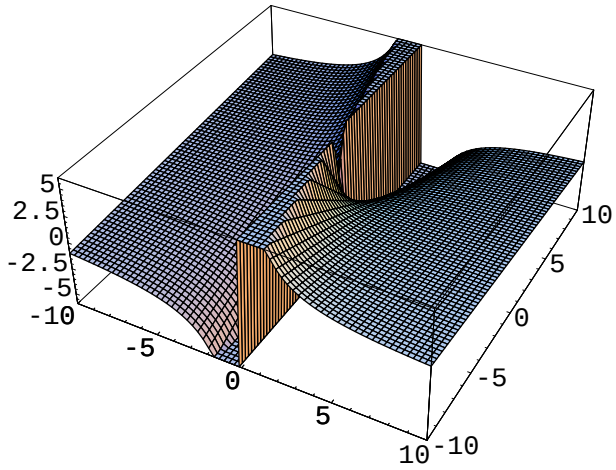


Figura 6.1: La función $f(x,y) = \frac{y}{x}$

$$y' = f(x, y(x)) \quad y(x_0) = y_0 \tag{6.1}$$

Por ejemplo

$$y' = -2x - y \quad y(0) = -1.$$

La solución numérica de una EDO será una tabla de valores aproximados de la función $y(x)$ en un conjunto de puntos x_i. Usando esa tabla podemos construir una aproximación de la función mediante, por ejemplo, los métodos de interpolación. En lo que sigue, trabajaremos con puntos x_i igualmente espaciados, es decir, separados cada uno del siguiente por un tamaño de paso h, con lo que $x_i = x_0 + ih$. $y(x_i)$ denotará el valor exacto de la función $y(x)$ en el punto x_i; y_i será un valor aproximado de $y(x_i)$.

6.2 Método de la serie de Taylor

Para resolver el problema (6.1), el método de Taylor de orden k propone sustituir la función desconocida $y(x)$ por su polinomio de Taylor de orden k

en un entorno de cada punto x_i, es decir

$$y(x_{i+1}) = y(x_i + h) \approx y(x_i) + y'(x_i)h + y''(x_i)\frac{h^2}{2} + \cdots + y^{k)}(x_i) + \frac{h^k}{k!}$$

Los valores de la función $y(x)$ y de sus derivadas en x_i son desconocidos en general, por lo que hay que sustituirlos por aproximaciones:

$$\begin{array}{rcl} y(x_i) & \approx & y_i \\ y'(x_i) & \approx & f(x_i, y_i) \\ y''(x_i) & \approx & \dfrac{\mathrm{d}\, f(x_i, y_i)}{\mathrm{d}\, x} \\ & \cdots & \\ y^{k)}(x_i) & \approx & \dfrac{\mathrm{d}^{\text{k-1}}\, f(x_i, y_i)}{\mathrm{d}\, x^{k-1}} \end{array}$$

De este modo obtenemos la expresión del método:

$$\boxed{y_{i+1} = y_i + f(x_i, y_i)h + \frac{\mathrm{d}\, f(x_i, y_i)}{\mathrm{d}\, x}\frac{h^2}{2} + \cdots + \frac{\mathrm{d}^{\text{k-1}}\, f(x_i, y_i)}{\mathrm{d}\, x^{k-1}}\frac{h^k}{k!}}$$

A la hora de derivar $f(x, y)$ respecto de x hay que tener en cuenta que es una función de dos variables y que y es función de x. De modo que, denotando f_x y f_y a las derivadas parciales de $f(x, y)$ respecto x e y, tendríamos, por ejemplo que

$$\frac{\mathrm{d}\, f(x, y)}{\mathrm{d}\, x} = f_x(x, y) + f_y(x, y)y'(x, y) = f_x(x, y) + f_y(x, y)f(x, y).$$

Elegido, el tamaño del paso h, el método de Taylor paso a paso sería el siguiente:

Paso 1. Partimos del punto x_0, obtenemos una aproximación de $y(x_1)$

$$y_1 = y_0 + f(x_0, y_0)h + \frac{\mathrm{d}\, f(x_0, y_0)}{\mathrm{d}\, x}\frac{h^2}{2} + \cdots + \frac{\mathrm{d}^{\text{k-1}}\, f(x_0, y_0)}{\mathrm{d}\, x^{k-1}}\frac{h^k}{k!}$$

Paso 2. Partimos de $x_1 = x_0 + h$ y calculamos una aproximación $y(x_2)$:

$$y_2 = y_1 + f(x_1, y_1)h + \frac{\mathrm{d}\, f(x_1, y_1)}{\mathrm{d}\, x}\frac{h^2}{2} + \cdots + \frac{\mathrm{d}^{\text{k-1}}\, f(x_1, y_1)}{\mathrm{d}\, x^{k-1}}\frac{h^k}{k!}$$

Paso i+1. Partimos de $x_i = x_{i-1} + h$ y calculamos una aproximación de $y(x_{i+1})$:

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{\mathrm{d}\, f(x_i, y_i)}{\mathrm{d}\, x}\frac{h^2}{2} + \cdots + \frac{\mathrm{d}^{\text{k-1}}\, f(x_i, y_i)}{\mathrm{d}\, x^{k-1}}\frac{h^k}{k!}$$

Ejemplo: *Proporciona tres valores aproximados de la función $y(x)$, solución de $y' = -2x - y$ con $y(0) = -1$ mediante el método de Taylor con $k = 4$ y tamaño de paso $h = 0.1$.*

Calculamos el desarrollo de Taylor hasta grado 4 de la función $y(x)$ en un punto cualquiera x_i:

$$y(x_i + h) \approx y(x_i) + y'(x_i)h + \frac{y''(x_i)}{2}h^2 + \frac{y'''(x_i)}{3!}h^3 + \frac{y^{4)}(x_i)}{4!}h^4$$

Denotando y_i al valor aproximado de $y(x_i)$ y dado que

$$\begin{aligned}
y'(x) &= f(x,y) \\
y''(x) &= [y'(x)]' = [-2x - y(x)]' = -2 - y'(x) = -2 - f(x,y) \\
y'''(x) &= [y''(x)]' = -y''(x) = 2 + y'(x) = 2 + f(x,y) \\
y^{4)}(x) &= [y'''(x)]' = [-y''(x)]' = -y'''(x) = -2 - y'(x) = -2 - f(x,y)
\end{aligned}$$

definimos el método

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{-2 - f(x_i, y_i)}{2}h^2 + \frac{2 + f(x_i, y_i)}{3!}h^3 + \frac{-2 - f(x_i, y_i)}{4!}h^4$$

Paso 1. Como $y_0 = -1$ y $f(x_0, y_0) = f(0, -1) = 1$, obtenemos el valor aproximado

$$y_1 = -1 + 0.1 - 1.5 \cdot 0.1^2 + 0.5 \cdot 0.1^3 - 0.125 \cdot 0.1^4 = -0.9645125.$$

Paso 2. Como $y_1 = -0.9145125$ y $f(x_1, y_1) = f(0.1, -0.9145125) = 0.7145125$, obtenemos el valor aproximado

$$y_2 = -0.8561927.$$

Paso 3. Como $y_2 = -0.8561927$ y $f(x_2, y_2) = f(0.2, -0.8561927) = 0.4561927$, obtenemos el valor aproximado

$$y_3 = -0.8224553$$

□

6.2.1 Método de Euler

Es el método anterior usando el polinomio de Taylor de orden 1

$$\boxed{y_{i+1} = y_i + f(x_i, y_i)h}$$

Ejemplo. *Resuelve la ecuación* $y' = -2x - y$ *con* $y(0) = -1$ *mediante el método de Euler.*

