Métodos Numéricos

Sergio Plaza[1]

Segundo semestre de 2007. Versión Preliminar en Revisión
Si el lector detecta errores, que ciertamente hay muchos,
le agradecere avisarme al e-mail: splaza@lauca.usach.cl

[1]Depto. de Matemática, Facultad de Ciencias, Universidad de Santiago de Chile, Casilla 307–Correo 2. Santiago, Chile. e–mail splaza@lauca.usach.cl, Homepage http://lauca2.usach.cl/~splaza

Contenidos

Capítulo 1

Teoría de Error

1.1 Representación punto flotante de números

Cada $x \in \mathbb{R}$, con $x \neq 0$, puede ser representado de modo único en su forma normalizada

$$x = \sigma \cdot \overline{x} \cdot 10^e \tag{1.1}$$

donde $\sigma = \pm 1$ es el *signo*, $e \in \mathbb{Z}$ es el *exponente* y $0.1 \leqslant \overline{x} < 1$ es la *mantisa* de x, es decir,

$$\overline{x} = 0.a_1a_2\ldots = a_1 10^{-1} + a_2 10^{-2} + \cdots + a_k 10^{-k} + \cdots .$$

con $a_1 \neq 0$. En lo que sigue, siempre consideramos los números en su forma normalizada,

Ejemplo 1

1. $x = 12.462 = 0.12462 \times 10^2$
2. $\pi = 3.14159265\ldots = 0.314159265\ldots \times 10^1$
3. $\sqrt[3]{2} = 1.2599210\ldots = 0.12599210\ldots \times 10^1$
4. $e = 0.27182818\ldots \times 10^1$
5. $\ln(2) = 0.6931471\ldots \times 10^0$
6. $x = 0.0000458 = 0.458 \times 10^{-4}$

La representación *decimal punto flotante* de un número $x \in \mathbb{R}$, con $x \neq 0$, es básicamente aquella dada en (1.1), con limitación del número de dígitos en la mantisa $\overline{x}$ y el tamaño del exponente e.

Si limitamos, por ejemplo, el número de dígitos de $\overline{x}$ a 6 y e entre -99 y $+99$, decimos que una computadora con tal representación tiene una aritmética decimal de punto flotante de 6 dígitos. Como consecuencia de la limitación del número de dígitos de $\overline{x}$, ningún número puede tener más que sus seis primeros dígitos almacenados con precisión.

Ahora consideremos un número x en su forma binaria normalizada, podemos escribir

$$\boxed{x = \sigma \cdot \overline{x} \cdot \ 2^e} \tag{1.2}$$

donde $\sigma = \pm 1$ es el signo, $e \in \mathbb{Z}$ es el exponente, y la mantisa $\overline{x}$ es un número en forma binaria con $(0.1)_2 \leqslant \overline{x} < 1$, esto es equivalente, en la forma decimal, a $\frac{1}{2} \leqslant \overline{x} < 1$, es decir,

$$\overline{x} = (0.a_1a_2\ldots)_2 = a_1 2^{-1} + a_2 2^{-2} + \cdots + a_k 2^{-k} + \cdots$$

con $a_1 = 1$ y $a_i = 0$ o 1 para $i \geqslant 2$.

Ejemplo 2 Sea $x = (1101.10111)_2$. Escribiendo este en la forma canónica, tenemos

$$\begin{aligned} x &= 2^3 + 2^2 + 2^0 + 2^{-1} + 2^{-3} + 2^{-4} + 2^{-5} \\ &= (2^{-1} + 2^{-2} + 2^{-4} + 2^{-5} + 2^{-7} + 2^{-8} + 2^{-9}) \times 2^4, \end{aligned}$$

de donde tenemos que $\sigma = +1$, $e = 4 = (100)_2$ y $\overline{x} = (0.110110111)_2$.

La representación *punto flotante de un número binario* x consiste básicamente en aquella dada en (1.2), con una restricción en el número de dígitos de $\overline{x}$ y en el tamaño del exponente e.

Para no estar restrictos a números en forma decimal o binaria, trabajaremos representaciones de números en una base $\beta > 1$, en general, sólo consideramos el caso en que β es un número entero positivo mayor que 1. Lo que desarrollaremos es completamente análogo a los casos bien conocidos cuando $\beta = 2$ (representación binaria) y $\beta = 10$ (representación decimal). Usar una base $\beta > 1$ arbitraria nos permite tratar en forma unificada conceptos y propiedades sin referirnos cada vez a una base determinada.

Dado un número entero $\beta > 1$, cada $x \in \mathbb{R}$, con $x \neq 0$, puede ser representado en la base β de modo único en su forma normalizada como

$$\boxed{x = \sigma\,(0.a_1a_2\ldots a_k\ldots)_\beta \times \beta^e = \sigma\,(a_1\beta^{-1} + a_2\beta^{-2} + \cdots + a_k\beta^{-k} + \cdots)\beta^e} \qquad (1.3)$$

donde $\sigma = \pm 1$ es el signo, $e \in \mathbb{Z}$ es el exponente, con $a_1 \neq 0$ y $0 \leqslant a_i \leqslant \beta - 1$, para todo $i = 1, 2, \ldots$. La notación $\overline{x} = (0.a_1a_2\ldots a_k\ldots)_\beta$ significa

$$\begin{aligned} \overline{x} &= (0.a_1a_2\ldots a_k\ldots)_\beta \\ &= \frac{a_1}{\beta} + \frac{a_2}{\beta^2} + \cdots + \frac{a_k}{\beta^k} + \cdots \\ &= a_1\beta^{-1} + a_2\beta^{-2} + \cdots + a_k\beta^{-k} + \cdots. \end{aligned}$$

Veremos en la próxima sección cómo obtener el número de máquina que representa a un número real $x \neq 0$.

1.2 Corte y redondeo

La mayoría de los números reales no pueden ser representados de forma exacta en la representación punto flotante, y por lo tanto deben ser aproximados por un número representable en la máquina, de la mejor forma que nos sea posible.

Dado $x \in \mathbb{R}$, denotamos por $fl(x)$ a la representación de x en la máquina, llamada comúnmente *representación punto flotante* (float point) de x. Existen principalmente dos maneras de obtener $fl(x)$ a partir de x, ellas son *corte* y *redondeo*, las cuales veremos a seguir.

Como vimos en las sección anterior, un número real $x \neq 0$ puede ser representado en forma exacta en una base entera $\beta > 1$ de la forma siguiente

$$x = \sigma \left(0.a_1 a_2 \ldots a_k a_{k+1} \ldots\right)_\beta \beta^e \tag{1.4}$$

donde, $\sigma = \pm$ es el signo, $a_1 \neq 0$ (x está normalizado), y $e \in \mathbb{Z}$ es el exponente.

El problema ahora es decidir cómo obtener la representación punto flotante, digamos con k dígitos para un número real.

Supongamos que $e \in \mathbb{Z}$ está acotado, digamos $L \leqslant e \leqslant U$.

Definición 1.1 *Sea $x \in \mathbb{R}$, $x \neq 0$, con representación en una base entera $\beta > 1$ dada por (1.4). La representación punto foltante por corte en el dígito k de x es dada por*

$$\boxed{fl(x) = \sigma \left(0, a_1 a_2 \ldots a_k\ \right)_\beta \beta^e} \tag{1.5}$$

donde $L \leqslant e \leqslant U$.

La razón para introducir corte, es que algunas máquinas usan corte después de cada operación aritmética.

Definición 1.2 *Sea $x \in \mathbb{R}$, $x \neq 0$, con representación en una base entera $\beta > 1$ dada por (1.4). La representación punto flotante por redondeo en el dígito k de x es dada por*

$$\boxed{fl(x) = \begin{cases} \sigma \left(0.a_1 a_2 \ldots a_k\right)_\beta \beta^e, & si \quad 0 \leqslant a_{k+1} < \frac{\beta}{2} \\ \\ \sigma \left[\left(0.a_1 a_2 \ldots a_k\right)_\beta + \left(0.00 \ldots \underset{\text{posición } k}{1}\right)_\beta \right] \beta^e, & si \quad \frac{\beta}{2} \leqslant a_{k+1} < \beta \end{cases}} \tag{1.6}$$

donde $L \leqslant e \leqslant U$.

Observación. Esta definición formal no es otra que la que usamos comúnmente al redondear cantidades a diario, por ejemplo al calcular el promedio final de las notas obtenidas en este curso.

Ejemplo 3 Los siguiente números están representados en forma decimal.

1. Tenemos que $\pi = 3.14159265\ldots = 0.314159265\ldots \times 10^1$. Esta es la representación exacta normalizada de π en forma decimal. Ahora usando corte en el dígito 4 la representación es dada por $\pi_A = 0.3141 \times 10^1$, y usando redondeo en el dígito 4 la represenatción es dada por $\pi_A = 0.3142 \times 10^1$, pues el dígito $a_5 = 5$.

2. Usando corte y redondeo a 3 dígitos para $x = 12.462 = 0.12462 \times 10^2$, tenemos $x_A = 0.124 \times 10^2$ en su representación por corte de x, y $x_A = 0.125 \times 10^2$ en su representación por redondeo de x.

3. Usando corte y redondeo a 5 dígitos para $x = 22.4622 = 0.224622 \times 10^2$, tenemos $x_A = 0.22462 \times 10^2$ en la representación por corte y en la representación por redondeo.

Para la mayoría de los números reales se tiene que $fl(x) \neq x$. Por esta razón introducimos el *error relativo* o *porcentaje de error*

Definición 1.3 *Sea $x \in \mathbb{R}$, con $x \neq 0$. El error relativo (o porcentaje de error) de x es dado por*

$$\boxed{E_R(x) = \frac{x - fl(x)}{x} = -\varepsilon} \tag{1.7}$$

donde

a) $-\beta^{-k+1} \leqslant \varepsilon \leqslant 0$, *(es decir,* $0 \leqslant -\varepsilon \leqslant \beta^{-k+1}$*) cuando* $fl(x)$ *es obtenido por corte desde* x*, y*

b) $-\frac{1}{2}\beta^{-k+1} \leqslant \varepsilon \leqslant \frac{1}{2}\beta^{-k+1}$ *cuando* $fl(x)$ *es obtenido por redondeo desde* x*.*

Lo anterior puede ser escrito en la forma

a') $0 \leqslant \left|\frac{x-fl(x)}{x}\right| \leqslant \beta^{-k+1}$ cuando $fl(x)$ es obtenido por corte desde x

b') $0 \leqslant \left|\frac{x-fl(x)}{x}\right| \leqslant \frac{\beta^{-k+1}}{2}$ cuando $fl(x)$ es obtenido por redondeo desde x.

De la fórmula $\frac{x-fl(x)}{x} = -\varepsilon$ obtenemos que $fl(x) = (1+\varepsilon)\, x$, donde ε es dado por a) o por b) anteriores, dependiendo del caso. Luego $fl(x)$ puede ser considerado como una pequeña perturbación relativa de x. La fórmula (1.7) nos permite tratar los efectos de corte y redondeo en las operaciones aritméticas del computador.

Se define el error de $fl(x)$ respecto a x como

$$\boxed{E(fl(x)) = x - fl(x)}$$

Notemos que desde la fórmula b'), para redondeo a k dígitos, tenemos que

$$|E(fl(x))| = |x - fl(x)| \leqslant \frac{1}{2}\beta^{-k+1}|x|\,,$$

ahora, como $x = \sigma \cdot \overline{x} \cdot \beta^e$ y $|\bar{x}| < 1$, obtenemos

$$|x - fl(x)| \leqslant \frac{1}{2}\beta^{-k+1}|x| = \frac{1}{2}\beta^{-k+1}|\overline{x}|\beta^e \leqslant \frac{1}{2}\beta^{e-k+1}\,, \tag{1.8}$$

es decir,

$$\boxed{|E(fl(x))| = |x - fl(x)| \leqslant \frac{1}{2}\beta^{e-k+1}} \tag{1.9}$$

De modo análogo, para corte a k dígitos, obtenemos que

$$\boxed{|E(fl(x))| = |x - fl(x)| \leqslant \beta^{e-k+1}} \tag{1.10}$$

1.2.1 Precisión de la representación punto flotante

Introduciremos ahora dos medidas que nos darán alguna idea de una posible precisión de la representación de punto flotante en las máquinas.

1.2.2 Unidad de redondeo de un computador

La unidad de redondeo de un computador es un número δ que satisface

1. δ es un número punto flotante positivo, y
2. δ es el menor número punto flotante tal que $fl\,(1+\delta) > 1$, es decir, si $\delta_0 < \delta$ entonces $fl\,(1+\delta_0) = 1$.

Luego, para cualquier otro número punto flotante positivo $\overline{\delta} < \delta$, se tiene que $fl\left(\overline{\delta}+1\right) = 1$, así dentro de la aritmética del computador $1+\overline{\delta}$ y 1 son iguales. Note que δ mide *"el ancho del cero"* en una máquina.

Esto da una medida de cuántos dígitos de precisión son posibles en la representación de un número. La unidad de redondeo δ para una máquina con k dígitos en la mantisa, es dada por

$$\boxed{\delta = \begin{cases} \beta^{-k+1} & \text{definición por corte de } fl\,(x) \\ \frac{1}{2}\beta^{-k+1} & \text{definición por redondeo de } fl\,(x) \end{cases}}$$

Por ejemplo, para una máquina con aritmética binaria con k dígitos y redondeo para la aritmética tenemos que $\delta = 2^{-k}$ (y por corte se tiene que $\delta = 2^{-k+1}$). En efecto, tenemos que

$$\begin{aligned} 1 + 2^{-k} &= \left[(0.100\ldots)_2 + \left(0.00\ldots0\underset{\text{posición } k+1}{1}0\right)_2\right] 2^1 \\ &= (0.10\ldots010)_2\, 2^1 \end{aligned} \tag{1.11}$$

Tomando $fl\left(1+2^{-k}\right)$ notamos que en la posición $k+1$ de la mantisa existe un 1. Luego

$$fl\left(1+2^{-k}\right) = (0.10...01)_2\, 2^1$$

por lo tanto $fl\left(1+2^{-k}\right) > 1$, y también tenemos que $fl\left(1+2^{-k}\right) \neq 1+2^{-k}$.

El hecho de que δ no puede ser menor que 2^{-k} se sigue de (1.11). Ahora, si $\overline{\delta} < \delta$ entonces $1+\overline{\delta}$ tiene un cero en la posición $k+1$ de la mantisa, y por definición se tiene entonces que $fl\left(1+\overline{\delta}\right) = 1$.

1.2.3 Mayor entero positivo representable en forma exacta

Una segunda medida de precisión posible es dada por el mayor entero positivo M para el cual se tiene que: para todo entero m con $0 \leqslant m \leqslant M$ tenemos que $fl(m) = m$, esto significa que $fl(M+1) \neq M+1$, es decir, M es el mayor entero positivo que es posible de representar en forma exacta en la aritmética de máquina que estamos usando.

Es fácil ver que $M = \beta^k$ en un computador con aritmética de k dígitos en la mantisa.

Observación. Los números M y δ varían de computador en computador. Todo usuario de computador debería conocer los números δ y M de su máquina, pues ellos representan el ancho del cero y el mayor entero representable en forma exacta en la máquina, respectivamente, lo cual da las medidas de precisión del computador en uso.

1.2.4 Underflow–overflow

Si una máquina trabaja con aritmática de k dígitos. Dado un número real $x \in \mathbb{R}$, $x \neq 0$, si al representar x en nuestra aritmática, las cotas L o U para los exponentes son violados, entonces el número de máquina asociado a x

$$fl(x) = \sigma (0.a_1 a_2 \ldots \tilde{a}_k)_\beta \beta^e ,$$

donde el dígito $\tilde{a}_k$ es dado por corte o redondeo, no puede ser representado en tal máquina.

Por ejemplo, el menor número positivo punto flotante es $x_L = (0.10\ldots0)_\beta \beta^L = \beta^{L-1}$, y usando la notación $\gamma = \beta - 1$, el mayor número positivo punto flotante es $x_U = (0.\gamma\ldots\gamma)_\beta \beta^U = \left(1 - \beta^{-k}\right)\beta^U$. Luego todo número punto flotante positivo x, representable en la máquina, deben satisfacer $x_L \leqslant x \leqslant x_U$.

Si durante las operaciones aritméticas obtenemos como resultado un valor de x tal que $|x| > |x_U|$ entonces aparecerá un *error fatal*, este error es denominado *overflow*, y los cálculos se detienen. Por otra parte, si $0 < |x| < x_L$ entonces $fl(x) = 0$ y los cálculos continuaran. Este error se denomina error de *underflow*.

1.3 Definición y fuentes de errores

Ahora daremos una clasificación de la principal manera en la cual el error es introducido en la solución de un problema, incluyendo algunos que caen fuera del ámbito de la matemática.

Al resolver un problema, buscamos una solución exacta o verdadera, la cual denotamos por x_T. Aproximaciones son usualmente hechas en la solución de un problema y como resultado obtenemos una solución aproximada x_A.

Definición 1.4 *El error en x_A respecto de x_T es dado por*

$$\boxed{E(x_A) = x_T - x_A} \tag{1.12}$$

El error absoluto en x_A es dado por $E_A(x_A) = |E(x_A)|$.

Para muchos propósitos es preferible estudiar el *porcentaje* o *error relativo* en x_A respecto de x_T, el cual es dado por

$$\boxed{E_R\left(x_A\right) = \frac{x_T - x_A}{x_T}, \quad x_T \neq 0} \tag{1.13}$$

Ejemplo. Sea $x_T = e = 2.7182818\ldots$ representado en forma exacta y sea $x_A = \frac{19}{7} = 2.71428587$ una aproximación. Tenemos $E\left(x_A\right) = 0.003996$ y el error relativo es $E_R\left(x_A\right) = 0.0147\ldots$

Ejemplo. Sean $x = 0.3721478693$ e $y = 0.3720230572$, si consideramos aritmética de puntos flotantes con 5 dígitos y redondeo, obtenemos

$$\begin{array}{l} fl\left(x\right) = 0.37215 \\ fl\left(y\right) = 0.37202 \\ E(fl(x) - fl(y)) = fl\left(x\right) - fl\left(y\right) = 0.00013 \\ |E_R\left(fl(x) - fl(y)\right)| = \left|\frac{x-y-(fl(x)-fl(y))}{x-y}\right| = \left|\frac{0.0001248121-0.00013}{0.0001248121}\right| \approx 4\%\,. \end{array}$$

Algunas veces usamos el *número de dígitos significativos* como una forma medir la precisión de la representación, el cual es calculado como sigue. Dados x_T y x_A como antes, si

$$\boxed{|E_R(x_A)| = \left|\frac{x_T - x_A}{x_T}\right| \leqslant 5 \times 10^{-m-1}} \tag{1.14}$$

con $m \in \mathbb{N}$. Decimos que x_A tiene al menos m dígitos significativos respecto de x_T.

1.3.1 Fuentes de error

Cuando resolvemos un problema científico–matemático, el cual envuelve cálculos computacionales, usualmente aparecen errores relacionados a este proceso, los cuales pueden ser clasificados de la siguiente forma.

E1) Errores de Modelación. Ecuaciones matemáticas son usadas para representar realidades físicas, este proceso se denomina modelación. La modelación introduce errores en el problema real que se está tratando de resolver, y caen fuera del alcance del análisis numérico.

Ilustramos la noción anterior con algunos modelos.

1) *Modelos poblacionales*. Modelos predador–presa, este es usado para el control de poblaciones, por ejemplo colonia de bacterias, y son expresados por ecuaciones relacionando el número de individuos existentes en cada especie, y se trata de mantener el equilibrio de modo que no se llegue a la extinción de una de las especies, que puede ser el alimento de la otra.

   El crecimiento de la población–versus la cantidad de alimento existente y que se puede producir, también entra en esta categoría.

2) *Modelos de la Física*. Ecuaciones son usadas para representar realidades, por ejemplo "la velocidad es igual la razón de la distancia recorrida por el tiempo empleado para hacerlo", las leyes de la gravitación, las leyes de la electricidad, etc.

3) *Modelos de tránsito en la ciudad*. Nos podemos plantear el problema de tránsito en Santiago y los problemas relacionados con éste , y nos podemos preguntar si detrás de ellos existe un modelo que lo respalde.

4) *Modelos de Aerodinámica*. Problemas de este tipo son la fabricación de aviones y autos que sean aerodinámicamente bueno.

E2) Desatinos o Equivocaciones. Estos son la segunda fuente de errores, muy familiar, estan más relacionados con problemas de programación. Para detectar tales errores es importante tener alguna forma de chequear la precisión de la salida después de la ejecución del programa, los cuales son llamados programas test y están basados en resultados conocidos de antemano.

E3) Datos Físicos. Estos contienen de modo natural errores observacionales. Como estos contienen errores, cálculos basados en ellos también los tendrán. El Análisis numérico, no puede remover estos errores pero debe sugerir procedimientos de cálculo que minimizan la propagación de ellos.

Por ejemplo, las ecuaciones de la Física sólo consideran parte de las variables que influyen en un fenómenos, despreciando otras que tienen menor influencia, por ejemplo las ecuaciones de la Física escolar, son simplificaciones brutales de la realidad, por otro lado, las ecuaciones de la Teoría de la Relatividad considera variables que a escala pequeña no tienen mayor importancia.

E4) Error de Máquina. Las máquinas naturalmente introducen errores principalmente por corte y redondeo. Estos errores son inevitables cuando se usa la aritmética de punto flotante y ellos forman la principal fuente de error en algunos problemas, por ejemplo soluciones de sistemas de ecuaciones lineales, como veremos en el respectivo capítulo, mostrando que en ejemplos sencillos podemos obtener soluciones absurdas para ecuaciones simples. También están los ejemplos de expresiones matemáticamente equivalentes, pero que al trabajar con aritmética de computador producen resultados diferentes.

E5) Error de Aproximaciones Matemáticas. Estos forman la mayor parte de los errores, estas provienen del hecho que el computador sólo puede trabajar con una cantidad finita de números, así por ejemplo al calcular el área de un círculo debemos usar el número π , pero la verdad es que usamos sólo una aproximación de este número, por lo tanto el resultado obtenido no puede ser exacto por esta razón. Otro ejemplo es que cuando trabajamos con funciones, debemos usar aproximaciones, en general dadas por polinomios de Taylor, por ejemplo las máquina para calcular el valor de $\text{sen}(x)$ para un dado valor de x usa tales aproximaciones, lo mismo ocurre cuando evalúa funciones elementales tales como $\cos(x)$, e^x , etc. Por ejemplo, si deseamos calcular el valor de $\int_0^1 e^{x^2} dx$ tenemos algunos problemas, pues esta integral no es expresable en términos de funciones elementales, pero usando series de Taylor se tiene que

$$e^{x^2} = 1 + x^2 + \frac{x^4}{2!} + \frac{x^6}{3!} + \cdots$$

de donde, usando una aproximación por un polinomio de Taylor obtenemos una aproximación para el valor de integral dado por

$$\int_0^1 e^{x^2} dx \approx \int_0^1 \left(1 + x^2 + \frac{x^4}{2!} + \frac{x^6}{3!}\right) dx = 1 + \frac{1}{3} + \frac{1}{10} + \frac{1}{42} = 1.457142285\ldots.$$

Observación. El valor de la integral $\int_0^1 e^{x^2} dx$ con 20 dígitos es 1.4626517459071816088

1.3.2 Pérdida de dígitos significativos

Para comenzar consideremos el problema de la evaluación de la función

$$f(x) = x\left(\sqrt{x+1} - \sqrt{x}\right)$$

para valores crecientes de x. Por ejemplo, tenemos $f(100) = 100\left(\sqrt{101} - \sqrt{100}\right)$, ahora como $\sqrt{101} - \sqrt{100} = 0.0498752\ldots$, evaluando $f(100)$ tiene que $f(100) = 100\left(\sqrt{101} - \sqrt{100}\right) = 4.987562\ldots$. Ahora si trabajamos con aritmética de 6 dígitos y redondeo, obtenemos los siguientes valores: $\sqrt{101} - \sqrt{100} = 10.0499 - 10 = 0.049900$ luego $f(100) = 100\left(\sqrt{101} - \sqrt{100}\right) = 4.99$ en nuestra aritmética.

El cálculo de $\sqrt{101} - \sqrt{100} = 0.049900$ cuyo valor aproximado es 0.0498756 tiene *perdida del significado del error*. Note que tres dígitos de precisión en $\sqrt{x+1} = \sqrt{101}$ fueron cancelados en la resta con los correspondientes dígitos de $\sqrt{x} = \sqrt{100}$. La perdida de precisión fue un subproducto de la forma de $f(x)$ y la precisión de la aritmética finita usada. Observe que $f(x)$ puede ser reescrita de la siguiente manera

$$f(x) = x\left(\sqrt{x+1} - \sqrt{x}\right) = \frac{x}{\sqrt{x+1} + \sqrt{x}},$$

esta ultima forma no tiene perdida de significado de error en su evaluación, pues

$$f(100) = \frac{100}{10.0499 + 10} = \frac{100}{20.0499} = 4.98756.$$

Observación. Es importante notar que las expresiones $x\left(\sqrt{x+1} - \sqrt{x}\right)$ y $\frac{x}{\sqrt{x+1}+\sqrt{x}}$ para la misma función, son matemáticamente equivalentes, pero numéricamente no lo son, pues la evaluación de ellas en nuestra aritmética nos arrojan resultados distintos.

Observación. Los programadores deben tener en cuenta esta posibilidad de este error y evitar este fenómeno de perdida de dígitos significativos.

La perdida de dígitos significativos es un subproducto de la resta de números parecidos. Esto redunda en "inestabilidad" en los cálculos numéricos.