Con estos datos se tiene que $x_0 = 0$, $y_0 = -1$, $f(x_0, y_0) = 1$ con lo que

$$y_1 = y_0 + f(x_0, y_0)h = -1 + 1 \cdot 0.1 = -0.9$$

Puesto que $f(x_1, y_1) = 0.7$, se deduce que

$$y_2 = y_1 + f(x_1, y_1)h = -0.9 + 0.7 \cdot 0.1 = -0.83$$

Puesto que $f(x_2, y_2) = 0.43$, se deduce que

$$y_3 = y_2 + f(x_2, y_2)h = -0.83 + 0.43 \cdot 0.1 = -0.787$$

Puesto que $f(x_3, y_3) = 0.187$, se deduce que

$$y_4 = y_3 + f(x_3, y_3)h = -0.787 + 0.187 \cdot 0.1 = -0.7683$$

□

6.2.2 Errores

Al resolver numéricamente una ecuación diferencial aparecen varios tipos de errores:

1. error de truncamiento local
2. error de truncamiento global
3. error de redondeo local
4. error de redondeo global
5. error total

El **error de truncamiento local** es el que aparece en un paso cuando reemplazamos un proceso infinito por uno finito. Por ejemplo, en el método de Taylor, usando el desarrollo hasta grado k, sustituimos la serie de Taylor (de infinitos términos) por los primeros k-términos. En este caso, para un cierto $c_i \in \mathbb{R}$, se tiene que

$$y(x_i + h) = y(x_i) + y'(x_i)h + \frac{y''(x_i)}{2!}h^2 + \cdots + \frac{y^{k)}(x_i)}{k!}h^k + \frac{y^{k+1)}(c_i)}{(k+1)!}h^{k+1}$$

y dado que en el algoritmo usamos la aproximación

$$y(x_i + h) \approx y(x_i) + y'(x_i)h + \frac{y''(x_i)}{2!}h^2 + \cdots + \frac{y^{k)}(x_i)}{k!}h^k$$

se deduce que el error local en el paso i-ésimo es proporcional a h^{k+1}, digamos $C_i h^{k+1}$. Un error de esa magnitud se dice que es $O(h^{k+1})$.

La acumulación de todos los errores de truncamiento local genera el **error de truncamiento global**. Si al aplicar un método el error de truncamiento local es $O(h^{k+1})$, el tamaño de paso es h, el punto inicial es x_0 y el punto final es x_n, entonces el número de pasos será $n = \frac{x_n - x_0}{h}$ y puesto que

$$\begin{aligned} C_1 h^{k+1} + C_2 h^{k+1} + \cdots + C_n h^{k+1} &= h^{k+1}(C_1 + C_2 + \cdots + C_n) \leq \\ &\leq h^{k+1} n \max\{C_1, C_2, \ldots, C_n\} = Ch^k \end{aligned}$$

siendo $C = (x_n - x_0)\max\{C_1 + C_2 + \cdots + C_n\}$, el error de truncamiento global será $O(h^k)$. Si un método tiene error de truncamiento global $O(h^k)$ decimos que el procedimiento numérico es de **orden** k. El método de Taylor usando el desarrollo hasta grado k es un método de orden k.

El **error de redondeo local** es el que en cada paso es provocado por la limitada precisión de nuestras máquinas de cómputo, como ya se comentó el tema inicial del curso. El **error de redondeo global** es la acumulación de los errores de redondeo local en los pasos anteriores. El **error total** es la suma de los errores de truncamiento global y redondeo global.

6.3 Métodos de Runge-Kutta

Los métodos de Taylor tienen el inconveniente de la dificultad de cálculo de las sucesivas derivadas de la función $y(x)$. En los métodos de Runge-Kutta se

elimina esa dificultad. Las ideas que dan origen a este conjunto de métodos se resumen en los siguientes pasos:

Paso 1. Consideramos la igualdad

$$y(x_{i+1}) - y(x_i) = \int_{x_i}^{x_{i+1}} y'(x)dx = \int_{x_i}^{x_{i+1}} f(x, y(x))dx.$$

y despejamos $y(x_{i+1})$

$$y(x_{i+1}) = y(x_i) + \int_{x_i}^{x_{i+1}} y'(x)dx = \int_{x_i}^{x_{i+1}} f(x, y(x))dx.$$

Paso 2. Realizamos el cambio de variable $x = x_i + \theta h$ en la integral anterior y obtenemos

$$y(x_{i+1}) = y(x_i) + h \int_0^1 f(x_i + \theta h, y(x_i + \theta h))d\theta. \tag{6.2}$$

Paso 3. Consideramos el conjunto de puntos $0 = \theta_0 \leq \theta_1 \leq \ldots \theta_m \leq 1$ del intervalo de integración $[0, 1]$ y estimamos la integral anterior mediante una fórmula de cuadratura del siguiente tipo

$$\int_0^1 F(\theta)d\theta \approx A_0F(\theta_0) + A_1F(\theta_1) + \cdots + A_mF(\theta_m)$$

con lo que, denotando $x_{i\theta_j} = x_i + \theta_j h$, en la ecuación (6.2) obtenemos

$$\begin{aligned} y(x_{i+1}) \approx y(x_i) \ + \ h \ \big[A_0 \ f\big(x_{i\theta_0}, y(x_{i\theta_0})\big) \ + \\ A_1 \ f\big(x_{i\theta_1}, y(x_{i\theta_1})\big) \ + \ \cdots + \ A_m \ f\big(x_{i\theta_m}, y(x_{i\theta_m})\big)\big]. \end{aligned} \tag{6.3}$$

La fórmula de cuadratura elegida puede variar, pero se le ha de exigir que al menos sea exacta para las funciones constantes, lo que se traduce en que los coeficientes han de cumplir la condición

$$A_0 + A_1 + \cdots + A_m = 1$$

El método de Runge-Kutta será la fórmula que resulta de sustituir en la ecuación (6.3) los valores $y(x_i)$, $y(x_{i\theta_0})$, $y(x_{i\theta_1})$, $\ldots$, $y(x_{i\theta_m})$ por sus valores aproximados. Si bien dispondremos en cada paso del valor y_i que aproxima a $y(x_i)$, no ocurre lo mismo con las aproximaciones de $y(x_{i\theta_0})$, $y(x_{i\theta_1})$, $\ldots$, $y(x_{i\theta_m})$.

Paso 4. En esta etapa conseguiremos obtener un algoritmo para las aproximaciones aludidas en el apartado anterior. De forma similar a como se obtuvo la ecuación (6.2), llegamos a la igualdad

$$y(x_{i\theta_j}) = y(x_i) + h\int_{x_i}^{\theta_j} f(x_i + \theta h, y(x_i + \theta h))d\theta$$

y aplicando una fórmula de cuadratura para cada j

$$\int_{x_i}^{\theta_j} F(\theta)d\theta \approx B_0^j F(\theta_0) + B_1^j F(\theta_1) + \cdots + B_{j-1}^j F(\theta_{j-1}) \tag{6.4}$$

a la integral anterior, se obtienen las aproximaciones buscadas

$$y(x_{i\theta_j}) \approx y(x_i) + h\sum_{k=0}^{j-1} B_k^j f(x_{i\theta_k}, y(x_{i\theta_k})).$$

Las fórmulas de cuadraturas usadas en (6.4) han de ser exactas al menos para las funciones constantes, por lo tanto los coeficientes deben verificar

$$B_0^j + B_1^j + \cdots + B_{j-1}^j = \theta_j \qquad j = 1, \ldots, m.$$

Paso 5. Si denotamos $y_{i\theta_j}$ a una estimación de $y(x_{i\theta_j})$, un método de Runge-Kutta quedarían así:

$$y_{i+1} = y_i + h[A_0 f(x_{i\theta_0}, y_{i\theta_0}) \ + A_1 f(x_{i\theta_1}, y_{i\theta_1}) + \cdots + A_m f(x_{i\theta_m}, y_{i\theta_m})] \tag{6.5}$$

siendo

$$\begin{aligned}
&y_{i\theta_0} = y_i \\
&y_{i\theta_1} = y_i + h\big[B_0^1 f(x_{i\theta_0}, y_{i\theta_0})\big] \\
&y_{i\theta_2} = y_i + h\big[B_0^2 f(x_{i\theta_0}, y_{i\theta_0}) + B_1^2 f(x_{i\theta_1}, y_{i\theta_1})\big] \\
&\cdots \\
&y_{i\theta_m} = y_i + h\big[B_0^m f(x_{i\theta_0}, y_{i\theta_0}) + B_1^m f(x_{i\theta_1}, y_{i\theta_1}) + \ldots \\
&\qquad\qquad\qquad\qquad \cdots + B_{m-1}^m f(x_{i\theta_{m-1}}, y_{i\theta_{m-1}})\big]
\end{aligned}$$

Previamente deben fijarse los valores

$$0 = \theta_0 \leq \theta_1 \leq \theta_2 \leq \ldots \theta_m \leq 1$$

así como los siguientes parámetros que han de cumplir con las condiciones de las sumas correspondientes que aparecen en el cuadro

A_0	A_1	A_2	$\ldots$	A_m	$\sum_{r=0}^{m} A_r = 1$
	B_0^1	B_0^2		B_0^m	
		B_1^2	$\ldots$	B_1^m	
				$\ldots$	
				B_{m-1}^m	
	$B_0^1 = \theta_1$	$B_0^2 + B_1^2 = \theta_2$	$\ldots$	$\sum_{k=0}^{m-1} B_k^m = \theta_m$	

Un método como el anterior se conoce con el nombre de método de Runge-Kutta de $m+1$ etapas, debido a que la fórmula (6.5) requiere $m+1$ evaluaciones de la función $f(x, y)$.

Ejemplo: Para $m = 0$ (una etapa), el método de Runge-Kutta que resulta es el **método de Euler**.

Para $m = 1$ (dos etapas), podemos obtener, entre otras posibilidades, el **método de Heun**:

$$\theta_0 = 0,\ \theta_1 = 1$$

$A_0 = 1/2$	$A_1 = 1/2$	$A_0 + A_1 = 1$
	$B_0^1 = 1$	
	$B_0^1 = \theta_1 = 1$	

$$y_{i+1} = y_i + h[\frac{1}{2}f(x_{i0}, y_{i0}) \ + \frac{1}{2}f(x_{i1}, y_{i1})]$$

siendo

$$\begin{aligned} y_{i0} &= y_i \\ y_{i1} &= y_i + hf(x_i, y_i) \end{aligned}$$

con lo que

$$\boxed{y_{i+1} = y_i + h[\frac{1}{2}f(x_i, y_i) \ + \frac{1}{2}f(x_{i+1}, y_i + hf(x_i, y_i))]} \qquad (6.6)$$

El llamado **método de Euler modificado** sale con la elección

$$\theta_0 = 0,\ \theta_1 = 1/2$$

$A_0 = 0$	$A_1 = 1$	$A_0 + A_1 = 1$
	$B_0^1 = 1/2$	
	$B_0^1 = \theta_1 = 1/2$	

$$y_{i+1} = y_i + h[f(x_{i\theta 1}, y_{i\theta 1})]$$

siendo

$$\begin{aligned} y_{i0} &= y_i \\ y_{i1/2} &= y_i + h\tfrac{1}{2} f(x_i, y_i) \end{aligned}$$

con lo que el método queda

$$\boxed{y_{i+1} = y_i + hf\big(x_i + \frac{h}{2}, y_i + \frac{h}{2} f(x_i, y_i)\big)} \tag{6.7}$$

Otra elección posible es

$$\theta_0 = 0,\ \theta_1 = 2/3$$

$A_0 = 1/4$	$A_1 = 3/4$	$A_0 + A_1 = 1$
	$B_0^1 = 2/3$	
	$B_0^1 = \theta_1 = 2/3$	

$$y_{i+1} = y_i + h[\frac{1}{4} f(x_{i\theta_0}, y_{i\theta_0}) \ + \frac{3}{4} f(x_{i\theta_1}, y_{i\theta_1})]$$

siendo

$$\begin{aligned} y_{i\theta_0} &= y_i \\ y_{i\theta_1} &= y_i + h\big[\tfrac{2}{3} f(x_{i\theta_0}, y_{i\theta_0})\big] \end{aligned}$$

con lo que el método queda así

$$\boxed{y_{i+1} = y_i + h[\frac{1}{4} f(x_i, y_i) \ + \frac{3}{4} f(x_i + \frac{2}{3}h, y_i + h\frac{2}{3} f(x_i, y_i)]}$$

El método clásico de Runge-Kutta es de $m = 3$ (cuatro etapas) definido por los parámetros

$$0 = \theta_0 \ \ \theta_1 = 1/2 \ \ \theta_2 = 1/2 \ \ \theta_3 = 1$$

$A_0 = 1/6$	$A_1 = 1/3$	$A_2 = 1/3$	$A_3 = 1/6$	$\sum_{r=0}^{3} A_r = 1$
	$B_0^1 = 1/2$	$B_0^2 = 0$ $B_1^2 = 1/2$	$B_0^3 = 0$ $B_1^3 = 0$ $B_2^3 = 1$	
	$B_0^1 = 1/2$	$B_0^2 + B_1^2 = 1/2$	$\sum_{k=0}^{2} B_k^3 = 1$	

que da lugar a

$$\boxed{y_{i+1} = y_i + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4)} \tag{6.8}$$

donde

$$K_1 = f(x_i, y_i) \qquad K_2 = f(x_i + \tfrac{h}{2}, y_i + \tfrac{h}{2}K_1)$$
$$K_3 = f(x_i + \tfrac{h}{2}, y_i + \tfrac{h}{2}K_2) \qquad K_4 = f(x_i + h, y_i + hK_3)$$

□

Ejercicio: *Aplica tres pasos del método clásico de Runge-Kutta de orden 4 con* $h = 1/128$ *para resolver la ecuación*

$$y' = \frac{yx - y^2}{x^2} \qquad y(1) = 2.$$

Paso 1. Partimos de $x_0 = 1$, $y_0 = 2$. Para aplicar la ecuación (6.8) previamente calculamos

$$K_1 = f(x_0, y_0) = f(1, 2) = -2 \qquad K_2 = f(x_0 + \tfrac{h}{2}, y_0 + \tfrac{h}{2}K_1) = -1.95355$$
$$K_3 = f(x_0 + \tfrac{h}{2}, y_0 + \tfrac{h}{2}K_2) = -1.95409$$
$$K_4 = f(x_0 + h, y_0 + hK_3) = -1.90898.$$

Con estos datos obtenemos

$$y_1 = y_0 + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) = 1.98473,$$

Paso 2. Partimos de $x_1 = 1 + \frac{1}{128}$, $y_1 = 1.98473$. Calculamos

$$K_1 = f(x_1, y_1) = -1.90899 \qquad K_2 = f(x_1 + \tfrac{h}{2}, y_1 + \tfrac{h}{2}K_1) = -1.86520$$
$$K_3 = f(x_1 + \tfrac{h}{2}, y_1 + \tfrac{h}{2}K_2) = -1.86570$$
$$K_4 = f(x_1 + h, y_1 + hK_3) = -1.82316.$$

Con estos datos obtenemos

$$y_2 = y_1 + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) = 1.97016.$$

Paso 3. Partimos de $x_2 = 1 + \frac{2}{128}$, $y_2 = 1.97016$. Calculamos

$$K_1 = f(x_2, y_2) = -1.82311 \qquad K_2 = f(x_2 + \tfrac{h}{2}, y_2 + \tfrac{h}{2}K_1) = -1.78186$$
$$K_3 = f(x_2 + \tfrac{h}{2}, y_2 + \tfrac{h}{2}K_2) = -1.78231$$
$$K_4 = f(x_2 + h, y_2 + hK_3) = -1.74215.$$

Con estos datos obtenemos

$$y_3 = y_2 + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) = 1.96088.$$

□

6.3.1 Errores

Los métodos de Runge-Kutta tienen también su error de truncamiento global, es decir, su orden. Existen métodos de Runge-Kutta de orden k con k menor o igual que el número de etapas $m + 1$. Se da la igualdad para $k = 1, 2, 3, 4$. Es decir, los métodos de Runge-Kutta de 1,2,3,4 etapas tienen orden 1,2,3,4, respectivamente. Para conseguir orden $k = 5$ es necesario 6 etapas; para $k = 6$, 7 etapas; en general para $k \geq 7$ se necesitan un número de etapas mayor o igual que $k + 2$.