Ejemplo 4 La ecuación de segundo grado

$$x^2 - 18x + 1 = 0$$

tiene soluciones $x_1 = 9 + \sqrt{80}$ y $x_2 = 9 - \sqrt{80}$. Si consideramos $\sqrt{80}$ con 4 dígitos decimales correctos, obtenemos

$$x_1 \quad = 69 + 8.9443 \pm 0.5 \times 10^{-4} = 17.9443 \pm 0.5 \times 10^{-4}$$

y

$$x_2 \quad = \quad 9 - 8.9443 \pm 0.5 \times 10^{-4} = 0.0557 \pm 0.5 \times 10^{-4}.$$

Luego, la aproximación de x_1 tiene 6 dígitos significativos y x_2 tiene sólo 3. La cancelación es evitada reescribiendo

$$x_2 = \frac{(9 - \sqrt{80})(9 + \sqrt{80})}{9 + \sqrt{80}} = \frac{1}{9 + \sqrt{80}} = \frac{1}{17.9443 \pm 0.5 \times 10^{-4}}$$

y entonces $\bar{x}_2 = \frac{1}{17.9443} = 0.055728002\ldots$.

Ejemplo 5 Consideremos la función $f(x) = \frac{1-\cos(x)}{x^2}$. Si usamos desarrollos en series de Taylor, obtenemos

$$\begin{aligned} f(x) &= \frac{1}{x^2}\left(1-\left(1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+R_6(x)\right)\right) \\ &= \frac{1}{2!}-\frac{x^2}{4!}+\frac{x^4}{6!}-\frac{x^6}{8!}\cos(x), \end{aligned}$$

de donde $f(0) = \frac{1}{2}$, y para $|x| < 0.1$, tenemos que

$$\left|\frac{x^6}{8!}\cos(x)\right| < \frac{10^{-6}}{8!} = 2.5 \times 10^{-11} \quad .$$

Luego la aproximación $f(x) = \frac{1}{2!} - \frac{x^2}{4!} + \frac{x^4}{6!}$, para $|x| < 0.1$ posee la precisión dada arriba. Esto nos brinda una manera mucho más fácil para evaluar $f(x)$.

Teorema 1.1 (Perdida de Precisión) *Si x e y son números representados en la base $\beta > 1$ en su forma normalizada de punto flotante, tales que $x > y$, con*

$$\beta^{-q} \leqslant 1 - \frac{y}{x} \leqslant \beta^{-p}.$$

Entonces en la resta $x - y$ se pierden a lo más q y al menos p dígitos significativos.

Ejemplo 6 Sea $y = x - \text{sen}(x)$. Como $\text{sen}(x) \approx x$ para valores pequeños de x, este cálculo tendrá una perdida de dígitos significativos ¿Cómo puede evitarse?

Una forma alternativa para la función $y = x - \text{sen}(x)$ resulta del desarrollo en serie de Taylor alrededor de $x = 0$. Luego,

$$y = x - \text{sen}(x) = x - \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\right) = \frac{x^3}{3!} - \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots$$

como x es próximo de cero, podemos truncar la serie, obteniendo, por ejemplo

$$y \approx \frac{x^3}{3!} - \frac{x^5}{5!} + \frac{x^7}{7!} - \frac{x^9}{9!}.$$

Consideremos el caso $\text{sen}(x) > 0$, como $x > \text{sen}(x) = y$, utilizando el teorema de Perdida de Precisión, obtenemos que la perdida de dígitos significativos en la resta dada por $y = x - \text{sen}(x)$ puede limitarse a un dígito significativo si restringuimos x de tal modo que $1 - \frac{\text{sen}(x)}{x} \geqslant \frac{1}{10}$, esto es, $\frac{9}{10} \geqslant \frac{\text{sen}(x)}{x}$, de donde tenemos que $\frac{\text{sen}(x)}{x} \leqslant \frac{9}{10}$ si $|x| \geqslant 1.9$. Para $|x| < 1.9$ usamos la representación en serie de Taylor truncada tomando $R_{10}(x) = \frac{x^{10}\text{sen}(x)}{10!}$, luego

$$|R_{10}(x)| = \frac{x^{10}|\text{sen}(x)|}{10!} < \frac{1.9^{10}}{10!} \approx 0.00017 \approx 1.7 \times 10^{-4}.$$

Ejemplo 7 ¿Cuántos dígitos significativos se pierden en la sustracción $1 - \cos(x)$ cuando $x = \frac{1}{4}$?

Consideremos $f(x) = 1 - \cos(x)$, por lo tanto $f\left(\frac{1}{4}\right) = 1 - \cos\left(\frac{1}{4}\right) = 0.0310875$. Como $\cos(x) < 1$, podemos utilizar el teorema de Perdida de Precisión y tenemos que $2^{-q} \leqslant 1 - \cos\left(\frac{1}{4}\right) \leqslant 2^{-p}$, con $q = 6$ y $p = 5$. Por lo tanto se pierden a lo más 6 y a lo menos 5 dígitos significativos.

Ejemplo 8 ¿Cuántos dígitos significativos se pierden en una computadora cuando efectuamos la sustracción $x - \text{sen}(x)$ para $x = \frac{1}{2}$?

Consideremos $f(x) = x - \text{sen}(x)$. Para $x = \frac{1}{2}$ obtenemos $f\left(\frac{1}{2}\right) = \frac{1}{2} - \text{sen}\left(\frac{1}{2}\right) = 0.020574$ como $x > \text{sen}(x)$ para $\text{sen}(x) > 0$, tenemos que $2^{-q} \leqslant 1 - \frac{\text{sen}(0.5)}{0.5} = 0.020574 \leqslant 2^{-p}$, de donde podemos deducir que $q = 5$ y $p = 4$, es decir, se peirden a lo más 5 y al menos 4 dígitos significativos en la resta $x - \text{sen}(x)$ para $x = 1/2$.

1.3.3 Propagación de error

Sea ω una operación aritmética $+$, $-$, $\times$, $/$ y sea $\overline{\omega}$ la correspondiente versión computacional, la cual incluye, usualmente, redondeo o corte. Sean x_A y y_A números usados en los cálculos y suponga que ellos ya tienen errores, siendo sus valores verdaderos $x_T = x_A + \varepsilon$ y $y_T = y_A + \eta$, dicho de otra forma $\varepsilon = E(x_A) = x_T - x_A$ y $\eta = E(y_A) = y_T - y_A$. Entonces $x_A \overline{\omega} y_A$ es el número calculado, para el cual el error es dado por

$$x_T\, \omega\, y_T - x_A\, \overline{\omega}\, y_A = \underbrace{x_T\, \omega\, y_T - x_A\, \omega\, y_A}_{\text{Error Propagado}} + \underbrace{x_A\, \omega\, y_A - x_A\, \overline{\omega}\, y_A}_{\text{Error de redondeo o corte}} \tag{1.15}$$

La cantidad $x_T\, \omega\, y_T - x_A\, \omega\, y_A$ es llamada *error propagado* y la cantidad $x_A\, \omega\, y_A - x_A\, \overline{\omega}\, y_A$ es llamada *error de redondeo* o *corte*.

Usualmente para el error de redondeo o corte tenemos que

$$\boxed{x_A \overline{\omega} y_A = fl\,(x_A \omega y_A)}$$

lo cual significa que $x_A\, \omega\, y_A$ es calculado correctamente y luego redondeado o cortado. Por lo tanto tenemos que

$$|\, x_A \omega y_A - x_A \overline{\omega} y_A| \leqslant \frac{1}{2}\ |x_A \omega y_A|\ \beta^{-k+1}$$

si usamos redondeo, y

$$|\, x_A \omega y_A - x_A \overline{\omega} y_A| \leqslant\ \ |x_A \omega y_A|\ \beta^{-k+1}$$

si usamos corte.

Lo anterior nos da el redondeo o corte verdadero usado.

Para el error propagado examinaremos algunos casos particulares.

Caso a. **Multiplicación**. Para el error en $x_A y_A$, tenemos $x_T = x_A + \varepsilon$ y $y_T = y_A + \eta$, es decir, $\varepsilon = E(x_A) = x_T - x_A$ y $\eta = E(y_A) = y_T - y_A$, por lo tanto

$$\begin{aligned} E(x_A y_A) &= x_T y_T - x_A y_A \\ &= x_T y_T - (x_T - \varepsilon)\,(y_T - \eta) \\ &= x_T \eta + y_T \varepsilon - \varepsilon\eta \end{aligned}$$

de donde,

$$\boxed{E(x_A y_A) = x_T E(y_A) + y_T E(x_A) - E(x_A)E(y_A)} \tag{1.16}$$

En particular, $E(x_A^2) = 2x_T E(x_A) - E(x_A)^2$.

Para el error relativo, tenemos

$$\begin{aligned} E_R(x_A y_A) &= \frac{x_T y_T - x_A y_A}{x_T y_T} \\ &= \frac{\eta}{y_T} + \frac{\varepsilon}{x_T} - \frac{\varepsilon\eta}{x_T y_T}, \end{aligned}$$

es decir,

$$\boxed{E_R(x_A y_A) = E_R(x_A) + E_R(y_A) - E_R(x_A)\, E_R(x_A)} \tag{1.17}$$

Observemos que si $|E_R(y_A)|$, $|E_R(x_A)| << 1$ ($<<$ significa "mucho menor que") entonces $E_R(x_A y_A) \approx E_R(x_A) + E_R(y_A)$.

Caso b. División.

$$\boxed{E_R\left(\frac{x_A}{y_A}\right) = \frac{E_R(x_A) - E_R(y_A)}{1 - E_R(y_A)}} \tag{1.18}$$

Si $|E_R(y_A)| << 1$ se tiene que $E_R\left(\frac{x_A}{y_A}\right) \approx E_R(x_A) - E_R(y_A)$.

Observación. Para las operaciones de multiplicación y división, los errores relativos no se propagan rápidamente.

Caso c. Adición y sustracción.

$$\begin{aligned} E(x_A \pm y_A) &= (x_T \pm y_T) - (x_A \pm y_A) \\ &= (x_T - x_A) \pm (y_T - y_A) = \varepsilon \pm \eta, \end{aligned}$$

por lo tanto

$$\boxed{E(x_A \pm y_A) = E(x_A) \pm E(y_A)} \tag{1.19}$$

Esto parece ser bastante bueno y razonable, pero puede ser engañoso. El error relativo en $x_A \pm y_A$ puede ser bastante pobre cuando es comparado con $E_R(x_A)$ y $E_R(y_A)$.

Ejemplo 9 Consideremos $x_T = \pi$, $x_A = 3.1416$, $y_T = \frac{22}{7}$, e $y_A = 3.1429$. Tenemos

$$E(x_A) = x_T - x_A \approx -7.35 \times 10^{-6}, \quad E_R(x_A) = -2.34 \times 10^{-6}$$

$$E(y_A) = y_T - y_A \approx -4.29 \times 10^{-5}, \quad E_R(y_A) = -1.36 \times 10^{-5}$$

$$E(x_A \pm y_A) = (x_T - y_T) - (x_A - y_A) \approx -0.0012645 - (-0.0013) \approx 3.55 \times 10^{-5}$$

Aunque el error en $x_A - y_A$ es pequeño, el error relativo en $x_A - y_A$ es mucho mayor que en x_A e y_A.

Ejemplo 10 Sea $x_A = 3.0015 = 0.30015 \times 10^1$ correctamente redondeado al número de dígitos señalados. Considere las expresiones matemáticamente equivalentes $u(x) = (3-x)^2$, $v(x) = (3-x)(3-x)$ y $w(x) = (x-6)x + 9$.

Si utilizamos en la evaluación de cada una de las expresiones anteriores un computador con aritmética decimal de redondeo a 4 dígitos ¿qué valores numéricos se obtendrían?

Es fácil realizar los cálculos numéricos con una máquina, por lo tanto se dejan a cargo del lector.

Estudiemos la propagación del error absoluto en cada una de las expresiones correspondientes, determinando cotas para los valores absolutos de cada uno de los errores respectivos.

Sabemos que $-\frac{\beta^{-k+1}}{2} \leqslant \frac{x_T - x_A}{x_T} \leqslant \frac{\beta^{-k+1}}{2}$, de donde $|x_T - x_A| \leqslant \frac{\beta^{-k+1}}{2}|x_T|$. Ahora, si escribimos x_T en su forma normalizada, es decir, $x_T = \sigma\, \overline{x}_T\, 10^e$, donde $0.1 \leqslant \overline{x}_T < 1$, obtenemos que $E(x_A) = |x_T - x_A| \leqslant \frac{10^{-k+1}}{2} |\, \overline{x}_T \,| \times 10^e$, aquí $k = 5$ y $e = 1$, por lo tanto $E(x_A) = |x_T - x_A| \leqslant \frac{10^{-3}}{2}$, de donde $-\frac{10^{-3}}{2} + x_A \leqslant x_T \leqslant \frac{10^{-3}}{2} + x_A$ luego $3.001 \leqslant x_T \leqslant 3.002$, por lo tanto

$$|E(u_A)| = \left|E(3-x_A)^2\right| = |E((3-x_A)\ (3-x_A))| = \left|2(3-x_T)E(3-x_A) - (E(3-x_A))^2\right|$$

$$\leqslant 2|3-x_T|\,|E(3-x_A)| + \left|E(3-x_A)^2\right| \leqslant 2 \times 0.002\,|E(x_A)| + |E(x_A)|^2$$

$$\leqslant 2 \times 0.002 \times \tfrac{10^{-3}}{2} + \left(\tfrac{10^{-3}}{2}\right)^2 = 0.225 \times 10^{-5}.$$

Observación. Desde la definición del error se tiene que $E(x_A + c) = E(x_A)$.

Falta analizar para v_A. Este se deja a cargo del lector.

Tenemos finalmente que

$$|E(w_A)| = |E((x_A - 6)x_A + 9)| = |E((x_A - 6)x_A)|$$

$$= |(x_T - 6)E(x_A) + x_T E(x_A - 6) - E(x_A)E(x_A - 6)|$$

$$\leqslant |x_T - 6|\,|E(x_A)| + |x_T|\,|E(x_A)| + |E(x_A)|^2$$

$$\leqslant 2.999 \times \tfrac{1}{2} \times 10^{-3} + 3.002\tfrac{1}{2} \times 10^{-3} + \left(\tfrac{1}{2} \times 10^{-3}\right)^2 = 0.00300075.$$

1.4 Propagación de error en evaluación de funciones

Sea $f(x)$ una función, la cual suponemos derivable, con derivada continua. Sea x_A un valor aproximado de x_T ¿Cómo aproxima $f(x_A)$ a $f(x_T)$?

Usando el Teorema del Valor Medio, tenemos

$$f(x_T) - f(x_A) = f'(c)\,(x_T - x_A)$$

donde c entre x_A y x_T. Como en general, x_A y x_T son bastante próximos, tenemos que

$$\begin{aligned} f(x_T) - f(x_A) &\approx f'(x_T)\,(x_T - x_A) \\ &\approx f'(x_A)\,(x_T - x_A)\,, \end{aligned}$$

luego,

$$\boxed{E(f(x_A)) \approx f'(x_A)E(x_A)} \tag{1.20}$$

Tenemos además que

$$\boxed{|E(f(x_A))| \leqslant |E(x_A)| \cdot \max\{\,|f'(x)| \;:\; x \text{ entre } x_T \text{ y } x_A\}} \tag{1.21}$$

Para el error relativo tenemos la fórmula

$$\boxed{E_R(f(x_A)) \approx \frac{f'(x_T)}{f(x_T)}\,(x_T - x_A) \;\approx\; \frac{f'(x_T)}{f(x_T)}\,x_T\;E_R(x_A)} \tag{1.22}$$

El número $\kappa(x) = \frac{f'(x)}{f(x)}\,x$ es llamado el *número de condición* de $f(x)$. Si $|\lim_{x\to x_T}\kappa(x)| = +\infty$, decimos que la evaluación de f para $x \approx x_T$ es *numéricamente inestable*. Caso contrario, decimos que la evaluacion de f para $x \approx x_T$ es *estable*.

Notemos que si $\min\{f(x) \;:\; x \text{ entre } x_T \text{ y } x_A\} \neq 0$, entonces

$$|E_R(x_A)| \leqslant \frac{\max\{\,|f'(x)| \;:\; x \text{ entre } x_T \text{ y } x_A\}}{\min\{\,|f(x)| \;:\; x \text{ entre } x_T \text{ y } x_A\}}\,\max\{|x| \;:\; x \text{ entre } x_T \text{ y } x_A\}\,|E_R(x_A)|\,.$$

Como antes podemos escribir,

$$f(x_T) - \bar{f}(x_A) = (f(x_T) - f(x_A)) + (f(x_A) - \bar{f}(x_A)),$$

donde $f(x_T) - f(x_A)$ es el error propagado y $f(x_A) - \bar{f}(x_A)$ es el error de la evaluación de $f(x)$ en el computador.

Ejemplo 11 Consideremos $f(x) = \tan(x)$. Para $x_0 = \pi/2$ tomamos aproximaciones $x_1 = \frac{\pi}{2} - 0.00001$ y $x_2 = \frac{\pi}{2} - 0.000001$. Evaluando, nos queda $y_1 = \tan(x_1) = 100.000$ e $y_2 = \tan(x_2) = 1.000.000$, luego $y_2 - y_1 = 900.000$ mientras que $x_2 - x_1 = 9 \times 10^{-6}$.

Ejemplo 12 Considere la función $f(x) = \frac{\text{sen}\left(\frac{\pi}{3}+x\right)-\text{sen}\left(\frac{\pi}{3}-x\right)}{2x}$.

Veamos si esta función es numéricamente estable la evaluación de $f(x)$ para $x \approx 0$.

Primero un argumento teórico de resta de números parecidos nos muestra que $f(x)$ es numéricamente inestable en su evaluación para $x \approx 0$. Vemos si hay perdida de dígitos significativos evaluando $f(x)$ para $x = 10^{-n}$ con $n = 1, 2, \ldots, 13$. Tenemos la siguiente tabla

x_n	$f(x)$
$x_1 = \frac{1}{10}$	0.017450368
$x_2 = \frac{1}{10^2}$	0.017450377
$x_3 = \frac{1}{10^3}$	0.017450377
$x_4 = \frac{1}{10^4}$	0.017450377
$x_5 = \frac{1}{10^5}$	0.01745038
$x_6 = \frac{1}{10^60}$	0.0174504
$x_7 = \frac{1}{10^7}$	0.0174505
$x_8 = \frac{1}{10^8}$	0.01745
$x_9 = \frac{1}{10^9}$	0.0174

Observe que el error entre las cantidades x_8 y x_9, lo cual denotamos por $E(x_8, x_9)$, es $E(x_8, x_9) = |0.01745 - 0.0174| = 0.00005$, como el error es un valor grande tenemos inestabilidad numérica. Recuerde que en números de partida próximos se debe tener imágenes próximas.

Determinemos ahora un polinomio de grado mayor o igual a uno, que aproxime a $f(x)$ y que sea numéricamente estable para $x \approx 0$. Desarrollando, obtenemos la expresión siguiente para f,

$$f(x) = \frac{\text{sen}(\frac{\pi}{3})\cos(x) + \text{sen}(x)\ \cos(\frac{\pi}{3}) - \text{sen}(\frac{\pi}{3})\cos(x) + \text{sen}(x)\ \cos\left(\frac{\pi}{3}\right)}{2x} = \frac{\sqrt{3}}{2}\frac{\text{sen}(x)}{x}.$$

Ahora, desarrollando $\text{sen}(x)$ en serie de Taylor tenemos que $\text{sen}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$, luego $f(x) = \frac{\sqrt{3}}{2}\frac{\text{sen}(x)}{x} = \frac{\sqrt{3}}{2}\left(1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots\right)$. Consideramos la siguiente aproximación polinomial $p(x) = \frac{\sqrt{3}}{2}\frac{\text{sen}(x)}{x} \approx \frac{\sqrt{3}}{2}\left(1 - \frac{x^2}{6}\right)$ para f, y vemos que en ella no existe resta de números parecidos luego es estable. Por otra parte, si calculamos el número condición obtenemos para nuestra aproximación polinomial, tenemos

$$\kappa(x) = x\frac{p'(x)}{p(x)} = \frac{x\ \left(-\frac{x}{3}\right)}{1 - \frac{x^2}{6}} = -\frac{2x^2}{6 - x^2}$$

y para valores $x \approx 0$ se tiene que $x\frac{p'(x)}{p(x)} \approx 0$, de donde concluimos que nuestra aproximación es numéricamente estable.

1.5 Errores en sumas

Veamos como se propaga el error en sumas repetidas. Sea $S = a_1 + a_2 + \cdots + a_n = \sum_{j=1}^{n} a_j$, donde cada a_j es un número punto flotante, para sumar estos valores en la máquina se necesitan $n-1$ sumas, cada una de las cuales, envuelve errores de redondeo o corte. Mas precisamente, definimos

$$S_2 = fl\,(a_1 + a_2)$$

la versión punto flotante de $a_1 + a_2$. Enseguida, definimos

$$\begin{aligned} S_3 &= fl\,(a_3 + S_2) \\ S_4 &= fl\,(a_4 + S_3) \\ &\vdots \\ S_n &= fl\,(a_n + S_{n-1}) \end{aligned}$$

S_n es la versión punto flotante de S. Tenemos

$$\begin{aligned} S_2 &= (a_1 + a_2)\ \ (1+\varepsilon_2) \\ S_3 &= (a_3 + S_2)\ \ (1+\varepsilon_3) \\ &\vdots \\ S_n &= (a_n + S_{n-1})\ \ (1+\varepsilon_n)\,. \end{aligned}$$

Los término en la expresión anterior pueden ser combinados, manipulados y estimados, obteniendo lo siguiente

$$\begin{aligned} S - S_n \ = \ & -a_1\,(\varepsilon_2 + \cdots + \varepsilon_n) \\ & -a_2\,(\varepsilon_2 + \cdots + \varepsilon_n) \\ & -a_3\,(\varepsilon_3 + \cdots + \varepsilon_n) \\ & \vdots \\ & -a_n\varepsilon_n \end{aligned}$$

observando esta fórmula y tratando de minimizar el error total de $S - S_n$, la siguiente puede ser una estrategia razonable: ordenar los términos $a_1,\ a_2, \ldots,\ a_n$ antes de sumarlos de tal modo que $0 \leqslant |a_1| \leqslant |a_2| \leqslant |a_3| \leqslant \cdots \leqslant |a_n|$. De esa manera los términos del lado derecho de la expresión arriba con el mayor número de ε_j son multiplicados por los menores valores de a_j. Luego esto minimiza $S - S_n$ sin costos adicionales en la mayoría de los casos.

1.6 Estabilidad en métodos numéricos

Muchos problemas matemáticos tienen soluciones que son bastante sensitivas a pequeños errores computacionales, por ejemplo errores de redondeo. Para tratar con este fenómeno, introduciremos los conceptos de estabilidad y número condición. El número condición de un problema está relacionado al máximo de precisión que puede ser obtenido en su solución cuando usamos

números de longitud finita y aritmética de computador. Estos conceptos son entonces extendidos a métodos numéricos usados para calcular la solución. En general, queremos que los métodos numéricos usados no tengan sensibilidad a pequeños errores más allá de los originados en el problema matemático mismo.

Para simplificar, supongamos que nuestro problema viene dado por

$$F(x, y) = 0. \tag{1.23}$$

La variable x es la incógnita buscada y la variable y son los datos de los cuales depende la solución. Por ejemplo F puede ser una función real de dos variables, o x puede ser una variable real e y puede ser un vector, puede ser una ecuación diferencial o integral o funcional, etc.

Diremos que el problema (1.23) es *estable* si la solución x depende continuamente de la variable y, esto significa que si $(y_n)_{n\in\mathbb{N}}$ es una sucesión de valores aproximándose a y, entonces los valores solución asociados $(x_n)_{n\in\mathbb{N}}$ deben aproximarse a x de la misma forma. Equivalentemente si hacemos pequeños cambios en y esos deben corresponder a pequeños cambios en x. El sentido en el cual los cambios son pequeños depende de la norma que se este usando en x e y, respectivamente, existen varias elecciones posibles, variando de problema en problema. Problemas estables son también llamados *well-posed problems*. Si un problema no es estable, este es llamado inestable o *ill-posed problem.*

Ejemplo 13 Considere el sistema de ecuaciones lineales

$$\begin{aligned} x + 2y &= 3 \\ 0.5x + 1.000001y &= 1.5 \end{aligned}$$

Multiplicando la primera ecuación por 0.5 y restando las ecuaciones, obtenemos $y = 0$, de donde $x = 3$.

Consideremos ahora el sistema próximo

$$\begin{aligned} x + 2y &= 3 \\ 0.4999999x + 1.000001y &= 1.5 \end{aligned}$$

Multiplicando la primera ecuación por 0.4999999 y restando las ecuaciones, obtenemos $-1.2 \times 10^{-6}y = -3 \times 10^{-7}$, de donde $y = 0.25$ y $x = 3 - 2y = 2.5$, soluciones que no son próximas, aún cuando los coeficientes de las ecuaciones en ambos problemas son próximos, de hecho, $0.5 - 0.4999999 = 10^{-7}$ y las soluciones varian del orden de 0.25.

Ejemplo 14 Consideremos la solución de

$$ax^2 + bx + c = 0, \quad a \neq 0$$

cualquier solución es un número complejo, los datos de este caso son $y = (a, b, c)$ vector de los coeficientes y la solución esta dado por

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

la cual varía continuamente con $y = (a, b, c)$.

Si el problema es inestable, entonces existen serias dificultades al tratar de resolverlo. Usualmente no es posible resolver tales problemas sin primero tratar de entender mas acerca de las propiedades de la solución, usualmente retornando al contexto en el cual el problema matemático fue formulado. Esto nos lleva a otras áreas de investigación en Matemática Aplicada y Análisis Numérico.

Para propósitos prácticos existen muchos problemas que son estables en el sentido anterior, pero que continúan fastidiosos con respecto de los cálculos numéricos. Para tratar esta dificultad introducimos una medida de la estabilidad, llamada *número condición*.

El número condición trata de medir los posibles malos efectos sobre la solución x del problema cuando la variable y es perturbada en una cantidad pequeña. Sea $y + \delta_y$ una perturbación de y, y sea $x + \delta_x$ la solución de la ecuación perturbada

$$F(x + \delta_x, y + \delta_y) = 0.$$

Definimos

$$\kappa(x) = \sup_{\delta_y} \frac{\frac{\|\delta_x\|}{\|x\|}}{\frac{\|\delta_y\|}{\|y\|}} \tag{1.24}$$

donde $\| \ \|$ es una norma en el espacio adecuado.

Problemas que son inestables nos llevan a $\lim_{\delta x \to 0} \kappa(x) = \infty$. El número $\kappa(x)$ es llamado número de condición del problema $F(x, y) = 0$. Este es una medida de la sensibilidad de la solución a pequeños cambios de la data y.

Si $\kappa(x)$ es bastante grande entonces existen cambios pequeños δ_y relativos a y que llevan a grandes cambios relativos a x. Pero si $\kappa(x)$ es pequeño, digamos $\kappa(x) \leqslant C$, con C una constante "razonable dependiendo del problema", por ejemplo, entonces cambios relativos en y siempre llevan cambios pequeños relativos en x. Como cálculos numéricos casi siempre envuelven una variedad de pequeños errores computacionales, no deseamos problemas con un número condición grande. Tales problemas son llamados *ill-condicioned problem*.

Ejemplo 15 Considere el problema $x - a^y = 0$, donde $a > 0$. Perturbando y por δ_y tenemos $\frac{\delta_x}{x} = \frac{a^{y+\delta_y} - a^y}{a^y} = a^{\delta_y} - 1$. Por lo tanto

$$\kappa(x) = \sup_{\delta_y} \frac{\left|\frac{\delta_x}{x}\right|}{\left|\frac{\delta_y}{y}\right|} = \sup_{\delta_y} \left| y \frac{a^{\delta_y} - 1}{\delta_y} \right|.$$

Restringiendo δ_y a valores pequeño, obtenemos que $\kappa(x) \approx |y \ \log(a)|$.

1.7 Inestabilidad numérica de métodos

Definición 1.5 *Decimos que un proceso numérico es inestable si pequeños errores que se producen en alguna etapa se agrandan en etapas posteriores degradando seriamente la exactitud del cálculo en su conjunto.*

Ejemplo 16 Consideremos la siguiente sucesión

$$\begin{cases} x_0 = 1, \;\; x_1 = \dfrac{1}{3} \\ \\ x_{n+1} = \dfrac{13}{3}x_n - \dfrac{4}{3}x_{n-1}, \;\; n \geqslant 1. \end{cases}$$

Observe que la sucesión es equivalente a la sucesión $x_{n+1} = \left(\frac{1}{3}\right)^{n+1}$ y que $x_n > 0$ para todo $n \in \mathbb{N}$. Es claro de esta última expresión que $\lim_{n\to\infty} x_n = 0$, mientras que en la evaluación numérica de x_n aparecen valores negativos, (que los alumnos hagan la tabla para la primera expresión de x_n, es decir, la fórmula dada inicialmente). La prueba de $x_{n+1} = \left(\frac{1}{3}\right)^{n+1}$, es fácil y se hace por inducción.

1.8 Ejemplos resueltos

Ejemplo 17 Considere la función definida por $f(x) = \frac{\sqrt{1-x^2} - \sqrt[3]{1-x^2}}{x^2}$.

1. ¿Se produce inestabilidad numérica al evaluar $f(x)$ en valores de $x \approx 0$?
2. Determine una expresión algebraica, matemáticamente equivalente a $f(x)$, la cual sea estable numéricamente para valores de $x \approx 0$.
3. Determine una aproximación polinomial de grado mayor o igual a tres para $f(x)$ en el intervalo $|x| < 0.5$, la cual sea numéricamente estable.

Solución. Para ver si existe inestabilidad numérica en la evaluación de $f(x)$ para $x \approx 0$ puede usarse una tabla de valores o argumentar la existencia de resta de números parecidos.

Para la segunda parte, tenemos

$$f(x) = \frac{\sqrt{1-x^2} - \sqrt[3]{1-x^2}}{x^2} = -\frac{1-\sqrt{1-x^2}}{x^2} + \frac{1-\sqrt[3]{1-x^2}}{x^2}$$

$$= -\frac{1-\sqrt{1-x^2}}{x^2}\;\frac{1+\sqrt{1-x^2}}{1+\sqrt{1-x^2}} + \frac{1-\sqrt[3]{1-x^2}}{x^2}\;\frac{1+\sqrt[3]{1-x^2}+\sqrt[3]{(1-x^2)^2}}{1+\sqrt[3]{1-x^2}+\sqrt[3]{(1-x^2)^2}}$$

$$= -\frac{1}{1+\sqrt{1-x^2}} + \frac{1}{1+\sqrt[3]{1-x^2}+\sqrt[3]{(1-x^2)^2}}$$

Finalmente, para encontrar una paroximación que sea numéricamente estable para $f(x)$, sean

$$F(x) = \sqrt{1-x^2}, \quad G(x) = \sqrt[3]{1-x^2}$$

por lo tanto se tiene si desarrollamos vía Taylor que

$$F(x) \approx 1 - \frac{x^2}{2} - \frac{x^4}{8} - \frac{x^6}{16}$$

$$G(x) \approx 1 - \frac{x^2}{3} - \frac{x^4}{9} - \frac{2x^6}{9}$$

por lo tanto

$$f(x) \approx \frac{1 - \frac{x^2}{2} - \frac{x^4}{8} - \frac{x^6}{16} - 1 + \frac{x^2}{3} + \frac{x^4}{9} + \frac{2x^6}{9}}{x^2} = -\frac{1}{6} - \frac{1}{72}x^2 - \frac{49}{1296}x^4$$

Ejemplo 18 Considere $x_A = 1.00011$ redondeado al número de dígitos que se señala. Considere las expresiones matemáticamente equivalentes $u = (x-1)(x-2)$ y $v = x(x-3)+2$.

1. Usando el Teorema del Valor Medio, estudie la propagación del error para u y v.
2. Determine buenas cotas para $E_R(u_A)$ y para $E_R(v_A)$.

Solución. Sea x_T el valor exacto y x_A una aproximación de x_T.

Definamos $f(x) = (x-1)(x-2)$, así $f'(x) = 2x-3$, luego

$|f(x_T) - f(x_A)| = |f'(c)|\ |x_T - x_A|$, con c entre x_T y x_A. Si $x_A \approx x_T$ entonces

$$|f(x_T) - f(x_A)| \approx |f'(x_A)|\,|x_T - x_A| = 0.099978\ \cdot |E(x_A)| \leqslant 4.9989\ \cdot 10^{-5}$$

ya que $x_A = 1.00011$ está redondeado a 6 dígitos, entonces se tiene que $|E(x_A)| \leqslant 0.5\ \cdot 10^{-6+1+1} = 0.5\ \cdot\ 10^{-4}$.

Definamos $g(x) = x(x-3)+2$, de donde se tiene que $g'(x) = 2x-3$. Luego

$$\begin{aligned} |g(x_T) - g(x_A)| &\approx |g'(x_A)|\ |x_T - x_A| \\ &= 0.99978\,|E_A(x_A)| \leqslant 0.99978\ \cdot 0.5\ \cdot 10^{-4} = 4.9989\ \cdot 10^{-5} \end{aligned}$$

Para el error relativo, tenemos

$$E_R(u_A) = \frac{|u_T - u_A|}{|u_T|} \leqslant \frac{4.9989\ \cdot 10^{-5}}{|x_T - 1|\ \ |x_T - 2|}$$

Por otro lado sabemos que $|E(x_A)| \leqslant 0.5\ \cdot\ 10^{-6+1+1} = 0.5\ \cdot\ 10^{-4}$ por lo tanto

$$-0.5 \cdot 10^{-4} + 1.00011 \leqslant x_T \leqslant 0.5 \cdot 10^{-4} + 1.00011 \Rightarrow$$

$$0.00006 \leqslant x_T - 1 \leqslant 0.00016 \quad \text{y} \quad -0.99994 \leqslant x_T - 2 \leqslant -0.99984$$

de donde se deduce que

$$\frac{1}{|x_T - 1|} \leqslant \frac{1}{0.00006} \leqslant 16666.67 \quad y \quad \frac{1}{|x_T - 2|} \leqslant \frac{1}{0.99984} \leqslant 1.0001600256\,.$$

por lo tanto

$$E_R(u_A) = \frac{|u_T - u_A|}{|u_T|} \leqslant 4.9989\ \cdot 10^{-5}\,(16666.67)\,(1.0001600256) \leqslant 0.8323283325329$$

1.9 Ejercicios

Problema 1.1 La ecuación de segundo grado $ax^2 + bx + c = 0$ se resuelve usualmente por las fórmulas

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}\,, \qquad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}\,.$$

Pruebe que una solución alternativa viene dada por

$$x_1 = \frac{2c}{-b - \sqrt{b^2 - 4ac}}\,, \quad x_2 = \frac{2c}{-b + \sqrt{b^2 - 4ac}}\,.$$

Escriba un programa en MatLab que resuelva las ecuaciones de segundo grado de las dos maneras indicadas. Ejecute el programa cuando

(i) $a = 1.0$, $b = -5.0$, $c = 6.0$, (ii) $a = 1.0$, $b = 12345678.03$, $c = 0.92$.

Problema 1.2 Usando MatLab calcule $x+y+z$ de las dos formas matematicamente equivalentes, $x+(y+z)$ y $(x+y)+z$ cuando (i) $x = 1.0$, $y = -5.0$, $z = 6.0$, (ii) $x = 1\times 10^{20}$, $y = -1 \times 10^{20}$, $z = 1.0$. Explique los resultados obtenidos.

Problema 1.3 Con MatLab, usando el formato largo, calcule $h = 1/3333$. Enseguida sume 3333 veces la cnatidad obtenida. También multiplique dicha cantidad por 3333. Explique los resultados obtenidos.

Problema 1.4 Considere la función $h\,(x) = \dfrac{\tan\left(\frac{\pi}{4} + x\right) - \tan\left(\frac{\pi}{4} - x\right)}{2x}$

1. ¿Es numéricamente estable la evaluación de $h(x)$ para $x \approx 0$? Justifique su respuesta.
2. Determine una expresión matemáticamente equivalente a $h(x)$ que sea numéricamente estable en su evaluación para $x \approx 0$.
3. Usando desarrollos de Taylor, determine un polinomio de grado mayor o igual que 4, que aproxime a $h(x)$ y que sea numéricamente estable en su evaluación para $x \approx 0$.

Problema 1.5 Considere las expresiones matemáticamente equivalente $u(x, y) = (x + y)^2 - (x-y)^2$ y $v(x, y) = 4xy$. Sabiendo que $x_A = 1.214301$ e $y_A = 0.24578$ están correctamente redondeados al número de dígitos señalados, estudie la propagación del error absoluto (valor absoluto del error) en la evaluación de u y de v , determinando una buena cota Cuál de ellas tiene menor propagación del error?

Problema 1.6 Considere la ecuación cúbica $x^3 - 31x - 31x - 1 = 0$. Al calcular las raíces de esta ecuación obtenemos $r_T^{(1)} = 15 + \sqrt{224}$, $r_T^{(2)} = 15 - \sqrt{168}$, y $r_T^{(3)} = 1$. Suponga que usa aritmética de redondeo a 5 dígitos y denote por $r_A^{(1)}$, $r_A^{(2)}$, y $r_A^{(3)}$ las soluciones aproximadas obtenidas para la ecuación.

1. Calcule el error absoluto y el error relativo (en valor absoluto) en las soluciones aproximadas.
2. Una de ellas tiene perdida del dígitos significativos. Proponga una expresión algebraica matemáticamente equivalente para evitar la perdida de dígitos significativos en ese caso. Justifique la respuesta calculando el nuevo error relativo (en valor absoluto).

Problema 1.7 Sea $w(x, y) = (x - y)(x^2 + xy + y^2)$, con $x_A = 1.2354$ e $y_A = 3.001$, número redondeado correctamente al número de dígitos señalados, determine una buena cota para $|E(x_A, y_A)|$.

Problema 1.8 Sea $f(x) = \dfrac{\operatorname{sen}(x) - x}{x}$.

1. Analice la estabilidad numérica en la evaluación de $f(x)$ para $x \approx 0$.
2. Encuentre una aproximación polinomial de grado mayor o igual que 3 para $f(x)$ y que sea estable numéricamente para la evaluación cuando $x \approx 0$.

Problema 1.9 Considere la función definida por $f(x) = 100^x$.

1. Estudie la estabilidad numérica en la evaluación de $f(x)$ de acuerdo a los valores de $x \in \mathbb{R}$.
2. Si $x_A = 1000.01$ es un número redondeado correctamente al número de dígitos señalados, determine una buena cota para $|E(f(x_A))|$,.

Problema 1.10 Considere las expresiones matemáticamente equivalentes: $u = (x-1)^3$ y $v = -1 + x(3 + x(-3 + x))$

1. Estudie la propagación del error absoluto en las expresiones u y v.
2. Qué puede decir acerca de la presición que se puede obtener para u y v, para un determinado valor de x?

Problema 1.11 Sea $x_A = 2.0015$, redondeado al número de dígitos señalados. Considere las expresiones matemáticamente equivalentes: $u = (x-4)^2$, $v = (x-4)(x-4)$, $w = (x-8)x+16$.

1. Si utiliza en la evaluación de cada una de las expresiones anteriores un computador con aritmética decimal de redondeo a 4 dígitos, qué valores numéricos se obtendrían en cada una de las expresiones?
2. Estudie la propagación del error (absoluto) en cada una de las expresiones correspondientes, determinando buenas cotas para los valores absolutos de cada uno de los errores respectivos.

Problema 1.12 Analice la estabilidad numérica en la evaluación de $f(x_A)$ para $x_A \approx 0$, donde

$$f(x) = \frac{\mathrm{e}^x - 1}{\operatorname{sen}(x)}.$$

Problema 1.13 Sean $x_A = 1.0005$ e $y_A = 0.999092$, redondeados correctamente al número de dígitos señalados. Considere las expresiones matemáticamente equivalentes: $u = x^3 + y^3$ y $w = (x+y)(x^2 - xy + y^2)$.

1. Si utiliza en la evaluación de cada una de las expresiones anteriores un computador con aritmética decimal de redondeo a 4 dígitos en cada etapa del cálculo, qué valores numéricos se obtendrían en cada una de las expresiones?
2. Estudie la propagación del error (absoluto) en la expresión u, determinando una buena cota para el valor absoluto del error (usando x_A).

Problema 1.14 Considere la función $f(x) = \dfrac{e^{2x}-1}{2xe^x}$.

i) Es numéricamente estable la evaluación de $f(x)$ para $x \approx 0$? Justifique brevemente y alternativamente evaluando $f(x)$, para $x = 10^{-n}$ con $n = 1, \ldots, 13$.

ii) Determine un polinomio de grado mayor o igual que 3, que aproxime a $f(x)$ y que sea estable numéricamente para $x \approx 0$.

Problema 1.15 Sea $x_A = 1.0015$, redondeado correctamente al número de dígitos señalados. Considere las expresiones matemáticamente equivalentes: $u = (x-1)^3$ y $w = ((x-3)x+3)x-1$.

1. Si utiliza en la evaluación de cada una de las expresiones anteriores un computador con aritmética decimal de redondeo a 4 dígitos en cada etapa del cálculo, qué valores numéricos se obtendrían en cada una de las expresiones?

2. Estudie la propagación del error (absoluto) en la expresión u, determinando una buena cota para el valor absoluto del error (usando x_A).

Problema 1.16 Considere la función $f(x) = \dfrac{\cos(\pi/3+x) - \cos(\pi/3-x)}{2x}$.

i) Es numéricamente estable la evaluación de $f(x)$ para $x \approx 0$? Justifique brevemente y alternativamente evaluando $f(x)$, para $x = 10^{-n}$ con $n = 1, \ldots, 13$.

ii) Determine un polinomio de grado 1, que aproxime a $f(x)$ y que sea estable numéricamente para $x \approx 0$.

Problema 1.17 Considere la función $f(x) = \sqrt{x+1/x} - \sqrt{x-1/x}$.

i) Estudie la estabilidad numérica en la evaluación de $f(x)$ para $x \approx 0$. Justifique.

ii) Proponga una expresión algebraica, que sea matemáticamente equivalente a $f(x)$ y que sea numéricamente estable en la evaluación para $x \approx 0$.

iii) Determine un polinomio de grado mínimo mayor que 0, que aproxime a $f(x)$, y que sea estable numéricamente para la evalución en $x \approx 0$. Además, obtenga una buena cota para el error correspondiente con $|x| \leqslant 0.1$ y $x \neq 0$.

Problema 1.18 Considere las expresiones equivalentes, $u_A = (x_A - 2)(x_A^2 - 3x_A + 1)$ y $v_A = ((x_A - 5)x_A + 7)x_A - 2$. Suponga que cada una de las operaciones aritméticas en las expresiones u_A y v_A deben ser redondeadas a 3 dígitos en la mantisa.

(a) Para $x_A = 2.01$, redondeado al número de dígitos señalados, evalué las expresiones equivalentes de acuerdo a lo expresado arriba, y compare las precisiones de ambos resultados.

(b) Estudie la propagación del error absoluto en ambas expresiones, u_A y v_A, acotando los errores correspondientes (en valor absoluto) Qué puede decir acerca de lo concluido en a)?

Capítulo 2

Ecuaciones no Lineales

En este capitulo estudiaremos uno de los problemas básicos de la aproximación numérica: *el problema de la búsqueda de raíces.* Este consiste en obtener una raíz exacta o una buena aproximación a la raíz exacta de una ecuación de la forma $f(x) = 0$, donde f es una función dada. Este es uno de los problemas de aproximación más antiguos, y sin embargo, la investigación correspondiente todavía continua. El problema de encontrar una aproximación a una raíz de una ecuación se remonta por lo menos al año 1700 a.C. Una tabla cuneiforme que pertenece a la Yale Banylonian Collection, y que data de este período, da la aproximación de $\sqrt{2}$, la cual puede calcularse con algunos de los método que veremos más adelante

2.1 Método de bisección

Este método se basa en el Teorema del Valor Intermedio, el cual enunciamos a seguir.

Teorema 2.1 (del valor intermedio) *Sea* $f : [a, b] \longrightarrow \mathbb{R}$ *una función continua. Supongamos que* $f(a)$ *y* $f(b)$ *tienen signos diferentes, entonces existe* $r \in (a, b)$ *tal que* $f(r) = 0$.

Aunque el procedimiento se aplica en el caso en que $f(a)$ y $f(b)$ tengan signos diferentes y exista más de una raíz en el intervalo (a, b), por razones de simplicidad suponemos que la raíz de este intervalo es única. El método requiere dividir varias veces en la mitad los subintervalos de $[a, b]$ y en cada paso localizar aquella mitad que contenga a la raíz r. Para comenzar consideremos $a_0 = a$ y $b_0 = b$, y sea c_0 el punto medio de $[a, b]$, es decir, $c_0 = \frac{a_0+b_0}{2}$. Si $f(c_0) = 0$, entonces $r = c_0$; si no, entonces $f(c_0)$ posee el mismo signo que $f(a_0)$ o que $f(b_0)$. Si $f(c_0)$ y $f(a_0)$ tienen igual signo, entonces $r \in (c_0, b_0)$ y tomamos $a_1 = c_0$ y $b_1 = b_0$. Si $f(c_0)$ y $f(a_0)$ tienen signos opuestos, entonces $r \in (a_0, c_0)$ y tomamos $a_1 = a_0$ y $b_1 = c_0$. Enseguida, volvemos a aplicar el proceso al intervalo $[a_1, b_1]$, y así sucesivamente.

A continuación describiremos algunos procedimientos de parada que pueden aplicarse en algún paso del algoritmo, o a cualquiera de las técnicas iterativas que se estudian en este capitulo. Se elige una tolerancia $\varepsilon > 0$ y generamos una sucesión de puntos $p_1, p_2, \ldots, p_N$, con $p_n \longrightarrow r$ hasta que se cumplan una de las siguientes condiciones

$$|p_N - p_{N-1}| \leqslant \varepsilon \tag{2.1}$$

$$\frac{|p_N - p_{N-1}|}{|p_N|} \leqslant \varepsilon, \quad p_N \neq 0 \tag{2.2}$$

$$|f(p_N)| \leqslant \varepsilon \tag{2.3}$$

Al usar cualquiera de estos criterios de parada pueden surgir problemas. Por ejemplo, existen sucesiones $(p_n)_{n\in\mathbb{N}}$ con la propiedad de que las diferencias $p_n - p_{n-1}$ convergen a cero, mientras que la sucesión diverge, esto se ilustra con la sucesión siguiente, sea $(p_n)_{n\in\mathbb{N}}$ la sucesión dada por $p_n = \sum_{k=1}^{n} \frac{1}{k}$, es conocido que $(p_n)_{n\in\mathbb{N}}$ diverge aún cuando se tiene $\lim_{n\to\infty} (p_n - p_{n-1}) = 0$. También es posible que $f(p_n)$ este cercano a cero, mientras que p_n difiere significativamente de r, como lo ilustra la siguiente sucesión. Sea $f(x) = (x-1)^{10}$, tenemos que $r = 1$, tomando $p_n = 1 + \frac{1}{n}$ es fácil ver que $|f(p_n)| < 10^{-3}$ para todo $n > 1$, mientras que $|r - p_n| < 10^{-3}$ sólo si $n > 1000$. En caso que no se conozca r, el criterio de parada (2.2) es el mejor al cual puede recurrirse, ya que verifica el error relativo.

Observe que para iniciar el algoritmo de bisección, tenemos que encontrar un intervalo $[a, b]$, de modo que $f(a) \cdot f(b) < 0$. En cada paso, la longitud del intervalo que se sabe contiene una raíz de f se reduce en un factor de $\frac{1}{2}$; por lo tanto, conviene escoger un intervalo inicial $[a, b]$ lo mas pequeño posible. Por ejemplo, si $f(x) = x^2 - 1$, entonces $f(0) \cdot f(2) < 0$ y también $f(0.75) \cdot f(1.5) < 0$, de manera que el algoritmo de bisección puede emplearse en ambos intervalos $[0, 2]$ o $[0.75, 1.5]$. Al comenzar el algoritmo de bisección en $[0, 2]$ o con $[0.75, 1.5]$, la cantidad de iteraciones necesarias para alcanzar determinada exactitud varia.

El siguiente ejemplo ilustra el algoritmo de bisección. La iteración se termina cuando el error relativo es menor que 0.0001, es decir, cuando

$$\frac{|c_n - c_{n-1}|}{|c_n|} < 10^{-4}.$$

Ejemplo 19 La ecuación $f(x) = x^3 + 4x^2 - 10$ posee una raíz en $[1, 2]$ ya que $f(1) = -5$ y $f(2) = 14$. El algoritmo de bisección puede ser resumido por la siguiente tabla

n	a_n	b_n	c_n	$f(c_n)$
1	1.0	2.0	1.5	2.375
2	1.0	1.5	1.25	−1.79687
3	1.25	1.5	1.375	0.16211
⋮	⋮	⋮	⋮	⋮
13	1.36499024	1.3671875	1.365112305	−0.00194

Después de 13 iteraciones, $c_{13} = 1.365112305$ aproxima a la raíz r con un error de $|r - c_{13}| < |b_{13} - a_{13}| = 0.000122070$, ya que $|a_{13}| < |r|$, obtenemos

$$\frac{|r - c_{13}|}{|r|} < \frac{|b_{13} - a_{13}|}{|c_{13}|} \leqslant 9 \times 10^{-5},$$

por lo tanto la aproximación será correcta por lo menos en cuatro dígitos significativos. El valor correcto de r con nueve cifras decimales correctas es $r = 1.365230013$. Observe que r_9 está más cerca de r que c_{13}. Pero lamentablemente no podemos verificar esto si no conocemos la respuesta correcta.

El método de bisección, aunque claro desde el punto de vista conceptual, ofrece inconvenientes importantes, como el de converger lentamente, es decir, la cantidad de iteraciones puede ser demasiado grande para poder obtener que c_n este lo próximo a r, además, inadvertidamente podemos desechar una aproximación intermedia. Sin embargo, tiene la importante propiedad de que siempre converge en una solución y por tal razón a menudo sirve para iniciar los métodos más eficientes que explicaremos más adelante.

2.1.1 Análisis del error

Denotemos los intervalos generados por el método de bisección por $[a_0, b_0]$, $[a_1, b_1]$, $[a_2, b_2]$, ..., de donde obtenemos que

$$a_0 \leqslant a_1 \leqslant \cdots \leqslant b_0$$

luego la sucesión $(a_n)_{n\in\mathbb{N}}$ es creciente y acotada superiormente. Tenemos también que

$$b_0 \geqslant b_1 \geqslant \cdots \geqslant a_0$$

luego la sucesión $(b_n)_{n\in\mathbb{N}}$ es decreciente y acotada inferiormente.