6.4 Métodos multipaso

Tanto los métodos de Taylor como los de Runge-Kutta son métodos de un paso, porque para calcular el valor aproximado y_{i+1} solo se usa, además de x_i y de h, el valor aproximado anterior y_i. Los métodos multipaso se caracterizan por obtener y_{i+1} a partir de los valores $h, x_i, y_i, x_{i-1}, y_{i-1}, \ldots, x_{i-k}, y_{i-k}$.

Partimos siempre de la ecuación

$$y' = f(x, y) \qquad y(x_0) = y_0.$$

Ejemplo: Un primer ejemplo sencillo de un método multipaso se puede obtener de la siguiente forma. Sabemos que

$$y'(x) \approx \frac{y(x + h) - y(x - h)}{2h}$$

con lo que

$$y(x_i) \approx \frac{y(x_{i+1}) - y(x_{i-1})}{2h}$$

y se deduce que

$$y(x_{i+1}) \approx y(x_{i-1}) + 2\ h\ y'(x_i),$$

lo que sugiere el método

$$\boxed{y_{i+1} = y_{i-1} + 2\ h\ f(x_i, y_i).}$$

Para poder obtener el punto y_{i+1} son necesarios los dos valores anteriores y_i, y_{i-1}. Es un método de dos pasos. Para empezar a aplicarlo tenemos que disponer de y_0 e y_1. Este último puede calcularse con un método de un paso. □

El que sigue es un proceso para obtener un conjunto de métodos multipasos. Partimos de la relación

$$y(x_{i+1}) - y(x_i) = \int_{x_i}^{x_{i+1}} y'(x)dx = \int_{x_i}^{x_{i+1}} f(x, y(x))dx,$$

y aplicamos el cambio de variable $x = x_i + \theta h$

$$y(x_{i+1}) - y(x_i) = h \int_0^1 f(x_i + \theta h, y(x_i + \theta h))d\theta.$$

Aproximamos la integral anterior por una fórmula de cuadratura

$$\int_0^1 F(\theta)d\theta = A_0 F(\theta_0) + A_1 F(\theta_1) + \cdots + A_k F(\theta_k) \tag{6.9}$$

y nos queda

$$\begin{aligned} y(x_{i+1}) - y(x_i) \approx h\ \big[&A_0 f(x_i + \theta_0 h, y(x_i + \theta h)) + \\ &+ A_1 f(x_i + \theta_1 h, y(x_i + \theta h)) + \\ &+ \cdots + A_k f(x_i + \theta_k h, y(x_i + \theta_k h))\big] \end{aligned} \tag{6.10}$$

lo cual sugiere un método que resultará de sustituir los valores de la función por valores aproximados.

Ejemplo: *Deduzcamos el método de orden 2*

$$\boxed{y_{i+1} = y_i + h\big[\frac{3}{2}f(x_i, y_i) - \frac{1}{2}f(x_{i-1}, y_{i-1})\big]} \tag{6.11}$$

Si en la fórmula de cuadratura (6.9) tomamos los puntos

$$\theta_0 = 0 \qquad \theta_1 = -1$$

y le imponemos la condición de que sea exacta para los polinomios de grado menor o igual que 1, obtendremos que los coeficientes A_0 y A_1 deben verificar las ecuaciones

$$A_0 + A_1 = 1 \qquad A_1 = -\frac{1}{2}$$

y la solución del sistema anterior es $A_0 = \frac{3}{2}$ y $A_1 = -\frac{1}{2}$. Sustituyendo en (6.10) nos queda

$$y(x_{i+1}) - y(x_i) \approx h\ \big[\frac{3}{2}f(x_i, y(x_i)) - \frac{1}{2}f(x_{i-1}, y(x_{i-1})\big]$$

lo cual sugiere el método

$$y_{i+1} = y_i + h\ \big[\frac{3}{2}f(x_i, y_i) - \frac{1}{2}f(x_{i-1}, y_{i-1}\big].$$

□

Ejercicio: A*plica el método anterior al cálculo de la solución de*

$$y' = \frac{yx - x^2}{x^2} \qquad y(1) = 2$$

Para aplicar la fórmula (6.11) necesitamos dos puntos iniciales y_0, y_1, y solo disponemos del valor $y_0 = 2$. Es habitual calcular los puntos necesarios para empezar mediante un método de un paso del mismo orden. Usamos el método de Euler modificado (ver ecuación (6.7)), que es un método de Runge-Kutta de orden 2:

$$y_1 = y_0 + hf\big(x_0 + \frac{h}{2}, y_0 + \frac{h}{2}f(x_0, y_0)\big) = 1.98474.$$

Ya disponemos de y_0 e y_1 y podemos calcular y_2 aplicando (6.11)

$$y_2 = y_1 + h\ \big[\frac{3}{2}f(x_1, y_1) - \frac{1}{2}f(x_0, y_0\big] = 1.97018.$$

Con y_1 e y_2 podemos calcular y_3

$$y_3 = y_2 + h\ \big[\frac{3}{2}f(x_2, y_2) - \frac{1}{2}f(x_1, y_1\big] = 1.95627.$$

Y así sucesivamente.

□

6.4.1 Fórmulas de Adams-Bashforth

Un método multipaso de la forma

$$\boxed{y_{i+1} = y_i + A_0 f(x_i, y_i) + A_1 f(x_{i-1}, y_{i-1}) + \cdots + A_{k-1} f(x_{i-k+1}, y_{i-k+1})}$$

es una fórmula de Adams-Bashforth de k pasos y orden k. Estos métodos se encuadran dentro de los métodos multipaso explícitos, debido a que la definición de y_{i+1} es explícita en función de los valores anteriores, a diferencia de los métodos implícitos que veremos en es siguiente apartado. Estos son algunos ejemplos de fórmulas de Adams-Bashforth:

1. de orden 2:

$$y_{i+1} = y_i + h\big[\tfrac{3}{2}f(x_i, y_i) - \tfrac{1}{2}f(x_{i-1}, y_{i-1})\big]$$

2. de orden 4:

$$y_{i+1} = y_i + \frac{h}{24}\big[55f(x_i, y_i) - 59f(x_{i-1}, y_{i-1}) + \\ + 37f(x_{i-2}, y_{i-2}) - 9f(x_{i-3}, y_{i-3})\big]$$

3. de orden 5:

$$y_{i+1} = y_i + \frac{h}{720}\big[1901f(x_i, y_i) - 2774f(x_{i-1}, y_{i-1}) + \\ + 2616f(x_{i-2}, y_{i-2}) - 1274f(x_{i-3}, y_{i-3}) + 251f(x_{n-4}, y_{n-4})\big]$$

Todas ellas se pueden obtener siguiendo el proceso descrito en la página 126. Dado que en dicho proceso se utiliza el cambio de variable $x = x_i + \theta\ h$, si en la fórmula final (6.10) deseo que aparezca la función $f(x, y)$ valorada en los puntos

$$x_i,\ x_{i-1},\ ...,\ x_{i-k+1},$$

tendré que elegir los valores

$$\theta_0 = 0,\ \theta_1 = -1,\ ...,\ \theta_{k-1} = -k+1$$

al construir la fórmula de cuadratura (6.9).

Para aplicar una fórmula de Adams-Bashforth de k pasos necesitaremos los valores iniciales $y_0,\ y_1,\ ...,\ y_{k-1}$, que podrán obtenerse mediante un método de un paso del mismo orden.

6.4.2 Fórmulas de Adams-Moulton

Un método multipaso de la forma

$$\boxed{\begin{aligned} y_{i+1} = y_i + Bf(x_{i+1}, y_{i+1}) + A_0 f(x_i, y_i) + A_1 f(x_{i-1}, y_{i-1}) + \dots \\ \dots + A_{k-1} f(x_{i-k+1}, y_{i-k+1}) \end{aligned}} \quad (6.12)$$

es una fórmula de Adams-Moulton de k pasos y orden $k+1$. Estos métodos se encuadran dentro de los métodos multipaso implícitos, debido a que en la fórmula aparece el valor y_{i+1} definido de forma implícita. Al igual que las fórmulas de Adams-Bashforth, las fórmulas de Adams-Moulton se pueden obtener siguiendo el proceso descrito en la página 126. Dado que en dicho proceso se utiliza el cambio de variable $x = x_i + \theta\ h$, si en la fórmula final (6.10) deseo que aparezca la función $f(x, y)$ valorada en los puntos

$$x_{i+1},\ x_i,\ x_{i-1},\ ...,\ x_{i-k+1},$$

tendré que elegir los valores

$$\theta = 1,\ \theta_0 = 0,\ \theta_1 = -1,\ ...,\ \theta_k = -k+1$$

al construir la fórmula de cuadratura (6.9).

Para aplicar una fórmula de Adams-Moulton de k pasos necesitaremos los valores iniciales $y_0,\ y_1,\ ...,\ y_{k-1}$, que podrán obtenerse mediante un método de un paso del mismo orden.

Supongamos que disponemos de los valores $y_i,\ y_{i-1},\ ...,\ y_{i-k+1}$. ¿Cómo calcular y_{i+1} mediante la fórmula (6.12)? Se puede hacer de dos formas:

1. Método predictor-corrector. Se estima un primer valor predictor y^*_{i+1} mediante una fórmula de Adams-Bashforth del mismo orden y después se calcula el valor corrector y_{i+1} aplicando

$$y_{i+1} = y_i + Bf(x_{i+1}, y^*_{i+1}) + A_0 f(x_i, y_i) + A_1 f(x_{i-1}, y_{i-1}) + \ldots \\ \cdots + A_{k-1} f(x_{i-k+1}, y_{i-k+1}).$$

2. Método iterativo de punto fijo. En la fórmula (6.12) todos los términos de la parte derecha de la igualdad son datos excepto y_{i+1}. En este sentido, y_{i+1} es un punto fijo de la función

$$\Phi(\mathbf{t}) = y_i + Bf(x_{i+1}, \mathbf{t}) + A_0 f(x_i, y_i) + A_1 f(x_{i-1}, y_{i-1}) + \ldots \\ \cdots + A_{k-1} f(x_{i-k+1}, y_{i-k+1}).$$

En consecuencia, y_{i+1} lo podemos obtener como límite de la sucesión

$$t_0 = y^*_{i+1}, \qquad t_n = \Phi(t_{n-1}) \quad n \geq 1$$

siendo y^*_{i+1} un valor inicial que se recomienda obtener mediante la fórmula de Adams-Bashforth del mismo orden. En la práctica, es suficiente dos o tres iteraciones para conseguir un valor aceptable de y_{i+1}.

Ejemplo: *Deduzcamos la fórmula de Adams-Moulton de orden dos*

$$\boxed{y_{i+1} = y_i + h\big[\frac{1}{2} f(x_{i+1}, y_{i+1}) + \frac{1}{2} f(x_i, y_i)\big]} \tag{6.13}$$

Aplicaremos el proceso descrito en la página 126. Si en la fórmula de cuadratura (6.9) tomamos los puntos

$$\theta_0 = 1 \qquad \theta_1 = 0$$

y le imponemos la condición de que sea exacta para los polinomios de grado menor o igual que 1, obtendremos que los coeficientes A_0 y A_1 deben verificar las ecuaciones

$$A_0 + A_1 = 1 \qquad A_0 = \frac{1}{2}$$

y la solución del sistema anterior es $A_0 = \frac{1}{2}$ y $A_1 = \frac{1}{2}$. Sustituyendo en (6.10) nos queda

$$y(x_{i+1}) - y(x_i) \approx h\ \big[\frac{1}{2}f(x_{i+1}, y(x_{i+1})) + \frac{1}{2}f(x_i, y(x_i)\big]$$

lo cual sugiere el método

$$y_{i+1} = y_i + h\ \big[\frac{1}{2}f(x_{i+1}, y_{i+1}) + \frac{1}{2}f(x_i, y_i\big].$$

□

Ejercicio: *Aplica el método anterior al cálculo de la solución de*

$$y' = \frac{yx - x^2}{x^2} \qquad y(1) = 2$$

El método a aplicar es de orden 2 y 1 paso, por lo que solo necesitamos el valor y_0 para empezar. Para estimar el resto de valores podemos optar por el método predictor-corrector o por el método iterativo de punto fijo.

1. Método predictor. Calculamos un valor predictor y_1^*. Si es posible, debe calcularse con una fórmula de Adams-Bashforth del mismo orden, en este caso la fórmula (6.11) que es de orden dos. Sin embargo, para aplicar (6.11) necesitamos los dos puntos iniciales y_0 e y_1, con lo que en este primer paso no podemos hacer uso de ella. Calculamos y_1^* mediante otro método, por ejemplo, por Runge-Kutta de orden 2 (ver 6.6):

$$y_1^* = y_0 + h\big[\frac{1}{2}f(x_0, y_0)\ + \frac{1}{2}f(x_1, y_0 + hf(x_0, y_0))\big] = 1.98474$$

Con este valor, calculamos el valor corrector con Adams-Moulton

$$y_1 = y_0 + h\ \big[\frac{1}{2}f(x_1, y_1^*) + \frac{1}{2}f(x_0, y_0\big] = 1.98473.$$

Para calcular y_2, primero calculo el valor predictor y_2^* con Adams-Bashforth de orden 2 (fórmula (6.11))

$$y_2^* = y_1 + h\big[\frac{3}{2}f(x_1, y_1) - \frac{1}{2}f(x_0, y_0)\big] = 1.97017$$

y después calculo el corrector con Adams-Moulton

$$y_2 = y_1 + h\ \big[\frac{1}{2}f(x_2, y_2^*) + \frac{1}{2}f(x_1, y_1\big] = 1.97015.$$

En el siguiente paso calculo el predictor y_3^* con Adams-Bashforth

$$y_3^* = y_2 + h\big[\frac{3}{2}f(x_2, y_2) - \frac{1}{2}f(x_1, y_1)\big] = 1.95624$$

y después calculo el corrector con Adams-Moulton

$$y_3 = y_2 + h\ \big[\frac{1}{2}f(x_3, y_3^*) + \frac{1}{2}f(x_2, y_2\big] = 1.95622.$$

El proceso anterior se repite para calcular el resto de valores.

2. Método iterativo de punto fijo. Dada la función

$$\Phi_1(t) = y_0 + h\big[\frac{1}{2}f(x_1, t) + \frac{1}{2}f(x_0, y_0)\big]$$

el valor y_1 que buscamos verifica

$$\Phi_1(y_1) = y_1.$$

Es, por tanto, un punto fijo de la ecuación

$$\Phi_1(t) = t.$$

Para localizarlo generamos una sucesión iterativa comenzando por un valor y_1^* obtenido, si es posible, por Adams-Bashforth del mismo orden. En este caso no es posible aplicar Adams-Bashforth puesto que se necesitaría dos puntos iniciales que no tenemos. Calculamos y_1^* mediante Runge-Kutta de orden 2 ($y_1^* = 1.98474$) y generamos la sucesión iterativa:

$$\begin{array}{rcl} t_0 & = & y_1^* = 1.98474 \\ t_1 & = & \Phi_1(1.98474) = 1.98473 \\ t_2 & = & \Phi_1(1.98473) = 1.98473 \end{array}$$

De forma que tomamos $y_1 = 1.98473$.

y_2 es un punto fijo de

$$\Phi_2(t) = y_1 + h\big[\frac{1}{2}f(x_2, t) + \frac{1}{2}f(x_1, y_1)\big]$$

que localizamos mediante la sucesión iterativa similar a la anterior que comienza en $y_2^* = 1.97017$, que a su vez es obtenido por Adams-Bashforth de orden 2

$$\begin{aligned} t_0 &= y_2^* = 1.97017 \\ t_1 &= \Phi_2(1.97017) = 1.97015 \\ t_2 &= \Phi_2(1.97015) = 1.97015 \end{aligned}$$

Tomamos $y_2 = 1.97015$. Y repetimos el proceso para el resto de valores.