Por lo tanto existen los límites $\lim_{n\to\infty} a_n \leqslant b_0$ y $\lim_{n\to\infty} b_n \geqslant a_0$. Además, como

$$b_n - a_n = \frac{1}{2}(b_{n-1} - a_{n-1}) = \frac{1}{4}(b_{n-2} - a_{n-2}) = \cdots = \frac{1}{2^n}(b_0 - a_0),$$

se sigue que $\lim_{n\to\infty}(b_n - a_n) = 0$, luego $\lim_{n\to\infty} b_n = \lim_{n\to\infty} a_n = r$. Ahora como $f(a_n) f(b_n) \leqslant 0$, haciendo $n \to \infty$ se tiene que $(f(r))^2 \leqslant 0$, pues f es continua, de donde $f(r) = 0$.

Si el proceso se detiene en la iteración n, entonces f posee una raíz en el intervalo $[a_n, b_n]$ y

$$|r - a_n| \leqslant 2^{-n}(b_0 - a_0) \quad \text{y} \quad |r - b_n| \leqslant 2^{-n}(b_0 - a_0).$$

Por otra parte, vemos que una mejor aproximación para la raíz r de $f(x) = 0$ es $c_n = \frac{a_n + b_n}{2}$, pues

$$|r - c_n| \leqslant \frac{1}{2}(b_n - a_n) = 2^{-(n+1)}(b_0 - a_0).$$

Resumiendo lo anterior, tenemos el siguiente resultado.

Teorema 2.2 *Sean* $[a_0, b_0]$, $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_n, b_n]$, ... *los intervalos obtenidos en el método de bisección, entonces* $\lim_{n\to\infty} b_n = \lim_{n\to\infty} a_n = r$ *y* r *es una raíz de* $f(x) = 0$. *Además, se tiene que* $|r - a_n| \leqslant 2^{-n}(b_0 - a_0)$ *y* $|r - b_n| \leqslant 2^{-n}(b_0 - a_0)$. *Por otra parte, si* $c_n = \frac{a_n + b_n}{2}$ *entonces* $r = \lim_{n\to\infty} c_n$ *y* $|r - c_n| \leqslant 2^{-(n+1)}(b_0 - a_0)$.

Es importante señalar que el teorema sólo nos proporciona una cota del error de aproximación y que esta puede ser extremadamente conservadora. Por ejemplo, cuando la aplicamos al problema del ejemplo anterior sólo garantiza que $|p - p_9| \leqslant \frac{2-1}{2^9} \approx 2 \times 10^{-3}$ pero el error real es mucho menor $|p - p_9| \approx 4.4 \times 10^{-6}$.

Ejemplo 20 Para determinar el número de iteraciones necesarias para resolver la ecuación $f(x) = 0$ donde $f(x) = x^3 + 4x^2 - 10$ con una exactitud de 10^{-3} en el intervalo $[1, 2]$ basta determinar un entero N tal que

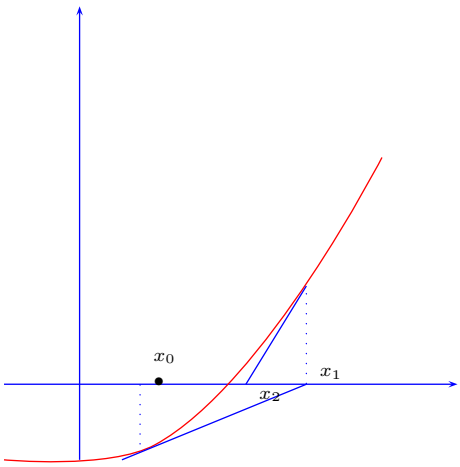
$$|c_N - r| \leqslant 2^{-(N+1)}(b-a) = 2^{-(N+1)} \leqslant 10^{-3}.$$

Aplicando logaritmo a la desigualdad $2^{-(N+1)} \leqslant 10^{-3}$ vemos que el valor del entero positivo N debe ser mayor que 9.96, por lo tanto para obtener una exactitud de 10^{-3} debemos iterar al menos 10 veces.

2.2 Método de Newton

El método Newton es uno de los métodos numéricos más populares para tratar un problema de búsqueda de raíces de una ecuación $f(x) = 0$. Una forma de introducir el método de Newton se basa en los polinomios de Taylor. En esta ocasión introduciremos el método de Newton geométricamente.

Consideremos una función derivable $f : [a, b] \longrightarrow \mathbb{R}$, que tiene un cero en $[a, b]$. Sea $x_0 \in (a, b)$ un punto arbitrario.



Para determinar la intersección x_1 de la recta tangente al gráfico de f en el punto $(x_0, f(x_0))$ con el eje x basta observar que dicha recta tiene por ecuación

$$y - f(x_0) = f'(x_0)(x - x_0)$$

así la intersección con eje x está dada por $-f(x_0) = f'(x_0)(x_1 - x_0)$, despejando x_1 obtenemos

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

si $f'(x_0) \neq 0$. Aplicamos ahora este procedimiento comenzando con x_1, para determinar la intersección x_2 de la recta tangente al gráfico de f en el punto $(x_1, f(x_1))$ con el eje x basta observar que dicha recta tiene por ecuación

$$y - f(x_1) = f'(x_1)(x - x_1)$$

así la intersección con eje x esta dada por $-f(x_1) = f'(x_1)(x_2 - x_1)$, despejando x_2 obtenemos

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \quad ,$$

si $f'(x_1) \neq 0$. De modo análogo, si comenzamos con el punto x_2, obtenemos un punto x_3 dado por

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}$$

si $f'(x_2) \neq 0$, y así sucesivamente. Este procedimiento da lugar a un proceso iterativo llamado *método de Newton*, dado por

$$\boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}} \tag{2.4}$$

si $f'(x_n) \neq 0$.

Ejemplo 21 Sea $f(x) = \frac{1}{x} - 1$. La ecuación $f(x) = 0$ tiene una raíz $r = 1$ en el intervalo $[0.5, 1]$ Qué ocurre al aplicar el método de Newton con $x_0 = 0.5$? Realizando los cálculos numericamente, tenemos

k	0	1	2	3	4
x_k	0.5	0.75	0.9374999998	0.9960937501	0.9999847417

Un buen ejercicio para el lector es esbozar una explicación para lo que está ocurriendo en este caso.

2.2.1 Análisis del Error

El error en el paso n es definido como

$$\boxed{e_n = x_n - r} \tag{2.5}$$

Supongamos ahora que f'' es continua y que r es un cero simple de f, es decir, $f(r) = 0$ y $f'(r) \neq 0$. Entonces tenemos que

$$e_{n+1} = x_{n+1} - r = x_n - \frac{f(x_n)}{f'(x_n)} - r = (x_n - r) - \frac{f(x_n)}{f'(x_n)} = e_n - \frac{f(x_n)}{f'(x_n)} = \frac{e_n f'(x_n) - f(x_n)}{f'(x_n)}$$

por otro lado se tiene que

$$0 = f(r) = f(x_n - e_n) \overset{\text{Taylor}}{=} f(x_n) - e_n f'(x_n) + \frac{1}{2} e_n^2 f''(\xi_n)$$

donde ξ_n está entre x_n y r. De esta última ecuación obtenemos

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

Luego

$$\boxed{e_{n+1} = \frac{1}{2} \frac{f''(\xi_n)}{f'(x_n)} e_n^2 \approx \frac{1}{2} \frac{f''(r)}{f'(r)} e_n^2 = C e_n^2}$$

Podemos formalizar lo anterior como sigue, si e_n es pequeño y $\frac{f''(\xi_n)}{f'(x_n)}$ no es muy grande, entonces e_{n+1} será más pequeño que e_n.

Definamos

$$C(\delta) = \frac{1}{2} \frac{\max\limits_{|x-r|<\delta} |f''(x)|}{\min\limits_{|x-r|<\delta} |f'(x)|} \tag{2.6}$$

donde $\delta > 0$ es tal que $|f'(x)| > 0$ para $|x - r| < \delta$. Eligiendo $\delta > 0$ más pequeño, si es necesario, podemos suponer que $\delta C(\delta) < 1$, podemos hacer esto pues cuando $\delta \longrightarrow 0$ se tiene que $C(\delta) \longrightarrow \left| \frac{f''(r)}{2f'(r)} \right|$, luego $\delta C(\delta) \longrightarrow 0$ si $\delta \longrightarrow 0$. Denotemos $\rho = \delta C(\delta)$. Comenzando las iteraciones del método de Newton con x_0 tal que $|x_0 - r| < \delta$ obtenemos que $|e_0| \leqslant \delta$ y $|\xi_0 - r| < \delta$ luego por la definición de $C(\delta)$, se tiene que $\left| \frac{1}{2} \frac{f''(\xi_0)}{f'(x_0)} \right| \leqslant C(\delta)$ de donde

$$|x_1 - r| = |e_1| \leqslant e_0^2 \, C(\delta) = |e_0| \, |e_0| \, C(\delta) = |e_0| \, \rho \leqslant |e_0| \leqslant \delta,$$

repetimos el argumento con x_1, $x_1, \ldots, x_n, \ldots$ obteniendo

$$\begin{aligned} &|e_1| \leqslant \rho \, |e_0| \\ &|e_2| \leqslant |e_1| \, \rho \leqslant \rho^2 \, |e_0| \\ &\vdots \\ &|e_n| \leqslant \rho^n \, |e_0| \longrightarrow 0, \quad \text{cuando } n \longrightarrow \infty \end{aligned}$$

así hemos obtenido el siguiente resultado.

Teorema 2.3 *Si f'' es continua y r es un cero simple de f, es decir, $f(r) = 0$ y $f'(r) \neq 0$, entonces existe un intervalo abierto J conteniendo a r y una constante $C > 0$ tal que si el método de Newton se inicia con un punto en J, se tiene*

$$|x_{n+1} - r| \leqslant C (x_n - r)^2 .$$

Teorema 2.4 *Si f'' es continua, y f es creciente, convexa y tiene un cero, entonces este cero es único y la iteración del método de Newton converge a él a partir de cualquier punto inicial.*

El siguiente teorema sobre la conducta local del método de Newton muestra que ella es muy buena, pero como veremos despues, desde el punto de vista global no lo es tanto.

Teorema 2.5 *Sea* $f : \mathbb{R} \longrightarrow \mathbb{R}$ *derivable. Si* $f'(x_0) \neq 0$ *, definimos* $h_0 = -\frac{f(x_0)}{f'(x_0)}$ *,* $x_1 = x_0 + h_0$ *,* $J_0 = [x_1 - |h_0|, x_1 + |h_0|]$ *y* $M = \sup_{x \in J_0} |f''(x)|$*. Si* $2\left|\frac{f(x_0)M}{(f'(x_0))^2}\right| < 1$ *entonces la ecuación* $f(x) = 0$ *tiene una única solución en* J_0 *y el método de Newton con condición inicial* x_0 *converge a dicha solución.*

Demostración. Sea $h_1 = -\frac{f(x_1)}{f'(x_1)}$. Necesitamos estimar h_1 , $f(x_1)$ y $f'(x_1)$. Tenemos $|f'(x_1)| \geqslant |f'(x_0)| - |f'(x_0) - f'(x_1)| \geqslant |f'(x_0)| - M|x_1 - x_0| \geqslant |f'(x_0)| - \frac{1}{2}|f'(x_0)| \geqslant \frac{|f'(x_0)|}{2}$, es decir,

$$|f'(x_1)| \geqslant \frac{|f'(x_0)|}{2} \qquad \text{(desigualdad 1)}$$

Ahora estimaremos $f(x_1)$.

$$\begin{aligned} f(x_1) &= f(x_0) + \int_{x_0}^{x_1} f'(x)dx \\ &= f(x_0) - [(x_1 - x)f'(x)]\,\big|_{x_0}^{x_1} + \int_{x_0}^{x_1} (x_1 - x)f''(x)dx \\ &= f(x_0) + (x_1 - x_0)f'(x_0) + \int_{x_0}^{x_1} (x_1 - x)f''(x)dx \end{aligned}$$

como $h_0 = x_1 - x_0$ y $f'(x_0)h_0 = -f(x_0)$, se tiene que

$$f(x_1) = \int_{x_0}^{x_1} (x_1 - x)f''(x)dx\,.$$

Haciendo el cambio de variables $th_0 = x_1 - x$. Tenemos $dx = -h_0 dt$, y

$$|f(x_1)| \leqslant M|h_0|^2 \int_0^1 tdt \;=\; \frac{M|h_0|^2}{2}\,.$$

es decir,

$$|f(x_1)| \leqslant \frac{M|h_0|^2}{2} \qquad \text{(desigualdad 2)}$$

De las desigualdades (1) y (2), obtenemos

$$|h_1| = \left|\frac{f(x_1)}{f'(x_1)}\right| \leqslant \frac{M|h_0|^2}{2}\left|\frac{2}{f'(x_0)}\right| = |h_0|^2\left|\frac{M}{f'(x_0)}\right|$$

y de $2\left|\frac{Mf(x_0)}{(f'(x_0))^2}\right| < 1$, tenemos

$$\left|\frac{M}{f'(x_0)}\right| < \frac{1}{2}\left|\frac{f'(x_0)}{f(x_0)}\right| = \frac{1}{2}\,\frac{1}{\left|\frac{f(x_0)}{f'(x_0)}\right|} = \frac{1}{2|h_0|}\,,$$

luego $|h_1| < \frac{|h_0|}{2}$, y

$$\begin{aligned}\left|\frac{Mf(x_1)}{(f'(x_1))^2}\right| &= \frac{M}{|f(x_1)|}\left|\frac{f(x_1)}{f'(x_1)}\right| = \frac{M}{|f'(x_1)|}|h_1| \\ &\leqslant \frac{M}{|f'(x_1)|}\frac{|h_0|}{2} \\ &\leqslant M\frac{|h_0|}{2}\frac{2}{|f'(x_0)|} = M\left|\frac{f(x_0)}{(f'(x_0))^2}\right|.\end{aligned}$$

Definamos $x_2 = x_1 + h_1$ y $J_1 = [x_2 - |h_1|, x_2 + |h_1|]$. Tenemos que $J_1 \subset J_0$, y las condiciones del teorema se verifican para x_1, h_1, y J_1. Definimos, inductivamente $h_k = -\frac{f(x_k)}{f'(x_k)}$, si $f'(x_k) \neq 0$ y $x_{k+1} = x_k + h_k$. Tenemos así que $x_k \in J_0$, y $(x_k)_{k\in\mathbb{N}}$ es una sucesión convergente, pues la serie $h_0 + h_1 + \cdots$ converge. Además, el límite $x_f = \lim_{k\to\infty} x_k$ es una raíz de la ecuación $f(x) = 0$, pues $\lim_{k\to\infty}(f(x_k) + f'(x_k)(x_{k+1} - x_k)) = 0$.

Observación. La desigualdad $h_{k+1} \leqslant \frac{h_k}{2}$, nos da la convergencia geométrica del método de Newton.

Teorema 2.6 *En las condiciones del teorema anterior, el método de Newton tiene convergencia cuadrática, esto es,* $|h_{k+1}| \leqslant \frac{M}{|f'(x_k)|}|h_k|^2$.

Demostración. Dividiendo la desigualdad $|f(x_{k+1})| \leqslant \frac{M|h_k|^2}{2}$ por $|f'(x_{k+1})|$, obtenemos

$$|k_{k+1}| = \frac{|f(x_{k+1}|}{|f'(x_{k+1})|} \leqslant \frac{M|h_k|^2}{2}\frac{2}{|f'(x_k)|} = \frac{M|h_k|^2}{|f'(x_k)|}.$$

Ejemplo 22 Sea $f(x) = x^3 - 2x - 5$. Tomando $x_0 = 2$, tenemos $f(x_0) = -1$, $f'(x_0) = 10$, $h_0 = 0.1$ y $J_0 = [2, 2.2]$, puesto que $f''(x) = 6x$ sobre J_0 el supremo M es 13.2. Como $\left|\frac{Mf(x_0)}{(f'(x_0))^2}\right| = 0.132 < 0.5 < 1$, el teorema garantiza que existe una raíz de la ecuación $f(x) = 0$ en el intervalo $[2, 2.2]$, y en consecuencia el método de Newton con condición inicial $x_0 = 2$ converge a dicha raíz.

Observación. Podemos usar como criterios de paradas unos de los siguientes. Seleccione una tolerancia $\varepsilon > 0$ y construya $p_1, p_2, \ldots, p_N$ hasta que se cumpla una de las siguientes desigualdades

$$|p_N - p_{N-1}| \leqslant \varepsilon \tag{2.7}$$

$$\left|\frac{p_N - p_{N-1}}{p_N}\right| \leqslant \varepsilon, \quad p_N \neq 0 \tag{2.8}$$

o bien

$$|f(x_N)| \leqslant \varepsilon. \tag{2.9}$$

El método de Newton es una técnica de iteración funcional de la forma $x_{n+1} = g(x_n)$, donde

$$x_{n+1} = g\left(x_n\right) = N_f(x_n) = x_n - \frac{f\left(x_n\right)}{f'\left(x_n\right)}, \quad n \geqslant 0.$$

En esta ecuación se observa claramente que no podemos continuar aplicando el método de Newton si $f'\left(x_n\right) = 0$ para algún n. Veremos que este método es más eficaz cuando $f'\left(r\right) > 0$.

Verificación de condiciones de convergencia. Cuando queremos aplicar en la práctica las condiciones de convergencia del método de Newton, tenemos el problema que no conocemos exactamente la raíz r de la ecuación $f(x) = 0$. Recordemos que las condiciones de convergencia del método de Newton en un intervalo abierto $J \ni r$ son: $f(r) = 0$; $f'(r) \neq 0$ y f'' continua. Para verificar que estas condiciones se cumplen cuando tomamos aproximaciones x_n para la raíz r, y procedemos como sigue. Verificamos

1. f'' es continua (esta condición depende sólo de la función f y no de las aproximaciones a la raíz).
2. Tomanos las aproximaciones x_n dadas dadas por el método de Newton, y aplicando un criterio de parada, obtenemos una aproximación x_N.
3. Para x_N verificamos que $f(x_N) \approx 0$ y $f'(x_N) \neq 0$.
4. Usando la continuidad de f y de sus derivadas, concluimos que existe un intervalo J que contiene a r tal que si $x_0 \in J$, entonces la sucesión de aproximaciones dada por el método de Newton convergen a la raíz r buscada.

Ejemplo 23 Para aproximar la solución de la ecuación $x = \cos(x)$, consideremos $f\left(x\right) = \cos(x) - x$. Tenemos $f\left(\frac{\pi}{2}\right) = -\frac{\pi}{2} < 0 < 1 = f\left(0\right)$, luego y por el Teorema del Valor Intermedio, existe un cero de f en el intervalo $\left[0, \frac{\pi}{2}\right]$. Aplicando el método de Newton obtenemos

$$x_{n+1} = x_n - \frac{\cos\left(x_n\right) - x_n}{-\operatorname{sen}\left(x_n\right) - 1}, \quad n \geqslant 0.$$

En algunos problemas es suficiente escoger x_0 arbitrariamente, mientras que en otros es importante elegir una buena aproximación inicial. En el problema en cuestión, basta analizar las gráficas de $h\left(x\right) = x$ y $k\left(x\right) = \cos\left(x\right)$ por lo que es suficiente considerar $x_0 = \frac{\pi}{4}$, y así obtenemos una excelente aproximación con sólo tres pasos, como muestra la siguiente tabla.

n	x_n
0	0.7853981635
1	0.7395361337
2	0.7390851781
3	0.7390851332
4	0.7390851332

Para asegurarnos que tenemos convergencia, verifiquemos esas condiciones. Tenemos que $f\left(x\right) = \cos(x) - x$, luego $f''\left(x\right) = -\cos(x)$, la cual es continua, ahora evaluando $f'\left(x\right)$ en el punto $x = 0.7390851332$ (en radianes) que es una aproximación a la raíz verdadera, se tiene el valor no cero siguiente, $f'\left(0.7390851332\right) = -\operatorname{sen}(0.7390851332) - 1 = -1.012899111\ldots \neq 0$.

Ejemplo 24 Para obtener una solución de la ecuación de $x^3 + 4x^2 - 10 = 0$ en el intervalo $[1, 2]$ mediante el método de Newton, generamos la sucesión $(x_n)_{n\in\mathbb{N}}$ dada por

$$x_{n+1} = x_n - \frac{x_n^3 + 4x_n^2 - 10}{3x_n^2 + 8x_{n-1}}, \quad n \geqslant 0$$

Tomando $x_0 = 1.5$ como condición inicial se obtiene el resultado $x_3 = 1.36523001$ este es correcto en ocho cifras decimales.

Para asegurarnos que tenemos convergencia, verifiquemos esas condiciones. Tenemos que $f(x) = x^3 + 4x^2 - 10$, luego $f''(x) = 6x + 8$, la cual es continua. Ahora evaluando $f'(x)$ en el punto $x_3 = 1.36523001$, que es una aproximación a la raíz verdadera, se tiene el valor no cero siguiente, $f'(1.36523001) = 3(1.36523001)^2 + 8 \cdot 1.36523001 = 16.1339902\ldots \neq 0$.

Ejemplo 25 Si queremos resolver $x^3 - x - \frac{\sqrt{2}}{2} = 0$ utilizando el método de Newton y comenzamos las iteraciones con $x_0 = 0.001$ obtenemos una sucesión que oscila entre valores cercanos a 0 y a $\frac{\sqrt{2}}{2}$, y de hecho, si comenzamos con la condición inicial $x_0 = 0$ obtenemos el ciclo periódico $\left\{0, \frac{\sqrt{2}}{2}\right\}$, es decir, $N_f(0) = \frac{\sqrt{2}}{2}$ y $N_f\left(\frac{\sqrt{2}}{2}\right) = 0$.

Observación. La derivación del método de Newton por medio de las series de Taylor, subraya la importancia de una aproximación inicial exacta. La suposición fundamental es que el término que contiene $(x - \overline{x})^2$ es, en comparación, tan pequeño que podemos suprimirlo. Esto evidentemente sería falso a menos que $\overline{x}$ sea una buena aproximación a la raíz r. En particular, si x_0 no es lo suficiente próximo a la raíz real, el método de Newton quizá no sea convergente a la raíz. Pero no siempre es así. Por ejemplo considere la ecuación $x^2 + 1 = 0$ que no tiene soluciones reales, y cuando le aplicamos el método de Newton obtenemos, para cualesquiera que sea la condición inicial una sucesión que no converge a ningún valor determinado, como puede comprobarse en la práctica sin mayores esfuerzos, realizando las iteraciones correspondientes.

2.3 Método de Newton multivariable

Consideremos el sistema de ecuaciones

$$\begin{cases} f_1(x_1, x_2) = 0 \\ f_2(x_1, x_2) = 0. \end{cases} \tag{2.10}$$

Si (x_1, x_2) es una solución aproximada y $(x_1 + h_1, x_2 + h_2)$ es la solución exacta, aplicando Taylor obtenemos

$$\begin{cases} 0 = f_1(x_1 + h_1, x_2 + h_2) \approx f_1(x_1, x_2) + h_1\dfrac{\partial f_1}{\partial x_1} + h_2\dfrac{\partial f_1}{\partial x_2} \\ \\ 0 = f_2(x_1 + h_1, x_2 + h_2) \approx f_2(x_1, x_2) + h_1\dfrac{\partial f_2}{\partial x_1} + h_2\dfrac{\partial f_2}{\partial x_2}. \end{cases}$$

Denotemos el jacobiano de $F(x_1, x_2) = (f_1(x_1, x_2), f_2(x_1, x_2))$ por $J(f_1, f_2)$, es decir,

$$J(f_1, f_2) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} \\ \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} \end{pmatrix}, \tag{2.11}$$

obtenemos

$$J(f_1, f_2)(x_1, x_2) \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = - \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix} \tag{2.12}$$

de donde, si $\det(J(f_1, f_2)(x_1, x_2) \neq 0$, se tiene que

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = -J^{-1}(f_1, f_2)(x_1, x_2) \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix}$$

con lo cual podemos definir el siguiente proceso iterativo

$$\begin{pmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \end{pmatrix} = \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \tag{2.13}$$

donde $\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$ es la colución de la ecuaciónn lineal (2.12), esto puede ser escrito en forma más explícita como

$$\boxed{\begin{pmatrix} x_1^{(k+1)} \\ \\ x_2^{(k+1)} \end{pmatrix} = \begin{pmatrix} x_1^{(k)} \\ \\ x_2^{(k)} \end{pmatrix} - \begin{pmatrix} \frac{\partial f_1}{\partial x_1}\left(x_1^{(k)}, x_2^{(k)}\right) & \frac{\partial f_1}{\partial x_2}\left(x_1^{(k)}, x_2^{(k)}\right) \\ \\ \frac{\partial f_2}{\partial x_1}\left(x_1^{(k)}, x_2^{(k)}\right) & \frac{\partial f_2}{\partial x_2}\left(x_1^{(k)}, x_2^{(k)}\right) \end{pmatrix}^{-1} \begin{pmatrix} f_1\left(x_1^{(k)}, x_2^{(k)}\right) \\ \\ f_2\left(x_1^{(k)}, x_2^{(k)}\right) \end{pmatrix}} \tag{2.14}$$

Observe que esta formulación es, sin embargo, poco útil para realizar los cálculos, ya que es poco práctico calcular la inversa de la matriz jacobiana. Sin embargo para sistemas de 2×2 es fácil obtener. Realizando las operaciones tenemos que

$$\begin{aligned} x_1^{(n+1)} &= x_1^{(n)} - \frac{f_1\left(x_1^{(n)}, x_2^{(n)}\right) \dfrac{\partial f_2}{\partial x_2}\left(x_1^{(n)}, x_2^{(n)}\right) - f_2\left(x_1^{(n)}, x_2^{(n)}\right) \dfrac{\partial f_1}{\partial x_2}\left(x_1^{(n)}, x_2^{(n)}\right)}{\det(J(f_1, f_2)(x_1^{(n)}, x_2^{(n)}))} \\ \\ x_2^{(n+1)} &= x_2^{(n)} - \frac{f_2\left(x_1^{(n)}, x_2^{(n)}\right) \dfrac{\partial f_1}{\partial x_1}\left(x_1^{(n)}, x_2^{(n)}\right) - f_1\left(x_1^{(n)}, x_2^{(n)}\right) \dfrac{\partial f_2}{\partial x_1}\left(x_1^{(n)}, x_2^{(n)}\right)}{\det(J(f_1, f_2)(x_1^{(n)}, x_2^{(n)}))} \end{aligned} \tag{2.15}$$

donde

$$\det J(f_1, f_2)\left(x_1^{(n)}, x_2^{(n)}\right) = \frac{\partial f_1}{\partial x_1}\left(x_1^{(n)}, x_2^{(n)}\right) . \frac{\partial f_2}{\partial x_2}\left(x_1^{(n)}, x_2^{(n)}\right) - \frac{\partial f_2}{\partial x_1}\left(x_1^{(n)}, x_2^{(n)}\right) . \frac{\partial f_1}{\partial x_2}\left(x_1^{(n)}, x_2^{(n)}\right) .$$

Es claro que podemos generalizar el método de Newton a más de dos variables, y de deja a cargo del lector deducir las fórmulas de iteración para un sistema de 3 ecuaciones en tres variables $(x,\ y,\ z)$.

Condiciones de convergencia del método de Newton multivariable. Supongamos que $f_1(r_1, r_2) = f_2(r_1, r_2) = 0$, $\det(J(f_1, f_2)(r_1, r_2)) \neq 0$ y la derivadas parciales segundas $\dfrac{\partial^2 f_1}{\partial x_1^2}$, $\dfrac{\partial^2 f_1}{\partial x_1 \partial x_2}$, $\dfrac{\partial^2 f_1}{\partial x_2^2}$, $\dfrac{\partial^2 f_2}{\partial x_1^2}$, $\dfrac{\partial^2 f_2}{\partial x_1 \partial x_2}$, $\dfrac{\partial^2 f_2}{\partial x_2^2}$ son continuas. Entonces existe un conjunto abierto $U \ni (r_1, r_2)$, tal que si $(x_1^0, x_2^0) \in U$ entonces el método de Newton comenzando con ese punto converge a (r_1, r_2).

Observación. En la práctica, comenzamos con $(x_1^{(0)}, x_2^{(0)})$ adecuados y generamos la sucesión dada por el método de Newton hasta satisfacer algún criterio de parada, obteniendo un punto $(x_1^{(N)}, x_2^{(N)})$. Verificamos entonces las condiciones para ese punto, es decir, $f_i(x_1^{(N)}, x_2^{(N)}) \approx 0$, $i = 1, 2$, $\det J(f_1, f_2)(x_1^{(N)}, x_2^{(N)}) \neq 0$. La condición de continuidad de las segundas derivadas parciales de f_1 y f_2 no depende de la sucesión de puntos generada. Si las condiciones son satisfechas para $(x_1^{(N)}, x_2^{(N)})$, entonces existe un conjunto abierto $U \subset \mathbb{R}^2$, con $U \ni (r_1, r_2)$, tal que si $(x_1^0, x_2^0) \in U$, entonces la sucesión generada por el método de Newton converge a (r_1, r_2).

Observación. En general, especialmente desde el punto de vista computacional, usar las ecuaciones (2.12) y (2.13) como una descompsición del método de Newton, en vez de las ecuaciones dadas en (2.14) o en su forma explícita (2.15). Más genral si deseamaos resolver el sistema de ecuaciones no lienales

$$\begin{cases} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ & \vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0 \end{cases} \tag{2.16}$$

usamos el siguiente algoritmo:

Dado $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$.

Primero: resolvemos el sistema de ecuaciones lineales

$$J_k(f_1, f_2, \ldots, f_n) \begin{pmatrix} h_1^{(k)} \\ h_2^{(k)} \\ \vdots \\ h_n^{(k)} \end{pmatrix} = - \begin{pmatrix} f_1(x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}) \\ f_2(x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}) \\ \vdots \\ f_n(x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}) \end{pmatrix} \tag{2.17}$$

para $H^{(k)} = (h_1^{(k)}, h_2^{(k)}, \ldots, h_n^{(k)})^T$, donde $J_k(f_1, f_2, \ldots, f_n)$ denota el jacobiano de $F = (f_1, f_2, \ldots, f_n)$ evaluado en el punto $(x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$.

Segundo. Realizamos la iteración

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + H^{(k)} \tag{2.18}$$

donde $\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$, y así sucesivamente hasta satisfacer alguna condición de parada.

Ejemplo 26 Consideremos el siguiente sistema

$$\begin{cases} f(x,y) = x^2y - xy^2 - 0.8 = 0 \\ \\ g(x,y) = \dfrac{x^2}{y^2} - 1 - \dfrac{x^2}{2} = 0 \end{cases}$$

y encontremos sus raíces.

Aplicando el método de Newton multivariable, primero calculamos las derivadas parciales de cada función obteniendo

$$\frac{\partial f}{\partial x} = 2xy - y^2 \qquad \frac{\partial f}{\partial y} = x^2 - 2xy$$

$$\frac{\partial g}{\partial x} = \frac{2x}{y^2} - x \qquad \frac{\partial g}{\partial y} = -\frac{2x^2}{y^3}$$

aplicando la fórmula anterior nos queda

$$x_{n+1} = \frac{-3x_n^2y_n + 2(x_ny_n)^2 - 0.5(x_ny_n)^3 - x_n^2y_n^4 - 1.6x_n - x_ny_n^3 + 2y_n^2}{x_ny_n(-6x_n + 6y_n + x_ny_n^2 - 2y_n^3)}$$

$$y_{n+1} = \frac{-6x_n^3y_n + 5(x_ny_n)^2 + (x_ny_n)^3 - 1.5x_n^2y_n^4 + 2x_ny_n^3 - y_n^4 - 1.6x_n + 8x_ny_n^2}{x_n^2(-6x_n + 6y_n + x_ny_n^2 - 2y_n^3)}$$

La siguiente tabla muestra los resultados obtenidos con el método de Newton multivariable

Iter.	x	y	$f(x,y)$	$g(x,y)$	$\lvert J\rvert$
0	1	1	-0.8	-0.5	-1
1	2.1	1.3	1.384	-0.59553	-14.7304
2	1.68036	1.11139	0.26257	-0.12583	-9.33548
3	1.55237	1.04843	0.02019	-0.01257	-7.94105
4	1.54040	1.04178	0.00016	-0.00011	-7.82963
5	1.54030	1.04173	1.06×10^{-8}	-6.9×10^{-9}	-7.82874
6	1.54030	1.04173	2.22×10^{-16}	6.66×10^{-16}	-7.82874

Se puede notar la rapidez a la cual converge el método, ya que en solamente 6 iteraciones se tiene un valor $(x_6, y_6) = (1.54030, 1.04173)$ para el cual $f(1.54030, 1.04173) \approx 0$ y $g(1.54030, 1.04173) \approx 0$, por otra parte, es claro que las segundas derivadas parciales de f y de g son continuas, pues ambas funciones son polinomios en las variables (x, y). Tenemos también que

$$\frac{\partial f}{\partial x}(x_6,y_6) = 2x_6y_6 - y_6^2 = 2.22631238 \qquad \frac{\partial f}{\partial y}(x_6,y_6) = x_6^2 - 2x_6y_6 = -0.85378829$$

$$\frac{\partial g}{\partial x}(x_6,y_6) = \frac{2x_6}{y_6^2} - x_6 = 1.1401168538 \qquad \frac{\partial g}{\partial y}(x_6,y_6) = -\frac{2x_6^2}{y_6^3} = -4.130734823$$

luego

$$\det J(f,g)(x_6,y_6) = \det \begin{pmatrix} 2.22631238 & -0.85378829 \\ 1.1401168538 & -4.130734823 \end{pmatrix} = -4.486785689\,,$$

por lo tanto el método iterativo de Newton es convergente en una vecindad de la solución exacta (x_T, y_T) del sistema.

2.4 Método de la secante

El método de Newton es una técnica poderosa, pero presenta un problema, a saber, la necesidad de conocer el valor de la derivada de f en cada paso de la aproximación. Con frecuencia es más difícil calcular $f'(x)$ y se requieren más operaciones aritméticas para calcularla que para $f(x)$. Si queremos evitar el problema de evaluar la derivada en el método de Newton, derivamos una pequeña variación. Por definición

$$f'(x_n) = \lim_{x \to x_n} \frac{f(x) - f(x_n)}{x - x_n}.$$

Haciendo $x = x_{n-1}$, tenemos

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

Al aplicar esta aproximación para $f'(x_n)$ en la formula de Newton, se obtiene el siguiente método iterativo,

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

el cual depende siempre de los valores de dos iteraciones anteriores, y como valores iniciales se tienen x_0 y x_1, los cuales deben ser elegidos con cierto criterio para tener la convergencia del método. Este método iterativo en llamado *método de la secante*.

Comenzando con las dos aproximaciones iniciales x_0 y x_1, la aproximación x_2 es la intersección del eje x y la recta que une los puntos $(x_1, f(x_1))$ y $(x_0, f(x_0))$. La aproximación x_3 es la intersección con el eje x y la recta que une los puntos $(x_1, f(x_1))$ y $(x_2, f(x_2))$, y así sucesivamente.

Las condiciones para garantizar la convergencia del método de la secante en un intervalo abierto J que contiene a la raíz r de la ecuación $f(x) = 0$, son las misma que se usan en el método de Newton, es decir, f'' debe ser continua y $f'(r) \neq 0$, pero como en general no conocemos r en forma exacta verificamos si $f(x_n) \approx 0$ y $f'(x_n) \neq 0$, donde x_n es nuestra aproximación a la raíz.

El siguiente ejemplo incluye un problema que vimos en un ejemplo anterior.

Ejemplo 27 Usando el método de la secante determinemos una raíz de $f(x) = \cos(x) - x$. En el ejemplo desarrollado anteriormente, usamos la aproximación inicial $x_0 = \frac{\pi}{4}$. Ahora

necesitamos dos aproximaciones iniciales. En la siguiente tabla aparecen los cálculos con $x_0 = 0.5$ y $x_1 = \frac{\pi}{4}$, y la fórmula

$$x_{n+1} = x_n - \frac{(x_n - x_{n-1})(\cos(x_n) - x_n)}{(\cos(x_n) - x_n) - (\cos(x_{n-1}) - x_{n-1})}, \quad n \geqslant 1.$$

n	x_n
0	0.5
1	0.7853981635
2	0.7363841388
3	0.7390581392
4	0.7390851493
5	0.7390851332

Observación. Al comparar los resultados de ahora con los del ejemplo anterior observamos que x_5 es exacto hasta la décima cifra decimal. Nótese que la convergencia del método de la secante es un poco más lenta en este caso que en el método de Newton, en el cual obtuvimos este grado de exactitud con x_3. Este resultado generalmente es verdadero.

El método de Newton o el método de la secante a menudo se usan para refinar las respuestas conseguidas con otra técnica, como el método de bisección. Dado que el método de Newton requiere de una buena aproximación inicial, pero por lo general da una convergencia más rápida, sirve perfectamente para el propósito antes mencionado.

2.4.1 Análisis del error

Recordemos que el error en el paso n es definido como $e_n = x_n - r$. Ahora como

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

obtenemos

$$\begin{aligned} e_{n+1} &= x_{n+1} - r \\ &= \frac{x_{n-1}f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})} - r \\ &= \frac{f(x_n)(x_{n-1} - r) - f(x_{n-1})(x_n - r)}{f(x_n) - f(x_{n-1})} \\ &= \frac{f(x_n)e_{n-1} - f(x_{n-1})e_n}{f(x_n) - f(x_{n-1})} \end{aligned}$$

podemos escribir ahora

$$e_{n+1} = \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \, \frac{\frac{f(x_n)}{e_n} - \frac{f(x_{n-1})}{e_{n-1}}}{x_n - x_{n-1}} \, e_n e_{n-1}.$$

Aplicando Taylor tenemos que

$$f(x_n) = f(r+e_n) = \underbrace{f(r)}_{=0} + e_n f'(r) + \frac{e_n^2}{2} f''(r) \ ,$$

luego $\frac{f(x_n)}{e_n} \approx f'(r) + \frac{e_n}{2} f''(r)$ y $\frac{f(x_{n-1})}{e_{n-1}} \approx f'(r) + \frac{e_{n-1}}{2} f''(r)$ de donde

$$\frac{f(x_n)}{e_n} - \frac{f(x_{n-1})}{e_{n-1}} \approx \frac{1}{2}(e_n - e_{n-1}) f''(r) \ .$$

Observemos que $x_n - x_{n-1} = e_n - e_{n-1}$, por lo tanto

$$\frac{\frac{f(x_n)}{e_n} - \frac{f(x_{n-1})}{e_{n-1}}}{x_n - x_{n-1}} \approx \frac{1}{2} f''(r)$$

y como $\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \cong \frac{1}{f'(r)}$, finalmente tenemos

$$\boxed{e_{n+1} \approx \frac{1}{2} \frac{f''(r)}{f'(r)} e_n e_{n-1} = C e_n e_{n-1}} \tag{2.19}$$

2.5 Método de la posición falsa

Este método es conocido también como método de *Regula Falsi*, con el generamos aproximaciones del mismo modo que el de la secante, pero mezclado con el método de bisección, ofrece por lo tanto una prueba para asegurarse de que la raíz quede entre dos iteraciones sucesivas.

Primero elegimos las aproximaciones iniciales x_0 y x_1 con $f(x_0) \cdot f(x_1) < 0$. La aproximación x_2 se escoge de la misma manera que en el método de la secante. Para determinar con cual secante calculamos x_3 verificamos el signo de $f(x_2) \cdot f(x_1)$, si este valor es negativo entonces $[x_1, x_2]$ contiene una raíz y eligiremos x_3 como la intersección del eje x con la recta que une $(x_1, f(x_1))$ y $(x_2, f(x_2))$; si no elegimos x_3 como la intersección del eje x con la recta que une $(x_0, f(x_0))$ y $(x_2, f(x_2))$.

Ejemplo 28 La siguiente tabla contiene los resultados del método de Regula Falsi aplicado a la función $f(x) = \cos(x) - x$ con las mismas aproximaciones iniciales que utilizamos para el método de la secante en el Ejemplo 27. Nótese que las aproximaciones son iguales en x_3 y que en el método de Regula Falsi requiere una iteración más para alcanzar la misma exactitud que la de la Secante.

2.6 Métodos iterativos de punto fijo

Describimos ahora otro tipo de métodos para encontrar raíces de ecuaciones, nos referimos a algunos *métodos iterativos de punto fijo*.

Un *punto fijo* de una función g es un punto p para el cual $g(p) = p$. En esta sección estudiaremos el problema de encontrar las soluciones a los problemas de punto fijo y la conexión entre estos y la búsqueda de la raíz que deseamos resolver.

n	x_n
0	0.5
1	0.7353981635
2	0.7363841388
3	0.7390581392
4	0.7390848638
5	0.7390851305
6	0.7390851332

El problemas de búsqueda de raíces y el problema de punto fijo son clases equivalentes en el siguiente sentido

"Dado un problema de buscar una raíz de $f(x) = 0$, *podemos definir una función* g *con un punto fijo en* p *de diversas formas; por ejemplo, como* $g(x) = x - f(x)$ *o como* $g(x) = x + 3f(x)$. *Si la función* g *tiene un punto fijo en* p, *es decir,* $g(p) = p$, *entonces la función definida, por ejemplo, como* $g(x) = x + f(x)$ *o más general* $g(x) = x + \psi(x)f(x)$, *con* $\psi(p) \neq 0$, *tiene un cero en* p*".*

Aunque los problemas que deseamos resolver vienen en forma de búsqueda de raíces, la forma de método iterativo de punto fijo puede ser más fácil de analizar; algunas opciones de punto fijo dan origen a técnicas poderosas de búsqueda de raíces.

Lo primero que debemos de hacer es acostumbrarnos a este tipo de problema, y decidir cuando una función tiene un punto fijo y cómo podemos aproximar los puntos fijos con determinado grado de precisión.

Ejemplo 29 Resolver la ecuación $3x^2 - e^x = 0$ es equivalente a $3x^2 = e^x$. Graficamente, vemos que existe una solución $x > 0$, la cual es la intersección de los gráficos de las funciones $f(x) = 3x^2$ y $g(x) = e^x$.

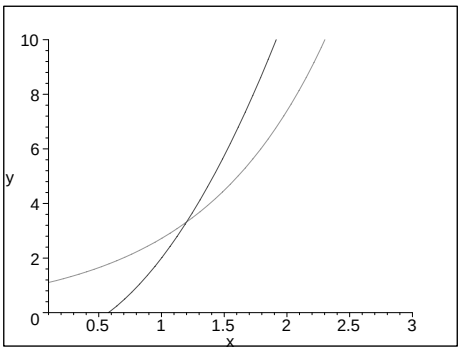


gráfico de $f(x) = 3x^2$ y $g(x) = e^x$

Despejando, obtenemos las funciones $x = \psi_1(x) = \log(3x^2)$, $x = \psi_2(x) = \frac{e^x}{3x}$, $x = \psi_3(x) = \frac{\sqrt{3}}{3}e^{x/2}$, etc.

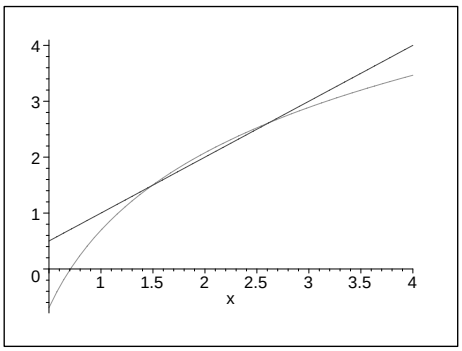


gráfico de $\psi_1(x) = \log(3x^2)$

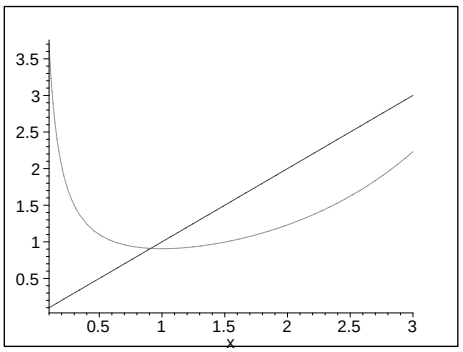


gráfico de $\psi_2(x) = \frac{e^x}{3x}$

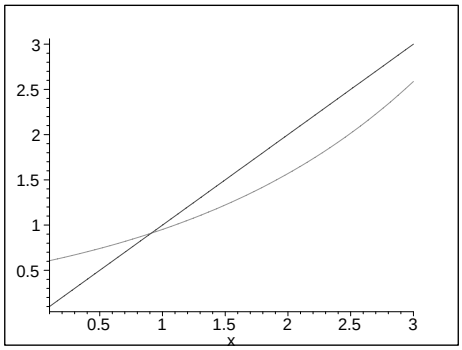


gráfico de $\psi_3(x) = \frac{\sqrt{3}}{3}e^{x/2}$

Ejemplo 30 La función $g\,(x) = x^2 - 2$ para $-2 < x < 3$, posee puntos fijos en $x = -1$ y $x = 2$, como se puede ver fácilmente resolviendo la ecuación cuadrática $g(x) = x$.

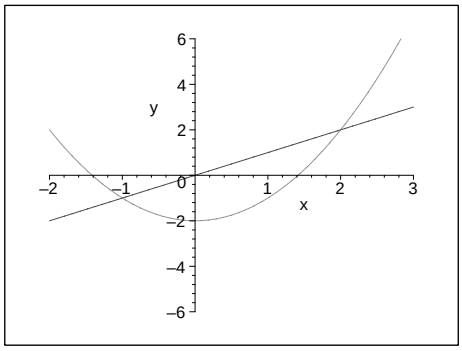


gráfico de $g(x) = x^2 - 2$

Para el problema de existencia de punto fijo, tenemos el siguiente resultado.

Teorema 2.7 *Sea* $g : [a, b] \longrightarrow [a, b]$ *una función continua, entonces* g *posee un punto fijo en* $[a, b]$.

Demostración. Usamos el Teorema del Valor Intermedio, para lo cual definimos la función $h(x) = g(x) - x$. Tenemos, $h(a) = g(a) - a \geqslant 0$ y $h(b) = g(b) - b \leqslant 0$, y siendo h continua, se sigue existe $c \in [a, b]$ tal que $h(c) = 0$, es decir, $g(c) = c$.

El siguiente teorema contiene condiciones suficientes para la existencia y unicidad del punto fijo.

Corolario 2.1 *Supongamos que* $g : [a,b] \longrightarrow [a,b]$ *es continua en* $[a,b]$ *y derivable en* (a,b). *Supongamos también que existe una constante positiva* $0 \leqslant \lambda < 1$ *con* $|g'(x)| \leqslant \lambda$, *para todo* $x \in (a,b)$ *(*$\lambda = \max\{|g'(x)| \ : \ x \in [a,b]\}$*), entonces el punto fijo* x_g *de* g *en* $[a,b]$ *es único. Además, si elegimos* $x_0 \in (a,b)$ *arbitrario, entonces la sucesión* $(x_n)_{n\in\mathbb{N}}$ *definida por*

$$x_{n+1} = g(x_n), \ n \geqslant 0$$

converge al único punto fijo x_g *de* g.

Más general, si $g : [a,b] \longrightarrow [a,b]$ *es tal que* $|g(x) - g(y()| \leqslant \lambda|x-y|$ *para todo* $x, y \in [a,b]$, *con* $0 \leqslant \lambda < 1$, *entonces* g *tiene un único punto fijo* $x_g \in [a,b]$. *Además, dado* $x_0 \in [a,b]$, *la sucesión*

$$x_{n+1} = g(x_n), \ n \geqslant 0$$

converge a x_g.

Demostración. Por el teorema anterior sabemos que g tiene un punto fijo. Ahora como $|g'(x)| \leqslant \lambda < 1$, por el Teorema del Valor Medio, tenemos

$$|g(x) - g(y)| = |g'(c)| \ |x-y| \leqslant \lambda|x-y|$$

para todo $x, y \in [a,b]$, donde c está entre x e y. Ahora bien, supongamos que existen dos puntos fijos x_g y $\overline{x}_g$ para g. Tenemos entonces que

$$|x_g - \overline{x}_g| = |g(x_g) - g(\overline{x}_g)| \leqslant \lambda|x_g - \overline{x}_g|$$

y como $0 \leqslant \lambda < 1$, se debe tener que $x_g = \overline{x}_g$.

Ahora, tenemos

$$\begin{aligned}
|x_n - x_g| &= |g(x_{n-1}) - g(x_g)| \leqslant \lambda|x_{n-1} - x_g| \\
|x_{n-1} - x_g| &= |g(x_{n-2}) - g(x_g)| \leqslant \lambda|x_{n-2} - x_g| \\
|x_{n-2} - x_g| &= |g(x_{n-3}) - g(x_g)| \leqslant \lambda|x_{n-3} - x_g| \\
&\vdots \\
|x_1 - x_g| &= |g(x_0) - g(x_g)| \ \leqslant \lambda|x_0 - x_g| \ ,
\end{aligned}$$

de donde obtenemos que $|x_n - x_g| \leqslant \lambda^n|x_0 - x_g|$ y como $\lambda^n \longrightarrow 0$ cuando $n \longrightarrow \infty$, el resultado se sigue.

Observación. El teorema anterior no sólo nos dice que bajo sus condiciones existe un punto fijo único, además nos dice cómo podemos encontrarlo, usando la sucesión generada a partir de la función g con una condición inicial arbitraria.

Corolario 2.2 *Si g satisface las hipótesis del Teorema anterior, cotas para el error que supone utilizar x_n para aproximar x_g son dadas por*

$$|x_n - x_g| \leqslant \lambda^n \max\{x_0 - a,\, b - x_0\}$$

y

$$|x_n - x_g| \leqslant \frac{\lambda^n}{1-\lambda}\,|x_1 - x_0|\,.$$

Demostración. Para tener la primera de las cotas, sólo basta observar que $|x_n - x_g| \leqslant \lambda^n |x_0 - x_g|$ y como x_g y x_0 están entre a y b se tiene que $|x_n - x_0| \leqslant \lambda^n |x_g - x_0| \leqslant \lambda^n \max\{x_0 - a, b - x_0\}$, como queríamos probar.