□

Ejercicio:

1. *Deduce la fórmula de Adams-Bashforth de orden 4:*

$$y_{i+1} = y_i + \frac{h}{24}\big[55f(x_i, y_i) - 59f(x_{i-1}, y_{i-1}) + \\ + 37f(x_{i-2}, y_{i-2}) - 9f(x_{i-3}, y_{i-3})\big]$$

   y úsala para resolver la ecuación

$$y' = \frac{yx - x^2}{x^2} \qquad y(1) = 2$$

   con $h = 1/128$

2. *Repite el apartado anterior para la fórmula de Adams-Moulton de orden 4:*

$$y_{i+1} = y_i + \frac{h}{24}\big[9f(x_{i+1}, y(i+1)) + 19f(x_i, y_i) - \\ - 5f(x_{i-1}, y_{i-1}) + f(x_{i-2}, y_{i-2})\big]$$

□

Ejercicio: *Usa la identidad*

$$y(x_{i+2}) - y(x_i) = \int_{x_i}^{x_{i+2}} y'(x)dx$$

y la fórmula de cuadratura de Simpson 1/3 para obtener la fórmula del siguiente método multipaso

$$y_{i+2} = y_i + \frac{h}{3}\big[f(x_i, y_i) + 4f(x_{i+1}, y_{i+1}) + f(x_{i+2}, y_{i+2})\big].$$

□

Problemas

Ecuaciones no lineales

1. Dada la ecuación
$$\sqrt{x}sen(x) - x^3 + 2 = 0$$
encuentra una aproximación de una raíz
   (a) Con el método de la bisección con un error menor que 1/30.
   (b) Con 4 iteraciones del método de Newton-Raphson.

2. Determina un intervalo y una función para aplicar el método del punto fijo a las siguientes ecuaciones:
   (a) $x^3 - x - 1 = 0$
   (b) $4 - x - tg(x) = 0$

   Realiza 4 iteraciones del método y determina en cada caso una cota del el error cometido.

3. Se considera la ecuación $x^2 - 1 - sen(x) = 0$.
   (a) Prueba que dicha ecuación tiene, al menos, una raíz positiva.
   (b) Encuentra un intervalo en el cual la iteración
$$x_n = \sqrt{1 + sen(x_{n-1})} \quad x \in \mathbb{N}$$
   converja, para cualquier valor inicial x_0 de dicho intervalo, a una raíz positiva de la ecuación anterior. ¿Cuántos pasos deben darse, a partir de $x_0 = \pi/2$, para obtener una aproximación de la raíz con un error inferior a una milésima?

4. Demuestra que la ecuación

$$e^x L(x) + x^3 - 2 = 0$$

tiene una única raíz positiva. Aplica el método de Newton-Raphson para encontrar una aproximación a la raíz. Realiza iteraciones hasta que se repitan las primeros cuatro decimales.

5. Determina un intervalo y una función para poder aplicar el método del punto fijo a las siguientes ecuaciones:

   (a) $x - L(1 + x) - 0.2 = 0$.
   (b) $x = -L(x)$.

   Calcula una aproximación a la solución con un error inferior a 10^{-5}.

6. Se considera la función

$$F(x) = e^{-x} - sen(x) - 2.$$

   (a) Demuestra que $F(x) < 0$ si $x \geq 0$.
   (b) Prueba que la ecuación $F(x) = 0$ tiene una única raíz negativa.
   (c) Determina un intervalo donde se pueda aplicar el método de Newton-Raphson para aproximar dicha raíz, así como los 4 primeros términos de la sucesión definida mediante dicho método. Calcula una cota del error cometido.

7. Dada la función $f(x) = e^x - 3x^2$,

   (a) Localiza un intervalo que contenga una solución de $f(x) = 0$. Enuncia el resultado teórico que utilices.
   (b) Encuentra, aplicando el método del punto fijo con cuatro iteraciones, una aproximación de un valor donde se anule la derivada de $f(x)$ y una cota del error cometido.

8. Dada la ecuación $x = arctg(x+1)$, aplica el método de la bisección para encontrar una solución con un error menor que $1/10$.

9. Calcula tres iteraciones del método de la secante para encontrar una raíz de la ecuación

$$cos(x) - x = 0$$

en el intervalo $[0.5, \pi/4]$

10. Dada la ecuación $2cos(2x) + 4x - k = 0$,

    (a) Determina el valor de k para que la ecuación tenga una única raíz triple en el intervalo $[0, 1]$.

    (b) Para $k = 3$, prueba que posee una única raíz simple en el intervalo $[0, 1]$ y calcúlala con el método de Newton con un error menor que 10^{-3}.

11. Dado el polinomio $P(x) = x^4 - 4x^3 + 7x^2 - 5x - 2$

    (a) Utiliza la sucesión de Sturm para separar las raíces del polinomio.

    (b) Utiliza el método de Newton-Raphson para encontrar una aproximación de la raíz real de $P(x)$ de menor valor absoluto.

12. Dado el polinomio $P(x) = 8x^3 - 20x^2 - 2x + 5$, utiliza su sucesión de Sturm para separar sus raíces en intervalos disjuntos. Encuentra aproximaciones de las raíces con un error menor que 10^{-3}.

13. Dado la ecuación $x^4 - 7x^3 + 18x^2 - 20x + 8 = 0$:

    (a) Demuestra con el teorema de Bolzano que tiene una solución en el intervalo $[1.5, 3]$. y utiliza el método de Newton para encontrar una aproximación con error menor que 10^{-6}

    (b) Utiliza la sucesión de Sturm para separar las distintas raíces del polinomio.

14. Cuestión de examen de febrero de 2004:
    Dada la ecuación $x^2 - e^{2x} = -1/2$,

    (a) Demuestra que tiene una **única** solución en el intervalo $[-2, 0]$.

    (b) Encuentra una aproximación a la solución mediante tres iteraciones del método del punto fijo partiendo del punto $x_0 = -1.5$

15. Cuestión de examen de junio de 2004:
    Dada la función $f(x) = \cos(x) - x$,

    (a) Calcula el número de pasos necesarios para encontrar una raíz de $f(x)$ con un error menor que 10^{-2} con el método de la bisección en el intervalo $[0.5, \pi/4]$. Aplica el método con el número de pasos calculado.

(b) Aplica tres iteraciones del método de la secante para encontrar una raíz de $f(x)$ en el intervalo $[0.5, \pi/4]$.

16. Cuestión de examen de septiembre de 2004:
Dado el polinomio $P(x) = -x^3 + x^2 - 2x + 5$,

(a) Utiliza el método de Sturm para separar las raíces en intervalos disjuntos.

(b) Empezando con $x = 6$ aplica un paso del método de Newton haciendo uso del algoritmo de Horner.

Sistemas de ecuaciones lineales

17. Para el conjunto de ecuaciones.

$$\begin{aligned} 7x_1 - 3x_2 + 8x_3 &= -49 \\ x_1 - 2x_2 - 5x_3 &= 5 \\ 4x_1 - 6x_2 + 10x_3 &= -84 \end{aligned}$$

a) Calcula la inversa de la matriz de los coeficientes haciendo uso de las operaciones fundamentales. b) Utiliza la inversa para obtener las soluciones del sistema. c) Calcula el número de condición de la matriz con las distintas normas que conozcas.

18. Dado el sistema de ecuaciones

$$\begin{aligned} 0.5x_1 - x_2 &= -9.5 \\ 0.28x_1 - 0.5x_2 &= -4.72 \end{aligned}$$

a) Resuélvelo por Gauss. b) Sustituye a_{11} por 0.55 y resuélvelo de nuevo. c) Interpreta los resultados anteriores en términos del condicionamiento del sistema. d) Calcula el número de condición de la matriz del sistema con la norma infinito.