Para obtener la segunda cota observemos que si $n > m$, digamos $n = m + k$, con $k \geqslant 1$, entonces

$$|x_n - x_m| = |x_{m+k} - x_m| \leqslant |x_{m+k} - x_{m+k-1}| + |x_{m+k-1} - x_{m+k-2}| + \cdots + |x_{m+1} - x_m| \quad (2.20)$$

Por otra parte tenemos que

$$\begin{aligned}
|x_{n+1} - x_n| &= |g(x_n) - g(x_{n-1})| \\
&\leqslant \lambda |x_n - x_{n-1}| \\
&\leqslant \lambda |g(x_{n-1}) - g(x_{n-2})| \\
&\leqslant \lambda^2 |x_{n-1} - x_{n-2}| \\
&\leqslant \lambda^2 |g(x_{n-2}) - g(x_{n-3})| \\
&\leqslant \lambda^3 |x_{n-2} - x_{n-3}| \\
&\vdots \\
&\leqslant \lambda^n |x_1 - x_0|\,,
\end{aligned}$$

es decir, $|x_{n+1} - x_n| \leqslant \lambda^n |x_1 - x_0|$ para todo $n \geqslant 0$. Aplicando esto a la desigualdad en (2.20) obtenemos

$$\begin{aligned}
|x_n - x_m| &= |x_{m+k} - x_m| \leqslant \lambda^{m+k-1}|x_1 - x_0| + \lambda^{m+k-2}|x_1 - x_0| + \cdots + \lambda^m |x_1 - x_0| \\
&= \lambda^m |x_1 - x_0|\,(\lambda^{k-1} + \lambda^{k-2} + \cdots + 1) \\
&\leqslant \lambda^m |x_1 - x_0|\,(1 + \lambda^2 + \cdots + \lambda^j + \cdots) \\
&= \frac{\lambda^m}{1-\lambda}\,|x_1 - x_0|
\end{aligned}$$

como deseábamos probar.

Observación Ambas desigualdades del Corolario relacionan la razón con la que $(x_n)_{n\in\mathbb{N}}$ converge en la cota λ de la primera derivada. La razón de convergencia depende del factor λ^n. Cuando más pequeño sea el valor de λ, más rápida será la convergencia, la cual puede ser

muy lenta sí λ es próxima de 1. En el siguiente ejemplo, consideraremos los métodos de punto fijo para algunos de los ejemplos vistos anteriormente.

Observación. Podemos usar como criterio de parada de un método iterativo de punto fijo como los anteriores los siguientes. Sea $\varepsilon > 0$ una tolerancia dada, entonces paramos las iteraciones si

1. $\lambda^n \max\{x_0 - a,\, b - x_0\} \leqslant \varepsilon$,
2. $\frac{\lambda^n}{1-\lambda} |x_1 - x_0| \leqslant \varepsilon$,
3. $|x_{n+1} - x_n| \leqslant \varepsilon$
4. otros que sean razonables de aplicar.

Ejemplo 31 Sea $g(x) = \frac{x^2-1}{3}$ en $[-1, 1]$. El mínimo absoluto de g se alcanza en $x = 0$ y se tiene que $g(0) = -\frac{1}{3}$. De manera análoga el máximo absoluto de g se alcanza en $x = \pm 1$ y $g(\pm 1) = 0$. Además, g es derivable y $|g'(x)| = \left|\frac{2x}{3}\right| \leqslant \frac{2}{3}$ para todo $x \in (-1, 1)$, luego $g(x)$ satisface las hipótesis del Teorema 2.1, en consecuencia $g(x)$ posee un único punto fijo x_g en $[-1, 1]$, y si elegimos $x_0 \in]-1, 1[$ arbitrario, la sucesión de puntos $x_{n+1} = g(x_n) \longrightarrow x_g$ cuando $n \to \infty$.

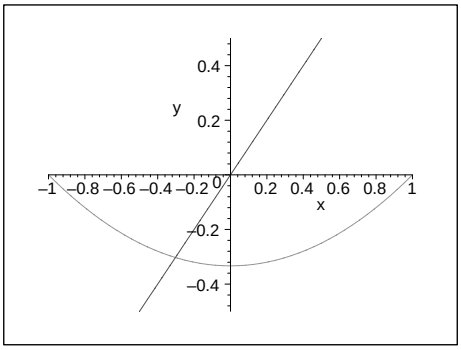


gráfico de $g(x) = \frac{x^2-1}{3}$

En este ejemplo el punto fijo p puede ser determinado algebraicamente por medio de la fórmula para resolver ecuaciones cuadráticas, obteniéndose $p = g(p) = \frac{p^2-1}{3}$, de donde $p^2 - 3p - 1 = 0$ y resolviendo la ecuación obtenemos que $p = \frac{1}{2}\left(3 - \sqrt{13}\right) \approx -0.3027756377$, la otra raíz es $q = \frac{1}{2}(3 + \sqrt{13}) \approx 3.302775638$ que está fuera del intervalo $[-1, 1]$. El lector puede verificar esto usando $x_0 = 0$ y $\varepsilon = 10^{-5}$.

Observación. Note que $g(x)$ también posee un punto fijo único $q = \frac{1}{2}\left(3 + \sqrt{13}\right)$ en el intervalo $[3, 4]$. Sin embargo, $g'(q) > 1$, así que g no satisface las hipótesis del Teorema en $[3, 4]$. Esto demuestra que esas hipótesis son suficientes pero no necesarias.

Ejemplo 32 Sea $g(x) = 3^{-x}$. Puesto que $g'(x) = -3^{-x} \ln 3 < 0$ en $[0, 1]$, la función es decreciente en ese intervalo. Por lo tanto $g(1) = \frac{1}{3} \leqslant g(x) \leqslant 1 = g(0)$ para todo $x \in [0, 1]$, por lo cual g posee un punto fijo. No podemos garantizar la unicidad del punto fijo usando el Teorema 2.1, sin embargo como g es estrictamente decreciente dicho punto fijo debe ser único.

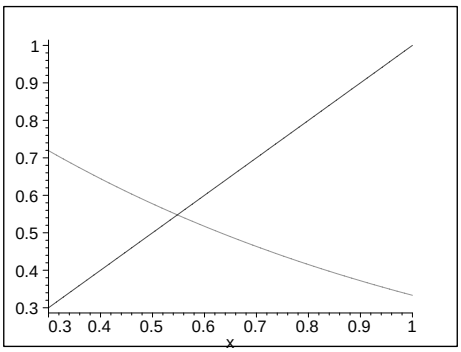


gráfico de $g(x) = 3^{-x}$

Ejemplo 33 La ecuación $x^3 + 4x^2 - 10 = 0$ posee una única raíz en $[1, 2]$. Hay muchas formas para convertir dicha ecuación en la forma $x = g(x)$ mediante simple manejo algebraico. Por ejemplo, para obtener la función que se describe en (c), podemos manipular ecuación $x^3 + 4x^2 - 10 = 0$, por ejemplo, podemos escribir $4x^2 = 10 - x^3$, por lo tanto $x^2 = \frac{1}{4}\left(10 - x^3\right)$, luego $x = g_3(x) = \pm\frac{1}{2}\left(10 - x^3\right)^{1/2}$.

Para obtener una solución positiva, elegimos $g_3(x)$ con el signo positivo. Se deja al lector deducir las funciones que se indica a continuación, pero debemos verificar que el punto fijo de cada una sea realmente una solución de la ecuación original $x^3 + 4x^2 - 10 = 0$.

(a) $x = g_1(x) = x - x^3 - 4x^2 + 10.$

(b) $x = g_2(x) = \left(\frac{10}{x} - 4x\right)^{1/2}$.

(c) $x = g_3(x) = \frac{1}{2}\left(10 - x^3\right)^{1/2}$.

(d) $x = g_4(x) = \left(\frac{10}{4+x}\right)^{1/2}$

(e) $x = g_5(x) = x - \frac{x^3+4x^2-10}{3x^2+8x}$.

Con $p_0 = 1.5$ la siguiente tabla proporciona los resultados del método de iteraciones de punto fijo para las cinco opciones de g.

La raíz exacta es $r = 1.365230013\ldots$, según se señalo anteriormente. Al comparar los resultados del algoritmo de bisección aplicado para resolver la misma ecuación, observamos que se obtuvieron excelentes resultados con las opciones (3), (4) y (5), ya que el método de bisección requiere 27 iteraciones para garantizar la exactitud. Conviene señalar que la opción (1) ocasiona divergencia y que la opción (2) se torna indefinida en $\mathbb{R}$ ya que contiene la raíz cuadrada de un número negativo.

Aún cuando las funciones de este ejemplo son problemas de punto fijo para el mismo problema de búsqueda de raíz, difieren como métodos para aproximar la solución a este tipo de problemas. Su propósito es ilustrar la pregunta que es preciso responder a la siguiente pregunta Cómo podemos encontrar un problema de punto fijo capáz de producir una sucesión convergente rápidamente a una solución de un problema de búsqueda de raíz?

Los siguientes ejemplos nos dan algunas pistas sobre los procedimientos que deberíamos seguir y quizá lo más importante, algunos que debemos excluir.

n	g_1	g_2	g_3	g_4	g_5
0	1.5	1.5	1.5	1.5	1.5
1	-0.875	0.8165	1.286953768	1.348399725	1.373333333
2	6.732	2.9969	1.402540804	1.367376372	1.365230015
3	-469.7	$(-8.65)^{1/2}$	1.345458374	1.364957015	1.365230014
4	1.03×10^8		1.375170253	1.365264748	1.365230013
5			1.360094193	1.365225594	
6			1.367846968	1.36523.576	
7			1.363887004	1.365229942	
8			1.365916734	1.365230022	
9			1.364878217	1.365230012	
10			1.365410062	1.365230014	
15			1.365223680	1.365230013	
20			1.365230236		
25			1.365230006		
30			1.365230013		

Ejemplo 34 Cuando $g_1(x) = x - x^3 - 4x^2 + 10$, tenemos que $g_1'(x) = 1 - 3x^2 - 8x$. No existe un intervalo $[a, b]$ que contenga a la raíz r para el cual se tenga $|g_1'(x)| < 1$. Aunque el Teorema 2.1 no garantiza que el método deba fallar para esta elección, tampoco tenemos razón para esperar que el método sea convergente.

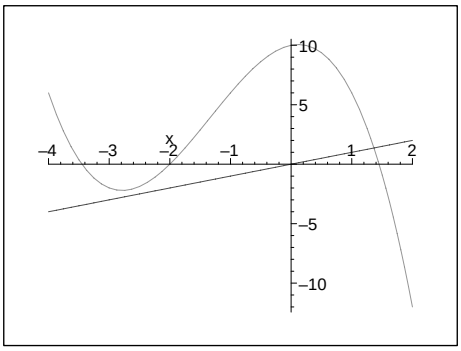


gráfico de $g_1(x) = x - x^3 - 4x^2 + 10$

Ejemplo 35 Con $x = g_2(x) = \left(\frac{10}{x} - 4x\right)^{1/2}$, podemos ver que g_2 no aplica $[1, 2]$ en $[1, 2]$ y que la sucesión $(x_n)_{n\in\mathbb{N}}$ no está definida para $x_0 > 1.58113883\ldots$. Además, tampoco existe un intervalo que contenga a r para el cual $|g_2'(x)| < 1$, pues $|g_2'(r)| \approx 3.4$.

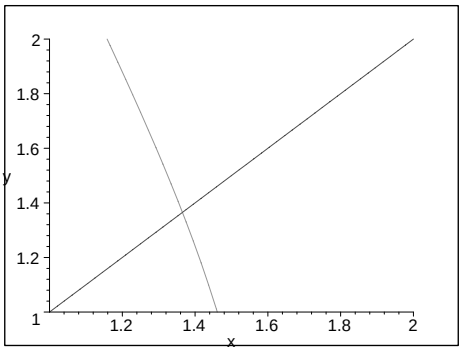


gráfico de $g_2(x) = \left(\frac{10}{x} - 4x\right)^{1/2}$

Ejemplo 36 Para $x = g_3(x) = \frac{1}{2}\left(10 - x^3\right)^{1/2}$, tenemos que $g_3'(x) = -\frac{3}{4}x^2\left(10 - x^3\right)^{-1/2} < 0$ en $[1,2]$, así que g_3 es estrictamente decreciente en $[1,2]$. Sin embargo $|g_3'(2)| \approx 2.12$ por lo cual la condición $|g_3'(x)| \leqslant \lambda < 1$ falla en $[1,2]$. Un análisis más minucioso de la sucesión $(x_n)_{n\in\mathbb{N}}$ con $p_0 = 1.5$ revela que basta considerar el intervalo $[1, 1.5]$ en vez de $[1,2]$. En este intervalo la función sigue siendo estrictamente decreciente, pero además $1 < 1.28 \approx g_3(1.5) \leqslant g_3(x) \leqslant g_3(1) = 1.5$ para todo $x \in [1, 1.5]$ en vez de $[1,2]$. Esto demuestra que g_3 aplica el intervalo $[1, 1.5]$ en si mismo. Además $|g_3'(x)| \leqslant 0.66 < 1$, por lo cual el Teorema 2.1 nos garantiza la unicidad del punto fijo y la convergencia del método iterativo.

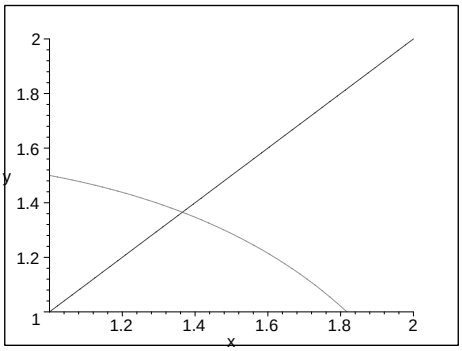


gráfico de $g_3(x) = \frac{1}{2}\left(10 - x^3\right)^{1/2}$

Ejemplo 37 Para $x = g_4(x) = \left(\frac{10}{4+x}\right)^{1/2}$, tenemos $|g_4'(x)| = \left|\frac{-5}{\sqrt{10}(4+x)^{3/2}}\right| \leqslant \frac{5}{\sqrt{10}(5)^{3/2}} < 0.15$ para todo $x \in [1,2]$. La cota en la magnitud de g_4' es mucho menor que magnitud de g_3' lo cual explica la convergencia mas rápida que se obtiene con g_4.

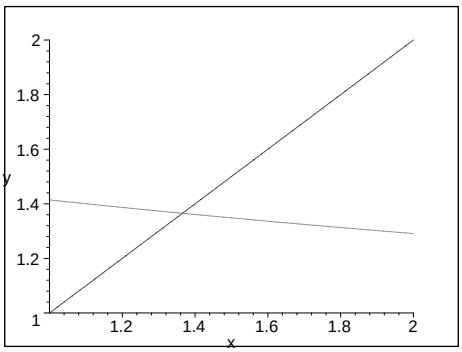


gráfico de $g_4(x) = \left(\frac{10}{4+x}\right)^{1/2}$

2.7 Métodos iterativos de punto fijo en varias variables

Extendemos ahora el resultado de convergencia métodos iterativos de punto fijo a funciones de varias variables. Para ellos tenemos el siguiente teorema.

Teorema 2.8 *Sea $C \subseteq \mathbb{R}^n$ un conjunto cerrado y acotado, y sea $F : C \longrightarrow \mathbb{R}^n$ una función continua. Supongamos que $F(C) \subseteq C$, entonces F posee un punto fijo. Además, si $\|JF(x)\| \leqslant \lambda < 1$ para todo $x \in C$ entonces el punto fijo es único, y el proceso iterativo*

$$x_{n+1} = F(x_n)\,, \qquad n \geqslant 0\,.$$

con $x_0 \in C$ arbitrario, converge al único punto fijo.

Demostración. Basta demostrar el siguiente resultado más débil .

Introducimos el siguiente concepto, que nos ayudara en los siguientes teoremas.

Definición 2.1 *Sea $A \subset \mathbb{R}^n$. Decimos que una función $F : A \longrightarrow \mathbb{R}^n$ es una contracción si existe una constante $0 \leqslant \lambda < 1$ tal que para cada $x, y \in A$ se tiene que*

$$\|F(x) - F(y)\| \leqslant \lambda \|x - y\|.$$

La menor constante λ que satisface la desigualdad anterior es llamada la constante de Lipschitz de F.

Teorema 2.9 *Sea $A \subset \mathbb{R}^n$ un conjunto cerrado y acotado, y sea $F : A \longrightarrow \mathbb{R}^n$ una contracción, tal que $F(A) \subseteq A$, entonces F tiene un único punto fijo en A. Además, dado $x_0 \in A$ arbitrario, la sucesión $(x_n)_{n\in\mathbb{N}}$ definida por $x_{n+1} = F(x_n)$, con $n \geqslant 0$, converge al único punto fijo x_F de F. Además, si elegimos x_N como una aproximación a x_F, entonces se tiene*

$$||x_N - x_F|| \leqslant \frac{\lambda^N}{1-\lambda}\,||x_0 - x_1|| \tag{2.21}$$

Demostración. Tenemos que $\|F(x) - F(y)\| \leqslant \lambda\|x - y\|$ para todo $x, y \in A$. Vamos a demostrar que dado $x_0 \in A$ arbitrario se tiene que la sucesión $(x_n)_{n\in\mathbb{N}}$ definida por $x_{n+1} =$

$F(x_n),\ n \geqslant 0$, es una sucesión de Cauchy, y siendo A cerrado y acotado ella es una sucesión convergente, cuyo límite es un punto fijo de F. Tenemos

$$\begin{aligned}
\|x_{n+1} - x_n\| &= \|F(x_n) - F(x_{n-1})\| \\
&\leqslant \lambda\|x_n - x_{n-1}\| \\
&\leqslant \lambda\|F(x_{n-1}) - F(x_{n-2})\| \\
&\leqslant \lambda^2\|x_{n-1} - x_{n-2}\| \\
&= \lambda^2\|F(x_{n-2}) - F(x_{n-3})\| \\
&\leqslant \lambda^3\|x_{n-2} - x_{n-3}\| \\
&\vdots \\
&\leqslant \lambda^n\|x_1 - x_0\| \ ,
\end{aligned}$$

es decir, $\|x_{n+1} - x_n\| \leqslant \lambda^n\|x_1 - x_0\|$ y como $0 \leqslant \lambda < 1$ se sigue que $\|x_{n+1} - x_n\| \longrightarrow 0$ cuando $n \longrightarrow \infty$. Ahora sean $m > n$, digamos $m = n + k$, con $k \geqslant 1$. Tenemos

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

esto es,

$$\|x_{n+k} - x_n\| \leqslant \frac{\lambda^n}{1-\lambda}\|x_1 - x_0\|$$

y como $\lim_{n\to\infty} \lambda^n = 0$ se sigue que $(x_n)_{n\in\mathbb{N}}$ es una sucesión de Cauchy, por lo tanto es una sucesión convergente. Sea $x_F = \lim_{n\to\infty} x_n$. Ahora como F es una contracción es continua, y $x_F = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} F(x_n) = F(\lim_{n\to\infty} x_n) = F(x_F)$. Finalmente, supongamos que existen dos puntos fijos para F, digamos, x_F y x'_F, entonces

$$\|x_F - x'_F\| = \|F(x_F) - F(x'_F)\| \leqslant \lambda\|x_F - x'_F\|$$

y siendo $0 \leqslant \lambda < 1$, se tiene que $x_F = x'_F$.

2.7.1 Análisis de error para métodos iterativos de punto fijo

En esta sección estudiaremos el orden de convergencia de los esquemas de iteración funcional y con el propósito de obtener una rápida convergencia, redescubriremos el método de Newton. También estudiaremos los métodos para acelerar la convergencia de Newton en casos especiales. Para poder realizar lo antes mencionado necesitamos un procedimiento para medir la rapidez con la que converge una sucesión dada.

Definición 2.2 *Supongamos que* $(p_n)_{n\in\mathbb{N}}$ *es una sucesión que converge a* p*, con* $p_n \neq p$ *para todo* n *suficientemente grande. Si existen constantes positivas* α *y* λ *tales que*

$$\lim_{n\to\infty} \frac{|p_{n+1}-p|}{|p_n-p|^{\alpha}} = \lambda \tag{2.22}$$

decimos que la convergencia a p *de la sucesión* $(p_n)_{n\in\mathbb{N}}$ *es de orden* α *con constante de error asintótica* λ*.*

Observación. De la definción de orden de convergencia (2.22), si n es suficientemente grande, entonces

$$\frac{|p_{n+1}-p|}{|p_n-p|} \approx \lambda \iff |p_{n+1}-p| \approx \lambda\,|p_n-p|^{\alpha}\,.$$

Definición 2.3 *Decimos que un método iterativo de punto fijo* $x_{n+1}=g(x_n)$ *es de orden* α *con constante de error asintótica* λ *si, la sucesión* $(x_n)_{n\in\mathbb{N}}$*, donde* $x_{n+1}=g(x_n)$*, con* $n \geqslant 0$*, es orden* α *con constante de error asintótica* λ*.*

Observación. En general una sucesión con un orden de convergencia alto, converge más rápidamente que una con un orden de convergencia más bajo. La constante asintótica influye en la rapidez de la convergencia, pero no es tan importante como el orden de convergencia.

Respecto al orden de convergencia, hay dos que son de interés.

1. Si $\alpha = 1$, la sucesión será linealmente convergente.
2. Si $\alpha = 2$, la sucesión será cuadráticamente convergente.

En el siguiente ejemplo se compara una sucesión linealmente convergente con una de orden cuadrático.

Ejemplo 38 Supongamos que $(p_n)_{n\in\mathbb{N}}$ y $(\hat{p}_n)_{n\in\mathbb{N}}$ convergen a cero, siendo la convergencia de $(p_n)_{n\in\mathbb{N}}$ lineal con $\lim_{n\to\infty}\frac{|p_{n+1}|}{|p_n|}=0.5$ y la convergencia de $(\hat{p}_n)_{n\in\mathbb{N}}$ es cuadrática con $\lim_{n\to\infty}\frac{|\hat{p}_{n+1}|}{|\hat{p}_n|^2}=0.5$. Por razones de simplicidad supongamos que $\frac{|\hat{p}_{n+1}|}{|\hat{p}_n|^2}\approx 0.5$ y que $\frac{|p_{n+1}|}{|p_n|}\approx 0.5$. Así en la convergencia lineal obtenemos que

$$|p_n-0| = |p_n| \approx 0.5\,|p_{n-1}| \approx (0.5)^2\,|p_{n-2}| \approx (0.5)^3\,|p_{n-3}| \approx \cdots \approx (0.5)^n\,|p_0|\,,$$

mientras que en la convergencia cuadrática se tiene que

$$|\hat{p}_n-0| = |\hat{p}_n| \approx 0.5\,|\hat{p}_{n-1}|^2 \approx 0.5\,(0.5|\hat{p}_{n-2}|^2)^2 = (0.5)^3|\hat{p}_{n-2}|^{2^2} \approx \cdots \approx (0.5)^{2^n-1}|\hat{p}_0|^{2^n}\,.$$

Claramente esta última converge más rápido que la primera.

La siguiente tabla muestra la rapidez de la convergencia lineal y cuadrática cuando $|p_0| = |\hat{p}_0| = 1$.

La sucesión con convergencia cuadrática es del orden de 10^{-38} en el séptimo término. Por otra parte, se necesitan 126 términos por lo menos para poder garantizar esta precisión en la sucesión con convergencia lineal.

n	Lineal	Cuadrática
1	5.0000×10^{-1}	5.0000×10^{-1}
2	2.5000×10^{-1}	1.2500×10^{-1}
3	1.2500×10^{-1}	7.8125×10^{-3}
4	6.2500×10^{-2}	3.0518×10^{-5}
5	3.1250×10^{-2}	4.6566×10^{-10}
6	1.5625×10^{-2}	1.0842×10^{-19}
7	7.8125×10^{-3}	5.8775×10^{-39}

Teorema 2.10 (orden de convergencia de punto fijo) *Sea* $g : [a,b] \longrightarrow [a,b]$ *continua. Supongamos que* g' *es continua en* $]a,b[$ *y que existe una constante positiva* $0 \leqslant \lambda < 1$ *tal que* $|g'(x)| \leqslant \lambda$ *para todo* $x \in]a,b[$. *Si* $g'(p) \neq 0$, *entonces para cualquier punto* $p_0 \in [a,b]$ *la sucesión*

$$p_n = g(p_{n-1}), \quad n \geqslant 1$$

converge linealmente al único punto fijo $p \in [a,b]$.

Observación. El teorema anterior establece que en el caso de los métodos de punto fijo, la convergencia de orden superior sólo puede ocurrir si $g'(p) = 0$. El siguiente resultado describe otras condiciones que garantizan la convergencia cuadrática que deseamos.

Teorema 2.11 *Sea* p *un punto fijo de* $g(x)$. *Supongamos que* $g'(p) = 0$ *y que* g'' *es continua y acotada por* M *en un intervalo abierto que contiene a* p. *Entonces existe* $\delta > 0$ *tal que para* $p_0 \in [p-\delta, p+\delta]$ *la sucesión definida por* $p_n = g(p_{n-1})$, *con* $n \geqslant 1$, *converge al menos cuadráticamente a* p. *Además, para valores suficientemente grandes de* n *se tiene que*

$$|p_{n+1} - p| < \frac{M}{2} |p_n - p|^2 .$$

Los teoremas anteriores nos indican que nuestra investigación de los métodos de punto fijo cuadráticamente convergentes deberían conducirnos a funciones cuyas derivadas son anulan en el punto fijo.

La manera más fácil de plantear un problema de punto fijo relacionado con la búsqueda de raíces de $f(x) = 0$ consiste en restar un múltiplo de $f(x)$ (que desaparecerá en la raíz). Por lo tanto, a continuación consideraremos un método iterativo de punto fijo de la forma

$$p_n = g(p_{n-1}), \quad n \geqslant 1,$$

con g es de la forma

$$g(x) = x - \phi(x) f(x),$$

donde $\phi(x)$ es una función derivable, y sobre la cual imponemos condiciones más adelante.

Para que el procedimiento iterativo dado por g sea cuadráticamente convergente, es necesario tener que $g'(p) = 0$. Dado que

$$g'(x) = 1 - \phi'(x) f(x) - f'(x) \phi(x)$$

y como $f(p) = 0$ tenemos que

$$g'(p) = 1 - f'(p)\,\phi(p)$$

entonces $g'(p) = 0$ si y sólo si $\phi(p) = \frac{1}{f'(p)}$.