19. Dado el sistema

$$\begin{aligned} -12x_1 + x_2 - 7x_3 &= -80 \\ x_1 - 6x_2 + 4x_3 &= 13 \\ -2x_1 - x_2 + 10x_3 &= 92 \end{aligned}$$

a) Resuélvelo por eliminación gaussiana simple usando dígitos con dos cifras decimales. b) Encuentra las matrices triangular inferior L, triangular superior U y de permutaciones P tal que $PA = LU$, siendo A la matriz de los coeficientes. c) Sustituye los resultados en las ecuaciones y, si es necesario, utiliza el método del refinamiento iterativo para mejorar la solución.

20. a) Resuelve los siguientes sistemas de ecuaciones mediante eliminación gaussiana con pivoteo parcial empleando dígitos con dos decimales. b) Encuentra las matrices P, L y U tal que $PA = LU$ como en el ejercicio anterior. c) Comprueba los resultados, y si no son exactos, emplea el refinamiento iterativo para mejorar la solución.

$$\text{1).}\quad \begin{aligned} 4x_1 + 5x_2 - 6x_3 &= 28 \\ 2x_1 - 7x_3 &= 29 \\ -5x_1 - 8x_2 &= -64 \end{aligned} \qquad \text{2).}\quad \begin{aligned} 3x_2 - 13x_3 &= -50 \\ 2x_1 - 6x_2 + x_3 &= 44 \\ 4x_1 + 8x_3 &= 4 \end{aligned}$$

21. Dado el sistema

$$\begin{pmatrix} 3 & 2 & 100 \\ -1 & 3 & 100 \\ 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 105 \\ 102 \\ 2 \end{pmatrix}$$

resuélvelo mediante descomposición LU de Gauss: a) sin escalamiento y b) con escalamiento previo de filas.

22. Dados

$$A = \begin{pmatrix} 4 & -3 & 0 \\ 2 & 2 & 3 \\ 6 & 1 & -6 \end{pmatrix} \; b = \begin{pmatrix} -7 \\ -2 \\ 6 \end{pmatrix} \; b' = \begin{pmatrix} 14 \\ 9 \\ -8 \end{pmatrix}$$

resuelve $Ax = b$ y $Ax = b'$ mediante una descomposición LU de Gauss. Utiliza 3 decimales en los cálculos.

23. Resuelve el sistema siguiente mediante descomposición LU:

$$\begin{aligned} -4x_1 + 2x_2 &= 0 \\ x_1 - 4x_2 &= -4 \\ x_2 - 4x_3 + x_4 &= -11 \\ x_3 - 4x_4 + x_5 &= 5 \\ 2x_4 - 4x_5 &= 6 \end{aligned}$$

24. Resuelve el sistema

$$\begin{pmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -4 \\ 0 \\ 4 \\ -4 \end{pmatrix}$$

mediante la descomposición de Cholesky. ¿Es la matriz de los coeficientes definida positiva?

25. Dado el sistema

$$\begin{pmatrix} 2 & 0 & -1 \\ -2 & -10 & 0 \\ -1 & -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -12 \\ 2 \end{pmatrix}$$

comprueba si se cumple alguna de las condiciones suficientes para que funcionen los métodos de Jacobi y Gauss-Seidel. Si se cumplen, realiza dos iteraciones para cada método tomando como valor inicial $(0, 0, 0)$.

26. Repite el ejercicio anterior para el sistema

$$\begin{aligned} x_1 - 2x_2 &= -1 \\ 3x_1 + x_2 &= 4 \end{aligned}$$

Intenta resolver el sistema con los métodos de Jacobi y Gauss-Seidel haciendo algún pequeño cambio en las ecuaciones.

27. Cuestión de examen de febrero de 2004:
Aplica el método de Jacobi a la resolución del sistema:

$$\begin{aligned} 7x - y + 4z &= 8 \\ 3x - 8y + 2z &= -4 \\ 4x + y - 6z &= 3 \end{aligned}$$

empleando 3 iteraciones y partiendo del punto $(0, 0, 0)$. ¿Es convergente el método?

28. Cuestión de examen de junio de 2004:
Resuelve por el método de Cholesky el sistema de ecuaciones:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 7 \\ 2x_1 + 5x_2 + 4x_3 &= 9 \\ 3x_1 + 4x_2 + 14x_3 &= 33 \end{aligned}$$

29. Cuestión de examen de septiembre de 2004:
Resuelve el siguiente sistema dos veces. Primero, utilizando la eliminación gaussiana simple y encontrando la factorización $A = LU$. Después, utilizando la eliminación gaussiana con pivoteo y encontrando la factorización $PA = LU$.

$$\begin{aligned} -x_1 + x_2 - 4x_3 &= 0 \\ 2x_1 + 2x_2 &= 1 \\ 3x_1 + 3x_2 + 2x_3 &= \tfrac{1}{2} \end{aligned}$$

Aproximación de funciones. Interpolación

30. De la función $f(x,t)$ que representa una onda que se desplaza sobre el plano se conocen cinco posiciones en un determinado instante:

$$(-1,0),(-0.5,-1),(0,0),(0.5,1),(1,0).$$

(a) Encuentra un polinomio que pase por esos puntos.

(b) Encuentra una función formada por trozos de recta que pase por esos puntos.

(c) Sabiendo que las derivadas de $f(x)$ en el mismo instante en los puntos $-1, -0.5, 0, 0.5$ y 1 son $-\pi, 0, \pi, 0$ y $-\pi$, calcula un polinomio que cumpla todas las condiciones.

(d) Calcula el spline natural para ese conjunto de datos.

31. ¿Cuál es el polinomio interpolador de $f(x) = x^3 - 2x - 5$ en los puntos $-0, 1, 2, 4$.?

32. Se ha estudiado el proceso de recepción de información de un ordenador y se ha observado que en los instantes $0, 1, 2$ medidos en segundos, la cantidad de información recibida era $0, 51, 104$ megabites y se recibía a una velocidad de $0, 52, 54$ megabites por segundo, respectivamente. Encuentra una función que se aproxime a la que representa el proceso.

33. Dada la función $f(x) = e^x$ y los puntos $0, 1, 2.$,

(a) Encuentra su polinomio de interpolación en los puntos y estima el error cometido.

(b) Repite el apartado anterior con los nodos de Chebyshev en el intervalo $[0, 2]$.

(c) Encuentra una función lineal a trozos que se aproxime a $f(x)$ en los puntos dados.

(d) Encuentra un polinomio que interpole a $f(x)$ y a sus derivadas en los puntos dados.

(e) Calcula el spline natural en los puntos dados.

34. La función $f(x) = L(x)$ se ha evaluado en varios puntos, obteniéndose lo siguientes datos:

$$\begin{array}{l|l} x_0 = 1 & f(x_0) = 0 \\ x_1 = 4 & f(x_1) = 1.3863 \\ x_2 = 6 & f(x_2) = 1.7918 \end{array}$$

(a) Utilícense los datos anteriores para calcular el polinomio de interpolación y el spline natural. Estímese el valor de $L(2)$.

(b) Agregando un cuarto punto $x_3 = 5, f(x_3) = 1.6094$, háganse de nuevo los cálculos y estimaciones.

35. Calculando los valores de la función $f(x) = \sqrt{x}$ en 1,4 y 9,

(a) Halla el polinomio interpolador de grado menor o igual que 2 por el método de Newton que pase por esos puntos.

(b) Halla el polinomio interpolador en los nodos de Chebyshev en [1,9].

(c) Aplica el método de Hermite en los puntos iniciales y en sus derivadas.

(d) Calcula el spline natural.

36. Cuestión de examen de junio de 2004:
Dada la función $f(x) = e^x$, determina los cuatro nodos de Chebyshev en $[0, 1]$, el polinomio interpolador en dichos nodos y un valor aproximado de $e^{0.5}$ mediante dicho polinomio. Realiza los cálculos con 4 decimales.

Polinomios de Chebysev: $T_0(x) = 1$, $T_1(x) = x$, $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$.