Es razonable suponer que $\phi(x) = \frac{1}{f'(x)}$, lo cual garantizará que $\phi(p) = \frac{1}{f'(p)}$. En este caso, el procedimiento natural para producir la convergencia cuadrática será

$$p_{n+1} = g(p_n) = p_n - \frac{f(p_n)}{f'(p_n)}$$

que es el método de Newton.

En el procedimiento anterior, impusimos la restricción de que $f'(p) \neq 0$, donde p es la solución de $f(x) = 0$. Conforme a la definición del método de Newton, es evidente que pueden surgir dificultades si $f'(p_n)$ tiene un cero simultáneamente con $f(p_n)$. En particular, el método de Newton y el de la secante ocasionaran problemas si $f'(p) = 0$ cuando $f(p) = 0$. Para examinar mas a fondo estas dificultades, damos la siguiente definición.

Definición 2.4 *Un cero p de $f(x) = 0$ es de multiplicidad m si para $x \neq p$ podemos escribir $f(x) = (x-p)^m q(x)$, con $\lim_{x\to p} q(x) \neq 0$.*

Teorema 2.12 *Una función $f : [a,b] \longrightarrow \mathbb{R}$, derivable m veces, tiene un cero de multiplicidad m en p si $f(p) = f'(p) = f''(p) = \cdots = f^{(m-1)}(p) = 0$ y $f^{(m)}(p) \neq 0$.*

El siguiente ejemplo muestra cómo la convergencia cuadrática posiblemente no ocurra si el cero no es simple.

Ejemplo 39 La función $f(x) = e^x - x - 1$ posee un cero de multiplicidad 2 en $p = 0$, pues $f(0) = f'(0) = 0$ y $f''(0) = 1$. De hecho podemos expresar $f(x)$ de la siguiente forma

$$f(x) = (x-0)^2\,\frac{e^x - x - 1}{x^2}$$

empleando la regla de L'Hospital, obtenemos que

$$\lim_{x\to 0} \frac{e^x - x - 1}{x^2} = \lim_{x\to 0} \frac{e^x - 1}{2x} = \lim_{x\to 0} \frac{e^x}{2} = \frac{1}{2}$$

la siguiente tabla muestra los términos que se generaron con el método de Newton aplicado a f con $p_0 = 1$.

Teorema 2.13 *Supongamos que F define un método iterativo de punto fijo y que $F(p) = p$, $F'(p) = F''(p) = \cdots = F^{(m-1)}(p) = 0$ y $F^{(m)}(p) \neq 0$, entonces el método iterativo de punto fijo definido por F tiene orden de convergencia m.*

Demostración. Sea p_0 un punto arbitrario y definamos la sucesión $(p_n)_{n\in\mathbb{N}}$ por $p_{n+1} = F(p_n)$, $n \geqslant 0$. Recordemos que el error es definido por

$$e_n = p_n - p.$$

n	p_n	n	p_n
0	1.0	9	2.7750×10^{-3}
1	0.58198	10	1.3881×10^{-3}
2	0.31906	11	6.9411×10^{-4}
3	0.16800	12	3.4703×10^{-4}
4	0.08635	13	1.7416×10^{-4}
5	0.04380	14	8.8041×10^{-5}
6	0.02206	15	4.2610×10^{-5}
7	0.01107	16	1.9142×10^{-6}
8	0.005545		

Tenemos entonces que

$$\begin{aligned} e_{n+1} &= p_{n+1} - p \\ &= F(p_n) - F(p) \\ &= F(p + e_n) - F(p) \\ &\overset{\text{Taylor}}{=} F(p) + F'(p)e_n + F''(p)\frac{e_n^2}{2!} + \cdots + F^{(m-1)}(p)\frac{e_n^{m-1}}{(m-1)!} + F^{(m)}(\widetilde{p})\frac{e_n^m}{m!} - F(p) \end{aligned}$$

con $\widetilde{p}$ entre p y $p + e_n$. Usando la hipótesis nos queda $e_{n+1} = F^{(m)}(\widetilde{p})\frac{e_n^m}{m!}$, de donde

$$\lim_{n\to\infty} \frac{\|e_{n+1}\|}{\|e_n\|^m} = \frac{\|F^{(m)}(p)\|}{m!}$$

pues $\widetilde{p} \to p$ cuando $p_n \to p$.

2.8 Raíces múltiples

Un método para tratar el problema de raíces múltiples consiste en definir la función $\mu(x)$ por medio de

$$\mu(x) = \frac{f(x)}{f'(x)}.$$

Si p es un cero de multiplicidad m entonces podemos escribir $f(x) = (x-p)^m q(x)$, con $q(p) \neq 0$. Reemplazando, nos queda

$$\mu(x) = \frac{(x-p)^m q(x)}{m(x-p)^{m-1} q(x) + (x-p)^m q'(x)} = (x-p)\frac{q(x)}{mq(x) + (x-p)q'(x)}$$

posee un cero en $x = p$. Pero como $q(p) \neq 0$, derivando y evaluando en $x = p$, obtenemos

$$\mu'(p) = \frac{1}{m} \neq 0,$$

por lo tanto p es un cero simple de $\mu(x)$. Así podemos aplicar el método de Newton a la función $\mu(x)$ obteniendo

$$g(x) = x - \frac{\mu(x)}{\mu'(x)} = x - \frac{\frac{f(x)}{f'(x)}}{\frac{(f'(x))^2 - f(x)\ f''(x)}{(f'(x))^2}}$$

desarrollando, nos queda la fórmula iterativa

$$g(x) = x - \frac{f(x) f'(x)}{(f'(x))^2 - f(x)\ f''(x)} \quad , \tag{2.23}$$

esta es conocida como *método iterativo de Schröder*.

Si g satisface las condiciones de diferenciablidad necesarias, la iteración funcional aplicada a g será cuadráticamente convergente sin importar la multiplicidad de la raíz de f. En teoría, el único inconveniente de este método es el cálculo adicional de $f''(x)$ y el procedimiento más laborioso con que se calculan las iteraciones. Sin embargo, en la práctica la presencia de un cero múltiple puede ocasionar serios problemas de redondeo porque el denominador del método arriba consta de la diferencia de dos números que están cercanos a la raíz, es decir, diferencia de números parecidos.

Ejemplo 40 La siguiente tabla contiene las aproximaciones de la raíz doble en $x = 0$ de la función $f(x) = e^x - x - 1$, utilizando el método de Schröder y una máquina con diez dígitos de precisión. Elegimos la aproximación inicial $p_0 = 1$ de manera que las entradas pueden compararse con las de la tabla del anterior. Lamentablemente no aparece en la tabla es que no se produce mejoramiento alguno en la aproximación de la raíz $-2.8085217 \times 10^{-7}$ en los cálculos subsecuentes cuando se usa Schröder y esta máquina, ya que el numerador y el numerador se acercan a cero.

n	p_n
1	$-2.3421061 \times 10^{-1}$
2	$-8.4582788 \times 10^{-3}$
3	$-1.1889524 \times 10^{-5}$
4	$-6.8638230 \times 10^{-3}$
5	$-2.8085217 \times 10^{-3}$

Ejemplo 41 En la sección anterior resolvimos la ecuación $f(x) = x^3 + 4x^2 - 10 = 0$ para la raíz $p = 1.36523001$. Para comparar la convergencia para una raíz simple por el método de Newton y el método de Schröder, sea

i. $p_n = p_{n-1} - \dfrac{p_{n-1}^3 + 4p_{n-1}^2 - 10}{3p_{n-1}^2 + 8p_{n-1}}$ (método de Newton)

y según la fórmula del método de Schröder

ii. $p_n = p_{n-1} - \dfrac{\left(p_{n-1}^3 + p_{n-1}^2 - 10\right)\left(3p_{n-1}^2 + 8p_{n-1}\right)}{\left(3p_{n-1}^2 + 8p_{n-1}\right)^2 - \left(p_{n-1}^3 + p_{n-1}^2 - 10\right)(6p_{n-1} + 8)}$.

Cuando $p_0 = 1.5$, las tres primeras iteraciones para (i) y (ii) se incluyen en la siguiente tabla. Los resultados muestran la convergencia rápida en ambos casos.

	(i)	(ii)
p_1	1.37333333	1.35689898
p_2	1.36526201	1.36519585
p_3	1.36523001	1.36523001

Otra manera de acelerar la convergencia del método de Newton para raíces múltiples es el siguiente. Sea p una raíz múltiple de multiplicidad m de $f(x) = 0$. Definamos el siguiente método iterativo

$$N_m(x) = x - m\,\frac{f(x)}{f'(x)}\,.$$

este método tiene convergencia cuadrática en un intervalo abierto que contiene a p y tiene convergencia lineal cerca de la raíces simples de $f(x) = 0$.

2.9 Aceleración de convergencia

Existen casos en los que no tenemos convergencia cuadrática, considerando esto a continuación estudiaremos una técnica denominada *método* Δ^2 *de Aitken*, la cual sirve para acelerar la convergencia de una sucesión que sea linealmente convergente.

Supongamos que $(p_n)_{n=0}^{\infty}$ es una sucesión linealmente convergente con limite p. Vamos a construir una sucesión $(\hat{p}_n)_{n=0}^{\infty}$ convergente a p más rápidamente que $(p_n)_{n=0}^{\infty}$. Supongamos primero que los signos de $p_n - p$, $p_{n+1} - p$ y $p_{n+2} - p$ son iguales y que n es suficientemente grande para que

$$\frac{p_{n+1} - p}{p_n - p} \approx \frac{p_{n+2} - p}{p_{n+1} - p}.$$

Entonces

$$(p_{n+1} - p)^2 \approx (p_n - p)\,(p_{n+2} - p)\,,$$

por lo tanto

$$(p_{n+1})^2 - 2p_{n+1}p + p^2 \approx p_{n+2}p_n - (p_n + 2p_n)\,p + p^2,$$

al despejar p obtenemos

$$p \approx \frac{p_{n+2}p_n - (p_{n+1})^2}{p_{n+2} - 2p_{n+1} + p_n}.$$

Si sumamos y restamos los términos p_n^2 y $2p_np_{n+1}$ en el numerador podemos escribir esta última expresión de la siguiente manera

$$p \approx p_n - \frac{(p_{n+1} - p_n)^2}{p_{n+2} - 2p_{n+1} + p_n}.$$

El método Δ^2 de Aitken se basa en la suposición de que la sucesión $(\hat{p}_n)_{n=0}^{\infty}$ definida por

$$\hat{p}_n = p_n - \frac{(p_{n+1} - p_n)^2}{p_{n+2} - 2p_{n+1} + p_n},$$

conocido como *método de acelaración de Aitken,* converge más rápido a p que la sucesión original $(p_n)_{n\in\mathbb{N}}$.

Ejemplo 42 La sucesión $(p_n)_{n\in\mathbb{N}}$ definida por $p_n = \cos\left(\frac{1}{n}\right)$, converge linealmente a $p = 1$. En la siguiente tabla se incluyen los primeros términos de las sucesiones $(p_n)_{n\in\mathbb{N}}$ y $(\hat{p}_n)_{n\in\mathbb{N}}$. Observe que $(\hat{p}_n)_{n\in\mathbb{N}}$ converge mas rápidamente a $p = 1$ que $(p_n)_{n\in\mathbb{N}}$. Esto queda claro en la tabla siguiente

n	p_n	$\widehat{p}_{\,n}$
1	0.54030	0.96178
2	0.87758	0.98213
3	0.94496	0.98979
4	0.96891	0.99342
5	0.98007	0.99541
6	0.98614	
7	0.98981	

La notación Δ asociada a esta técnica tiene su origen en la siguiente definición.

Definición 2.5 *Dada la sucesión $(p_n)_{n\in\mathbb{N}}$ la diferencia Δp_n es definida por*

$$\Delta p_n = p_{n+1} - p_n.$$

Las potencias mayores se denotan por $\Delta^k p_n$ y se definen recursivamente por

$$\Delta^k p_n = \Delta\left(\Delta^{k-1} p_n\right), \quad k \geqslant 2.$$

Tomando en cuenta la definición anterior obtenemos que

$$\Delta^2 p_n = \Delta\left(p_{n+1} - p_n\right) = \Delta p_{n+1} - \Delta p_n = p_{n+2} - 2p_{n+1} + p_n\,,$$

así la fórmula de aceleración de convergencia de Aitken se puede escribir como

$$\widehat{p}_{\,n} = p_n - \frac{(\Delta p_n)^2}{\Delta^2 p_n}\,. \tag{2.24}$$

Teorema 2.14 *Supongamos que la sucesión $(p_n)_{n\in\mathbb{N}}$ converge linealmente a p y que para todos los valores de n suficientemente grandes se tiene que $(p_n - p)(p_{n+1} - p) > 0$. Entonces la sucesión $(\hat{p}_n)_{n\in\mathbb{N}}$ construida por el método de aceleración de convergencia de Aitken converge con mayor rapidez a p que la sucesión $(p_n)_{n\in\mathbb{N}}$ en el sentido que*

$$\lim_{n\to\infty} \frac{\hat{p}_n - p}{p_n - p} = 0.$$

Al aplicar un método Δ^2 modificado de Aitken a una sucesión linealmente convergente obtenida mediante iteración de punto fijo podemos acelerar la convergencia. A este procedimiento se le conoce con el nombre de método de Steffensen, y difiere un poco de la aplicación del método Δ^2 de Aitken directamente a la sucesión de iteraciones de punto fijo que convergen linealmente. El método Δ^2 de Aitken deberá construir los términos en el orden:

$$p_0, \quad p_1 = g(p_1), \quad p_2 = g(p_1), \quad \hat{p}_0 = \left\{\Delta^2\right\} p_0, \quad p_3 = g(p_2), \quad \hat{p}_1 = \left\{\Delta^2\right\} p_1$$

donde $\left\{\Delta^2\right\}$ indica que se usa le método Δ^2 de Aitken.

El método de Steffensen construye los mismos primeros cuatro términos p_0, p_1, p_2 y $\hat{p}_0$. No obstante, en este paso supone que $\hat{p}_0$ es una mejor aproximación de p que p_2 y aplica la iteración de punto fijo a $\hat{p}_0$ en vez de a p_2. Al aplicar esta notación, la secuencia generada será

$$p_0^{(0)}, \ p_1^{(0)} = g\left(p_0^{(0)}\right), \ \ p_2^{(0)} = g\left(p_1^{(0)}\right),$$

$$p_0^{(1)} = \left\{\Delta^2\right\} p_0^{(0)}, \ \ p_1^{(1)} = g\left(p_0^{(1)}\right), \ \ p_2^{(1)} = g\left(p_1^{(1)}\right),$$

$$p_0^{(2)} = \left\{\Delta^2\right\} p_0^{(1)}, \ \ p_1^{(2)} = g\left(p_0^{(2)}\right), \ \ \ldots$$

Ejemplo 43 Para resolver $x^3+4x^2-10=0$ mediante el método de Steffensen, sea $x^3+4x^2=10$ y resolvamos para x dividiendo entre $x+4$. Con este procedimiento se produce el método de punto fijo

$$g(x) = \left(\frac{10}{x+4}\right)^{1/2},$$

y $x = g(x)$ implica que $x^3 + 4x^2 - 10 = 0$.

En la siguiente tabla se muestran los valores obtenidos con el método de Steffensen con $p_0 = 1.5$.

k	p_0^k	p_1^k	p_2^k
0	1.5	1.348399725	1.367376372
1	1.365265224	1.365225534	1.365230583
2	1.365230013		

La exactitud de la iteración $p_0^2 = 1.365230013$ es de nueve cifras decimales. En este ejemplo, con el método de Steffensen se obtuvo casi la misma rapidez de convergencia que con el método de Newton.

Teorema 2.15 *Supongamos p que es un punto fijo de $g(x)$, con $g'(p) \neq 1$. Si existe $\delta > 0$ tal que g es 3 veces derivable y que g''' es continua en $[a-\delta, p+\delta]$, entonces con el método de Steffensen obtendremos convergencia cuadrática para cualquier $x \in [a-\delta, p+\delta]$.*

2.10 Problemas resueltos

Problema 2.1 Sea $g(x) = 0.1 + 0.6\cos(2x)$.

(a) Pruebe que $g(x)$ define un método iterativo convergente, encontrando explícitamente un intervalo cerrado y acotado $I = [a, b]$ tal que $g(I) \subset I$ y una constante $0 \leqslant \lambda < 1$ tal que $|g'(x)| \leqslant \lambda$ para todo $x \in I$.

(b) Estime el número de iteraciones que debe hacerse con este método de modo que se tengan 4 decimales de precisión para el punto fijo de g en I si tomamos como condición inicial el punto $x_0 = 0.4$.

(c) Use el método de Newton para encontrar una aproximación al punto fijo de $g(x)$ en I. Use como condición inicial $x_0 = 0.4$ y como criterio de parada $|x_n - g(x_n)| \leqslant 10^{-5}$.

(d) Considere una pequeña variación $\tilde{g}(x)$ de $g(x)$, donde $\tilde{g}(x) = 0.2 + 0.6\cos(2x)$. Denote por α y $\tilde{\alpha}$ los puntos fijos de g y $\tilde{g}$ respectivamente. Demuestre que

$$|\alpha - \tilde{\alpha}| \leqslant \frac{0.1}{1 - L}$$

donde L es la constante de Lipschitz de la función g Qué ocurre con las iteraciones $x_{n+1} = \tilde{g}(x_n)$, donde $x_0 = 0.45$? Explique

Solución

(a) Tenemos que $g(x) = 0.1 + 0.6\cos(2x)$. Para probar que esta función define un método iterativo convergente, debemos mostrar que

(i) Existe un intervalo cerrado y acotado I que es aplicado en si mismo por g.

(ii) Existe una constante $0 \leqslant \lambda < 1$ tal que $|g'(x)| \leqslant \lambda$ para todo $x \in I$.

Es fácil ver que podemos satisfacer ambas condiciones por separado. Por ejemplo, como

$$-1 \leqslant \cos(2x) \leqslant 1$$

para todo x, se sigue que

$$-0.5 \leqslant g(x) \leqslant 0.7\,,$$

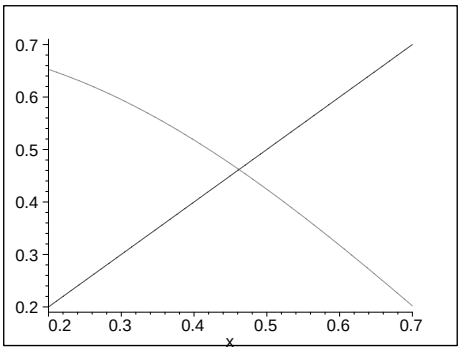


gráfico de g en $[0.2, 0.7]$

y como $g(x)$ tiene un máximo en $x = 0$, se tiene que

$$0.2 < g(0.7) \leqslant g(g(x)) \leqslant g(0) \leqslant 0.7\,.$$

También, $g'(x) = -1.2\,\mathrm{sen}(2x)$, luego (b) se satistace si

$$|x| \leqslant 0.5\,\mathrm{arcsen}(1/1.2) \approx 0.49255539\,.$$

Sin embargo ninguno de esos intervalos mencionados satisface ambas condiciones. Debemos encontrar un intervalo adecuado donde se satisfacen ambas condiciones.

No se gana mucho considerando un intervalo que no esté contenido en $[0.2, 0.7]$, pues todo los valores de $g(g(x))$ están en este intervalo. Notemos que $g(x)$ es decreciente sobre este intervalo (considere la derivada) y como $g(0.2) = 0.6526365964...$ y $g(0.7) = 0.2019802857...$ se tiene que $g([0.2, 0.7]) = [0.2019802857..., 0.6526365964...] \subset [0.2, 0.7] = I$. Luego, como puntos extremos de I son llevados en I por g, (a) se satisfecha, pero (b) requiere que el extremo derecho no sea mayor que $0.5\,\mathrm{arcsen}(1/1.2) \approx 0.49255539$, luego el punto extremo izquierdo no puede ser menor que la preimágen de este valor bajo g, la cual es aproximadamente 0.4288. Ahora, tomando $x_0 = 0.45$ se tiene que $x_1 = g(x_0) \approx 0.4729659810 < 0.473 = \tilde{x}_1$, y podemos tomar el intervalo $J = [0.45, 0.473] \subset I$, y se tiene que $g(0.45) \approx 0.4729659810$ y $g(0.473) \approx 0.4509592536$, y siendo g decreciente en ese intervalo, se tiene que $g(J) = [0.4509592536..., 0.4729659810...] \subset [0.45, 0.473]$, es decir, $g(J) \subset J$. En este intervalo J se tiene que $\max\{|g'(x)| : x \in J\} = 0.9732987256... = \lambda < 1$. Luego, el método propuesto converge para cada $x_0 \in J$.

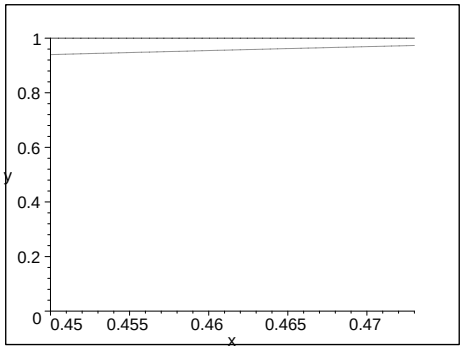


gráfico de $|g'(x)|$ en J

(b) Usaremos la fórmula $|x_n - x_g| \leqslant \frac{\lambda^n}{1-\lambda}|x_0 - x_1|$, donde $\lambda = \max\{|g'(x)| : x \in J\} = 0.9732987256...$ y x_g es el punto fijo de g que andamos buscando. Para simplificar, tomamos un $\tilde{\lambda} = 0.98 > \lambda$ en el numerador, pues tenemos que $1/(1-\lambda) \approx 37.45139595$ y $1/(1-0.98) \approx 50.00$. Luego

$$|x_n - x_g| \leqslant \frac{\lambda^n}{1-\lambda}|x_0 - x_1| \leqslant \frac{1}{1-\tilde{\lambda}}\tilde{\lambda}^n\,|x_0 - x_1| = 50\,\tilde{\lambda}^n|x_0 - x_1|$$

tomando un valor $\bar{x}_1 < x_1|$, agrandamos la diferencia $|x_0 - x_1$, por ejemplo, tomamos $\tilde{x}_1 = 0.473$, y tememos $0.72965981 \times 10^{-1} \approx |x_0 - x_1| \leqslant |x_0 - \tilde{x}_1| = |0.4 - 0.473| = 0.073$. Reemplazando esos valores en la última parte de la desigualdad anteior, obtenemos

$$50 \times (0.98)^n \times 0.073 \leqslant 10^{-4}$$

es decir,

$$3.65 \times (0.98)^n \times \leqslant 10^{-4}$$

de donde

$$(0.98)^n \leqslant 0.27397 \times 10^{-4}$$

ahora, tomando logarítmo natural, nos queda

$$n \ln(0.98) \leqslant \ln(0.27397 \times 10^{-4}) \approx -10.50507704$$

de donde, $n \geqslant \ln(0.8695652174 \times 10^{-4}) / \ln(0.98) \approx 519.9836$,. Luego, debemos tomar $n \geqslant 520$ como el número de iteraciones para asegurar que tenemos 4 decimales de precisión.

(c) Para encontrar el punto fijo de g debemos resolver la ecuación $g(x) = x$, es decir, $0.1 + 0.6\cos(2x) = x$, de donde debemos encontrar una raíz de la ecuación $f(x) = 0.1 + 0.6\cos(2x) - x = 0$. Usando el método de Newton, tenemos

$$\begin{aligned} N_f(x) &= x - \frac{f(x)}{f'(x)} \\ &= x - \frac{0.1 + 0.6\cos(2\,x) - x}{-1.2\sin(2\,x) - 1} \\ &= \frac{0.5\,(12\,x\sin(2\,x) + 1 + 6\cos(2\,x))}{6\sin(2\,x) + 5} \end{aligned}$$

Comenzando con las iteraciones, y usando que $x_{n+1} = g(x_n)$, y trabajando con 12 dígitos se tiene que

$x_1 = N_f(0.4) = 0.463425566162$, $e_1 = |x_1 - x_0| = 0.13425566162 \times 10^{-1}$

$x_2 = N_f(x_1) = 0.461786298387$, $e_2 = |x_2 - x_1| = 0.1639267775 \times 10^{-2}$

$x_3 = N_f(x_2) = 0.461785307874$, $e_3 = |x_3 - x_2| = 0.990513 \times 10^{-6}$

$x_4 = N_f(x_3) = 0.461785307873$, $e_4 = |x_4 - x_3| = 0.1 \times 10^{-11}$.

Además, $g(x_4) = 0.461785307873 = x_4$, por lo tanto, salvo errores de orden mayor que 10^{-11} se tiene que el punto fijo es $x_g = 0.461785307873$

(d) Es fácil ver graficamente que $\tilde{g}$ tiene un punto fijo $\tilde{\alpha}$.