37. Cuestión de examen de septiembre de 2004:
Determina si la función

$$f(x) = \begin{cases} 1 + x - x^3 & \text{si} \quad x \in [0, 1] \\ 1 - 2(x-1) - 3(x-1)^2 + 4(x-1)^3 & \text{si} \quad x \in [1, 2] \\ 4(x-2) + 9(x-2)^2 - 3(x-2)^3 & \text{si} \quad x \in [2, 3] \end{cases}$$

es el spline cúbico natural que interpola a los puntos de la tabla:

x	0	1	2	3
y	1	1	0	10

Diferenciación e integración numérica

38. De la función $f(x) = e^x - 2x^2 + 3x - 1$ se obtienen los siguientes datos:

x	0.0	0.2	0.4	0.6
$f(x)$	0.00000	0.74140	1.3718	1.9021

(a) Utilizando la fórmula más adecuada de entre las siguientes:
$f'(x) = \frac{f(x+h)-f(x)}{h}$ $f'(x) = \frac{f(x+h)-f(x-h)}{2h}$ $f''(x) = \frac{f(x+h)-2f(x)+f(x-h)}{h^2}$,
calcula valores aproximados de la primera y segunda derivadas de los puntos.

(b) Calcula los errores reales y las cotas de error por medio de las fórmulas de error para las estimaciones anteriores.

(c) Encuentra una aproximación de $f'(0.2)$ haciendo uso de la interpolación de la función en los cuatro puntos.

(d) Sabiendo que $f(-0.2) = -0.861267$, aplica la interpolación de Richardson para calcular una aproximación de $f'(0.2)$.

39. De una función $f(x)$ se tienen los siguientes datos:

x	0.0	0.5	1	1.5	2	2.5	3
$f(x)$	2.25	3.75	5	6	5.5	6	7.25

Calcula un valor aproximado de $\int_0^3 f(x)dx$ utilizando los siguientes métodos:

(a) Trapecios.

(b) Simpson 1/3.

(c) Simpson 3/8.

(d) Trapecios compuestos.

(e) Simpson compuesto.

(f) Romberg.

40. Determina el número de subintervalos que deben tomarse para aproximar el valor de la integral $\int_{0.5}^{1} cos(\sqrt{x})dx$ con un error menor que $\frac{1}{2}10^{-2}$

(a) con el método de los trapecios compuesto,

(b) con el método de Simpson compuesto.

41. Repite el ejercicio anterior para

(a) $\int_1^2 Ln(x)dx$ (b) $\int_2^3 \frac{e^x}{x}dx$

42. Calcula $\int_0^1 sen(x^2)dx$:

(a) con un error menor que 10^{-2} con el método de los trapecios compuestos

(b) con un error menor que 10^{-2} con el método de Simpson compuesto,

(c) con el algoritmo de Romberg hasta $R(4,4)$,

(d) con el método de cuadratura adaptativa con nivel de tolerancia 10^{-3}.

43. Repite el ejercicio anterior con $\int_1^2 e^{1/x}dx$.

44. (a) Calcula la integral $\int_0^1 sen(x^2)dx$ mediante cuadratura gaussiana de dos términos.

(b) Plantea cómo resolver el problema de calcular la $\int_0^1 sen(x^2)dx$ mediante cuadratura gaussiana de tres términos. Calcula la integral sabiendo que los nodos (t_i) en este caso son las raíces del polinomio $\frac{5x^3-3x}{2}$.

45. Aplica algún método de integración numérica para resolver las siguientes integrales:

(a) $\int_0^1 \frac{sen(x)}{x}dx$ (b) $\int_{-1}^1 \frac{cos(x)-e^x}{sen(x)}dx$ (c) $\int_1^\infty (xe^x)^{-1}dx$.

46. Determina los valores de A, B y C que hagan que la fórmula

$$\int_0^2 xf(x) \approx Af(0) + Bf(1) + Cf(2)$$

sea exacta para todos los polinomios de grado tan alto como sea posible. ¿Cuál es el grado máximo?

47. Evalúa la integral de $x^3 sen(x)$ entre $x = 0$ y $x = 2$ usando un valor de tolerancia inferior a 0.1% de la respuesta verdadera (3.791197).

48. Cuestión de examen de febrero de 2004:
Dada la función $f(x) = L(x)$ calcula $\int_2^3 L(x)dx$ empleando el método de cuadratura de Gauss con tres términos y sabiendo que los polinomios de Legendre se calculan así:

$$L_0(x) = 1,\ L_1(x) = x,$$

$$L_{n+1}(x) = \frac{(2n+1)xL_n(x) - nL_{n-1}(x)}{n+1} \quad \forall\ n \geq 1.$$

49. Cuestión de examen de junio de 2004:
Encuentra los coeficientes A_0, A_1, A_2 y A_3 para que la fórmula

$$\int_{-1}^{3} f(x)dx = A_0 f(-1) + A_1 f(0) + A_2 f(1) + A_3 f(2)$$

sea exacta para los polinomios de grado menor o igual que 3 y calcula un valor aproximado de $\int_{-1}^{3}(x^3 + x)dx$ haciendo uso de dicha fórmula.

50. Cuestión de examen de septiembre de 2004:
Encuentra los coeficientes A_1 y A_2 para que la fórmula

$$\int_0^{2\pi} f(x)dx = A_1 f(0) + A_2 f(\pi)$$

sea exacta para cualquier función que tenga la forma $a + b\cos(x)$.
Calcula $\int_0^{2\pi} 2 + 3\cos(x)dx$ haciendo uso de dicha fórmula.

Resolución numérica de ecuaciones diferenciales ordinarias

51. Cuestión de examen de junio de 2004:
Aplicando el método de Taylor de grado 2 con dos pasos, calcula valores aproximados de $y(0.1)$ y $y(0.2)$, sabiendo que

$$y' + 2y = x^2 + 2x \qquad y(0) = 1$$

52. Cuestión de examen de febrero de 2005:

Resuelve la ecuación $y' = y^2 - t^2 + 1$, $y(0) = 0.5$, siendo $0 \leq t \leq 1.2$, mediante el método de Taylor de grado 2 con 5 pasos.

53. (a) Deduce la fórmula de Adams-Bashforth de orden 4:

$$y_{i+1} = y_i + \frac{h}{24}\big[55f(x_i, y_i) - 59f(x_{i-1}, y_{i-1}) + \\ + 37f(x_{i-2}, y_{i-2}) - 9f(x_{i-3}, y_{i-3})\big]$$

y úsala para resolver la ecuación

$$y' = \frac{yx - x^2}{x^2} \qquad y(1) = 2$$

con $h = 1/128$

(b) Repite el apartado anterior para la fórmula de Adams-Moulton de orden 4:

$$y_{i+1} = y_i + \frac{h}{24}\big[9f(x_{i+1}, y(i+1)) + 19f(x_i, y_i) - \\ - 5f(x_{i-1}, y_{i-1}) + f(x_{i-2}, y_{i-2})\big]$$

54. Usa la identidad

$$y(x_{i+2}) - y(x_i) = \int_{x_i}^{x_{i+2}} y'(x)dx$$

y la fórmula de cuadratura de Simpson 1/3 para obtener la fórmula del siguiente método multipaso

$$y_{i+2} = y_i + \frac{h}{3}\big[f(x_i, y_i) + 4f(x_{i+1}, y_{i+1}) + f(x_{i+2}, y_{i+2})\big].$$

55. Aplica el método clásico de Runge-Kutta para estimar 10 valores aproximados de la solución de la ecuación siguiente en el intervalo $[0, 1]$

$$y' + 2y = x^2 + 2x \qquad y(0) = 1$$

56. A partir de la identidad

$$y(x_{i+1}) - y(x_i) = \int_{x_i}^{x_{i+1}} y'(x)dx$$

deduce el método de Runge-Kutta de orden 2

$$y_{i+1} = y_i + h\big[\frac{1}{2}f(x_i, y_i) \ + \frac{1}{2}f(x_{i+1}, y_i + hf(x_i, y_i))\big]$$