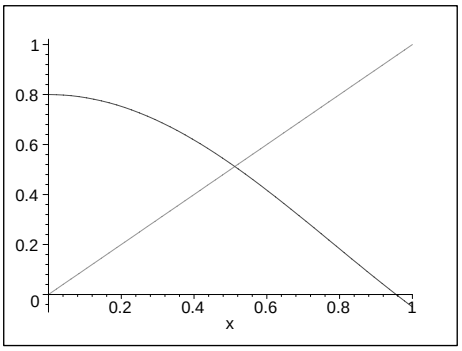


gráfico de $\tilde{g}$ en $[0, 1]$

Solución Numérica. Podemos usar el método de Newton para aproximar $\tilde{\alpha}$. Para ello debemos resolver la ecuación $\tilde{g}(x) = x$, es decir, si definimos la función $r(x) = 0.2 + 0.6\,\cos(2x) - x$, dedemos encontrar una raíz de r. Ahora, como $r(x) = 0.2 + 0.6\cos(2\,x) - x$, se tiene que $r'(x) = -1.2\sin(2\,x) - 1$ y

$$N_r(x) = x - \frac{r(x)}{r'(x)}$$

viene dada por

$$N_r(x) = x - \frac{0.2 + 0.6\cos(2\,x) - x}{-1.2\sin(2\,x) - 1} = \frac{6\,x\sin(2\,x) + 1 + 3\cos(2\,x)}{6\sin(2\,x) + 5}$$

Tomando $y_0 = 0.4$, tenemos

$$y_1 = N_r(y_0) = 0.5171651042\,,\ E_1 = 0.1171651042$$
$$y_2 = N_r(y_1) = 0.5119943202\,,\ E_2 = 0.0051707840$$
$$y_3 = N_r(y_2) = 0.5119861755\,,\ E_3 = 0.81447\,10^{-5}$$
$$y_4 = N_r(y_3) = 0.5119861754\,,\ E_4 = 0.1\,10^{-9}$$

y podemos considerar que salvo un error, $\tilde{\alpha} = y_4 = 0.5119861754$, pues $\tilde{g}(y_4) = 0.5119861753$ y $\tilde{g}(y_4) - y_4 = -0.1\,10^{-9}$, es decir, consideremos $\tilde{\alpha} = y_4 = 0.5119861753$

Ahora, calculemos la constante de Lipschitz de $g(x) = 0.1 + 0.6\,\cos(2x)$. Tenemos,

$$\begin{aligned} |g(x) - g(y)| &= |0.6\,\cos(2x) - 0.6\,\cos(2y)| \\ &= 0.6|\cos(2x) - \cos(2y)| \\ &= 0.6|2\mathrm{sen}(2\zeta)|\,|x - y| \end{aligned}$$

donde ζ está entre x e y, y $x, y \in [0.45, 0.473]$. Debemos maximizar la expresión $\mathrm{sen}(2\zeta)$ para $\zeta \in [0.45, 0.473]$. Se tiene que $\max\{|\mathrm{sen}(2\zeta)| : \zeta \in [0.45, 0.473]\} \approx 0.8110822713$, luego

$$|g(x) - g(y)| \leqslant 1.2 \times 0.8110822713\,|x - y|$$

es decir,

$$|g(x) - g(y)| \leqslant 0.9732987256\,|x - y|\,.$$

y tomamos $L = 0.973298726$. Tememos $\alpha = x_g = 0.461785307873$ y $\tilde{\alpha} = 0.5119861753$, luego

$$|\alpha - \tilde{\alpha}| = 0.050200867 \leqslant \frac{0.1}{1 - L} = \frac{0.1}{1 - 0.973298726} = 3.745139651\,.$$

Solución Teórica.

Tenemos

$$\begin{aligned} |\alpha - \tilde{\alpha}| &= |\tilde{g}(\tilde{\alpha}) - g(\alpha)| \\ &= |0.2 + 0.6\cos(2\tilde{\alpha}) - 0.1 - 0.6\cos(2\alpha)| \\ &= |0.1 + 0.6\cos(\tilde{\alpha}) - 0.6\cos(\alpha)| \\ &\leqslant 0.1 + |0.1 + 0.6\cos(2\tilde{\alpha}) - (0.1 + 0.6\cos(\alpha))| \\ &= 0.1 + |g(\tilde{\alpha}) - g(\alpha)| \\ &\leqslant 0.1 + L|\tilde{\alpha} - \alpha|\,. \end{aligned}$$

Es decir, hemos probado que $|\tilde{\alpha} - \alpha| \leqslant 0.1 + L|\tilde{\alpha} - \alpha|$, de donde, $|\tilde{\alpha} - \alpha|(1 - L) \leqslant 0.1$, y de aquí, $|\tilde{\alpha} - \alpha| \leqslant \frac{1}{1-L}$ como se pedía.

Problema 2.2 Sea $f(x) = (x-1)\ln(x)$.

(a) Determine la fórmula de Newton $x_{n+1} = g(x_n)$. Tomando como valor inicial $x_0 = e$, calcule las primeras dos iteraciones x_1 y x_2 del método.

(b) Demuestre que el método de Newton usado en la parte (a) tiene un orden de convergencia $p = 1$.

(c) Dé un método alternativo de iteración que garantice un orden de convergencia $p = 2$ y demuestre que efectivamente este es su orden de convergencia.

Solución.

(a) La frmula del método de Newton para encontrar la raíz de $f(x) = (x-1)\ln x = 0$ es:

$$\begin{aligned} x_{n+1} &= g(x_n) = x_n - \frac{(x_n - 1)\ln x_n}{\ln x_n + \frac{x_n - 1}{x_n}} = \frac{x_n(x_n \ln x_n + (x_n - 1) - (x_n - 1)\ln x_n)}{x_n \ln x_n + (x_n - 1)} \\ &= x_n \frac{x_n - 1 + \ln x_n}{x_n - 1 + x_n \ln x_n}. \end{aligned}$$

Tomando $x_0 = e$ y simplificando al máximo, obtenemos

$$x_1 = \frac{e^2}{2e-1} \quad \text{y} \quad x_2 = \frac{e^2}{2e-1} \cdot \frac{(e+1)^2 - 2 - \ln(2e-1)}{e^2(2 - \ln(2e-1)) + (e-1)^2}$$

(b) La raíz de $f(x) = 0$ es evidentemente $x = 1$ y es una raíz doble, es decir, $f(1) = 0$, $f'(1) = 0$, y $f''(1) \neq 0$. En efecto, es claro que $f(1) = 0$. Además, $f'(x) = \ln x + 1 - \frac{1}{x}$. Así, $f'(1) = 0$. Por último, $f''(x) = \frac{1}{x} + \frac{1}{x^2}$ y por lo tanto $f''(1) = 2 \neq 0$. Esto muestra que $x = 1$ es una raíz doble de $f(x) = 0$. Como se vi en clases, esto implica que la funcin de iteracin del método de Newton $g(x)$ satisface que

$$g'(1) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Luego, como $|g'(1)| < 1$, el método de Newton es convergente, partiendo de un punto x_0 suficientemente cercano a 1, pero su orden de convergencia es slo 1, pues $g'(1) \neq 0$.

(c) Como es sabido de lo visto en clases, la forma de "reparar" el orden de convergencia cuadrático del método de Newton es mediante el método de Newton modificado, el cual se obtiene multiplicando el cuociente $\frac{f(x)}{f'(x)}$ (en la frmula de Newton) por la multiplicidad de la raíz (que en este caso es 2), es decir,

$$x_{n+1} = \widetilde{g}(x_n) = x_n - 2\frac{(x_n - 1)\ln x_n}{\ln x_n + \frac{x_n - 1}{x_n}} = x_n \frac{x_n - 1 + (2 - x_n)\ln x_n}{x_n \ln x_n + x_n - 1}.$$

Con esto, por lo que se vi en clases, obtenemos que

$$\widetilde{g}'(1) = 1 - 2 \cdot \frac{1}{2} = 0,$$

lo cual implica que el método de Newton modificado converge cuadráticamente, partiendo de un punto x_0 suficientemente cercano a 1.

Problema 2.3 Considere la ecuación

$$\cos(x) = \tan(x),$$

de la cual se sabe que tiene una solución $\bar{x} = 0.6662394321$, con un error de 0.9×10^{-9}.

(a) Defina un método de punto fijo (diferente a Newton) y demuestre su convergencia (sin hacer iteraciones) en el intervalo que usted defina.

(b) Cuántas iteraciones necesitaría para obtener el error de 0.9×10^{-9} comenzando con $x_0 = 0.6$?

(c) Use el método de Newton con el punto inicial $x_0 = 0.6$ y obtenga x_1 y x_2 Cuál es la velocidad de convergencia del método de punto fijo propuesto por usted en la parte (a)?

Solución.

(a) Propuesta del método.

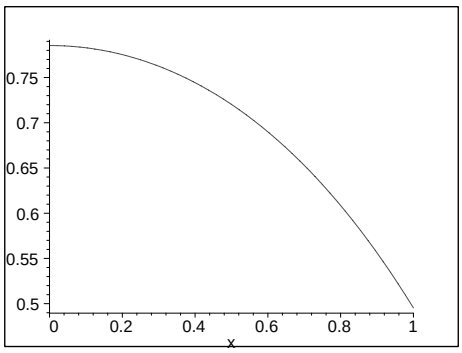
Despejando de la ecuación tenemos

$$x = \arctan(\cos(x)) = g(x)\,,$$

es decir,

$$x_{k+1} = g(x_k) = \arctan(\cos(x_k))\,.$$

Elegimos el intervalo $[0,1]$, tenemos que $g(x) = \arctan(\cos(x))$ es decreciente en ese intervalo, pues $g'(x) = \dfrac{d}{dx}\arctan(\cos(x)) = -\dfrac{\operatorname{sen}(x)}{1+\cos^2(x)}$ y como $g(0) = \arctan(\cos(0)) = \dfrac{\pi}{4}$ y $g(1) = \arctan(\cos(1)) \approx 0.4953672892$, vemos que $g([0,1]) \subset [0,1]$. Por lo tanto, g tiene un único punto fijo x_g en $[0,1]$.



Ahora, $g'(x) = \dfrac{d}{dx}\arctan(\cos(x)) = -\dfrac{\operatorname{sen}(x)}{1+\cos^2(x)}$. Luego,

$$|g(\bar{x})| = |g'(0.6662394321)| \approx 0.381964618 < 1$$

y el método iterativo propuesto es convergente.

(b) Para estimar el número de iteraciones podemos usar la fórmula para cotas del error

$$|x_n - \bar{x}| \leqslant \frac{\lambda^n}{1-\lambda}\,|x_0 - x_1|\,,$$

donde $0 \leqslant \lambda < 1$ es tal que $|g'(x)| \leqslant \lambda$ para todo $x \in]\,a, b\,[$. Ahora como $|g'(x)| = \frac{|\mathrm{sen}(x)|}{1+\cos^2(x)} = \frac{\mathrm{sen}(x)}{1+\cos^2(x)}$, para $x \in [0,1]$, es una función creciente, se sigue que

$$|g'(x)| \leqslant \frac{\mathrm{sen}(1)}{1+\cos^2(1)} \approx 0.5463024898 \leqslant 0.55$$

y podemos usar $\lambda = 0.55$.

Tenemos que $x_0 = 0.6$, luego $x_1 = 0.6899997083$ y $|x_0 - x_1| = 0.0899997083$. Luego

$$\begin{aligned} \frac{0.55^n}{1-0.55} \times 0.0899997083 &\leqslant 0.9 \times 10^{-9} \\ 0.55^n &\leqslant 4.50001458 \times 10^{-9} \\ n\ln(0.55) &\leqslant \ln(4.50001458 \times 10^{-9}) \\ n(-0.5978370008) &\leqslant -19.2191852 \end{aligned}$$

de donde $n \geqslant 32.14786836$, por lo tanto basta tomar $n = 33$ iteraciones para tener lo pedido.

(c) Tenemos que $N_f(x) = x - \frac{f(x)}{f'(x)}$, donde en este caso, $f(x) = \tan(x) - \cos(x)$. Como $f'(x) = 1 + \tan^2(x) + \mathrm{sen}(x)$, se tiene que

$$N_f(x) = x - \frac{\tan(x) - \cos(x)}{1 + \tan^2(x) + \mathrm{sen}(x)}$$

luego, calculando obtenemos $x_1 = 0.6694641628$ y $x_2 = 0.6660244822$. Ahora, como $f'(\bar{x}) \approx 2.236067976 \neq 0$ el método de Newton tiene convergencia cuadrática, es decir, orden $p = 2$.

Para analizar la convergencia del método propuesto en (a) tenemos que buscar la primera derivada no nula en el punto fijo, y por los cálculos ya realizados se tiene que $g'(\bar{x}) \neq 0$, por tanto el método tiene orden de convergencia $p = 1$. Por lo tanto, el método de Newton converge mucho más rápido, en este caso.

Problema 2.4 Se desea resolver la ecuación $x^3 - \ln(1+2x) = 0$.

1. Compruebe que esta ecuación tiene exactamente una solución en el intervalo $[1,2]$.

2. Proponga un método iterativo de punto fijo (diferente de Newton) que converja a la solución de la ecuación en el intervalo $[1,2]$. Justifique.

3. Partiendo de $x_0 = 1$, determine el número mínimo de iteraciones que se necesitarían para alcanzar una precisión de 10^{-4}, utilizando el método iterativo propuesto anteriormente.

Solución.

1. Tenemos la ecuación

$$f(x) = x^3 - \ln(1+2x) = 0\,.$$

Es claro que el dominio de f es $\{x \in \mathbb{R} : x > -1/2\}$, en el cual f es continua y diferenciable.

Ahora,

$$f'(x) = 3x^2 - \frac{2}{1+2x} = \frac{3x^2(1+2x)-2}{2x+1} = \frac{6x^3+3x^2-2}{2x+1}.$$

El denominador de f es positivo en el intervalo [1,2] y su numerador $h(x) = 6x^3+3x^2-2$ también es positivo en dicho intervalo, pues $h'(x) = 18x^2+6x = 0$ si y sólo si $x = 0$ o $18x+6 = 0$, es decir, $x = -1/3$. Con esto, es claro que $h'(x) > 0$ en el intervalo $[1,2]$, luego h es estrictamente creciente en $[1,2]$. Además, como $h(1) = 6+3-2 = 7$, se sigue que $h(x) > 0$ en todo $[1,2]$. Por lo tanto $f'(x) > 0$ en $[1,2]$, en consecuencia si tiene un cero en dicho intervalo, este sería único. Ahora, como $f(1) = 1-\ln(3) \approx -0.09861\ldots < 0$ y $f(2) = 8-\ln(5) \approx 6.39056\ldots > 0$, por el Teorema del Valor Intermedio, f tiene un cero en $[1,2]$, y como vimos, este es único.

2. Despejando x desde la igualdad

$$x^3 - \ln(1+2x) = 0$$

tenemos

$$x = \sqrt[3]{\ln(1+2x)}\,.$$

Proponemos entonces el método iterativo de punto fijo

$$x_{n+1} = g(x_n) = \sqrt[3]{\ln(1+2x_n)}\,.$$

Tenemos

$$g'(x) = \frac{2}{3}\frac{1}{(\ln(1+2x))^{2/3}}\cdot\frac{1}{2x+1} > 0$$

en $[1,2]$. Por lo tanto, g es creciente en $[1,2]$.

Ahora, para ver que $g([1,2]) \subseteq [1,2]$, basta ver que $1 \leqslant g(1) \leqslant g(2) \leqslant 2$. Tenemos

$$1 < \underbrace{1.03184584\ldots}_{g(1)} \leqslant \underbrace{1.171902307\ldots}_{g(2)} \leqslant 2$$

de donde $g([1,2]) \subseteq [1,2]$ y como g es estrictamente creciente, en ese intervalo, g tiene un único punto fijo $x_g \in g([1,2]) \subset [1,2]$.

Para ver la convergencia del método propuesto, debemos probar que $|g'(x)| \leqslant \lambda < 1$ para todo $x \in [1,2]$.

Ahora, $g'(x)$ es decreciente en $[1,2]$ y $g'(1) = 0.208717132 > g'(x) > g'(2) = 0.0970858457$. Por lo tanto $\lambda = \max\{|g'(x)| : x \in [1,2]\} = 0.2087170132$ y es claro que $\lambda < 1$.

Por lo tanto el método propuesto es convergente al único punto fijo de g en $[1,2]$.

3. Tomando $x_0 = 1$, tenemos que $x_1 = g(1) = \sqrt[3]{\ln(3)} = 1.03184584\ldots$.Úsando la fórmula

$$|x_g - x_n| \leqslant \frac{\lambda^n}{1-\lambda}\,|x_1 - x_0|,$$

si queremos precisión de 10^{-4}, imponemos entonces que

$$\frac{\lambda^n}{1-\lambda}\,|x_1 - x_0| \leqslant 10^{-4}\,.$$

Como $|x_1 - x_0| = |0.0318458398\ldots| < 0.032$ (para evitar el cálculo con tantos decimales) y $\lambda = 0.2087170132$, se tiene $1 - \lambda = 0.7912829868\ldots > 0.79$ (misma razón anterior)
Así $\dfrac{1}{1-\lambda} < \dfrac{1}{0.79}$, y como $\lambda < 0.21$ nos queda

$$\frac{\lambda^n}{1-\lambda} \leqslant \frac{(0.21)^n}{0.79}\,.$$

Luego

$$\frac{\lambda^n}{1-\lambda}\,|x_1 - x_0| \leqslant \frac{(0.21)^n}{0.79} \times 0.032 = (0.21)^n \times 0.0405063\ldots \leqslant (0.21)^n \times 0.041$$

y hacemos $0.041 \times (0.21)^n < 10^{-4}$, de donde $(0.21)^n < \dfrac{10^{-4}}{0.041}$, y entonces
$n\,\ln(0.21) \leqslant \ln\left(\dfrac{10^{-4}}{0.041}\right)$ y como $\ln(0.21) < 0$, se sigue que

$$n \;\geqslant\; \frac{\ln\left(\dfrac{10^{-4}}{0.041}\right)}{\ln(0.21)} \approx 3.8549104\ldots$$

Por lo tanto, con al menos 4 iteraciones tenemos garantizado lo pedido.

Problema 2.5 Considere el siguiente sistema de ecuaciones no lineales

$$\begin{cases} x_1^2 + x_2 - 37 &= 0 \\ x_1 - x_2^2 - 5 &= 0 \\ x_1 + x_2 + x_3 - 3 &= 0 \end{cases}$$

(a) Utilizando el método de Newton para sistemas no lineales y tomando como punto inicial $\mathrm{x}^0 = (0,0,0)^T$, obtenga el punto x^1.

(b) Cuando se utiliza el método de Newton para sistemas no lineales, una alternativa para no calcular la inversa de una matriz en cada iteración es resolver un sistema auxiliar de ecuaciones linwales en cada iteración, lo cual es desde el punto de vista numérico siempre resulta más conveniente.

Describa el sistema de ecuaciones lineales que debería resolver para obtener el punto x^1 de la parte (a) y calcule el número de condicionamiento de la matriz del sistema con respecto al redio espectral puede inferir algo sobre la estabilidad de las soluciones del sistema?

Solución. Tenemos el sistema de ecuaciones no lineales siguiente

$$\left\{\begin{array}{rcl} x_1^2+x_2-37 &=& 0 \\ x_1-x_2^2-5 &=& 0 \\ x_1+x_2+x_3-3 &=& 0 \end{array}\right.$$

a) Método de Newton. Llamemos

$$\begin{aligned} f_1(x_1,x_2,x_3) &= x_1^2+x_2-37 \\ f_2(x_1,x_2,x_3) &= x_1-x_2^2-5 \\ f_3(x_1,x_2,x_3) &= x_1+x_2-x_3-3 \end{aligned}$$

y $F(x_1,x_2,x_3)=(f_1(x_1,x_2,x_3),f_2(x_1,x_2,x_3),f_3(x_1,x_2,x_3))$.

El método de Newton es dado por

$$\begin{pmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{pmatrix} = \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{pmatrix} - \left(JF(x_1^{(k)},x_2^{(k)},x_3^{(k)})\right)^{-1} \begin{pmatrix} f_1(x_1^{(k)},x_2^{(k)},x_3^{(k)}) \\ f_2(x_1^{(k)},x_2^{(k)},x_3^{(k)}) \\ f_3(x_1^{(k)},x_2^{(k)},x_3^{(k)}) \end{pmatrix}$$

Ahora,

$$JF = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} \end{pmatrix}.$$

Derivando, tenemos

$$\begin{aligned} \frac{\partial F}{\partial x_1} &= \left(\frac{\partial f_1}{\partial x_1},\frac{\partial f_2}{\partial x_1},\frac{\partial f_3}{\partial x_1}\right)=(2x_1,1,1) \\ \frac{\partial F}{\partial x_2} &= \left(\frac{\partial f_1}{\partial x_2},\frac{\partial f_2}{\partial x_2},\frac{\partial f_3}{\partial x_2}\right)=(1,-2x_1,1) \\ \frac{\partial F}{\partial x_3} &= \left(\frac{\partial f_1}{\partial x_3},\frac{\partial f_2}{\partial x_3},\frac{\partial f_3}{\partial x_3}\right)=(0,0,1) \end{aligned}$$

Luego,

$$JF(x_1,x_2,x_3) = \begin{pmatrix} 2x_1 & 1 & 0 \\ 1 & -2x_2 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Evaluando en $\mathrm{x}^{(0)}=(0,0,0)^T$ nos queda la matriz

$$JF(0,0,0) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} .$$

Para calcular $(JF(0,0,0))^{-1}$, lo podemos hacer en forma simple, escribiendo

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

de donde $a_{21} = 1$, $a_{22} = 0$, $a_{23} = 0$, $a_{11} = 0$, $a_{13} = 0$, $a_{11} + a_{21} + a_{31} = 0$, de donde, $a_{31} = -1$; $a_{12} + a_{22} + a_{32} = 0$, de donde, $a_{32} = -1$; $a_{13} + a_{23} + a_{33} = 1$, de donde, $a_{33} = 1$. Por lo tanto,

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & -1 & 1 \end{pmatrix}$$

Tenemos, $F(0,0,0) = (-37,-5,-5)$. Por lo tanto,

$$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} -37 \\ -5 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ 37 \\ -39 \end{pmatrix}$$

b) Para obtener la fórmula pedida, recordemos lo hecho en clases para sistemas de ecuaciones no lineales, en este caso de 3×3. Sean (x_1, x_2, x_3) una solución aproximada del sistema de ecuaciones no lineales, y sea $(x_1 + h_1, x_2 + h_2, x_3 + h_3)$ la solución exacta. Aplicando Taylor, obtenemos

$$\begin{aligned} 0 &= f_1(x_1 + h_1, x_2 + h_2, x_3 + h_3) \approx f_1(x_1, x_2, x_3) + h_1 \frac{\partial f_1}{\partial x_1} + h_2 \frac{\partial f_1}{\partial x_2} + h_3 \frac{\partial f_1}{\partial x_3} \\ 0 &= f_2(x_1 + h_1, x_2 + h_2, x_3 + h_3) \approx f_2(x_1, x_2, x_3) + h_1 \frac{\partial f_2}{\partial x_1} + h_2 \frac{\partial f_2}{\partial x_2} + h_3 \frac{\partial f_2}{\partial x_3} \\ 0 &= f_3(x_1 + h_1, x_2 + h_2, x_3 + h_3) \approx f_3(x_1, x_2, x_3) + h_1 \frac{\partial f_3}{\partial x_1} + h_2 \frac{\partial f_3}{\partial x_2} + h_3 \frac{\partial f_3}{\partial x_3} \end{aligned}$$

donde las derivadas parciales están evaluadas en (x_1, x_2, x_3). De lo anterior, tenemos que

$$JF(x_1, x_2, x_3) \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = -F(x_1, x_2, x_3) .$$

Ahora, usando los datos, tenemos

$$F(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}) = \begin{pmatrix} -37 \\ -5 \\ -3 \end{pmatrix} ,$$

$$JF(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}) = JF(0,0,0) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Por lo tanto el sistema de ecuaciones lineales es dado por

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = - \begin{pmatrix} -37 \\ -5 \\ -3 \end{pmatrix}$$

De aquí, $h_2 = 37$, $h_1 = 5$ y $h_1 + h_2 + h_3 = 3$, de donde $h_3 = -39$.

Ahora, como

$$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{pmatrix} = \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix},$$

se tiene que

$$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{pmatrix} = \begin{pmatrix} 5 \\ 37 \\ -39 \end{pmatrix}$$

el cual coincide, obviamente, con el cálculo hecho en la parte (a).

El número de condición con respecto al radio espectral, significa que usando la norma subordinada $||\cdot||_2$, se tiene que

$$k(A) = \sqrt{\frac{\max\{|\lambda| \ : \ \lambda \text{ valor propio de } \ AA^T\}}{\min\{|\lambda| \ : \ \lambda \text{ valor propio de } \ AA^T\}}}.$$

LLamando

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

Se tiene que

$$A^T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Luego

$$AA^T = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

$$AA^T - \lambda I = \begin{pmatrix} 1-\lambda & 0 & 1 \\ 0 & 1-\lambda & 1 \\ 1 & 1 & 3-\lambda \end{pmatrix}$$

y

$$\det(AA^T - \lambda) = (1-\lambda)^2(3-\lambda) - (1-\lambda) - (1-\lambda) = (1-\lambda)((1-\lambda)(3-\lambda) - 2)$$

luego, $\det(AA^t - \lambda I) = 0$ si $\lambda_1 = 1$ o $\lambda^2 - 4\lambda + 3 = 0$, de donde $\lambda_2 = 3$ y $\lambda_3 = 1$. Luego, $\max\{|\lambda| : \lambda \text{ valor propio de } AA^T\} = 3$ y $\min\{|\lambda| : \lambda \text{ valor propio de } AA^T\} = 1$. Por lo tanto, $k(A) = \sqrt{\frac{3}{1}} = \sqrt{3}$. Como el número de condición de A es relativamente pequeño, esl sistema está bien condicionado y es estable.

Problema 2.6 Resuelva usando el método de Newton para varias variables, el siguiente sistema no lineal:

$$\begin{cases} 3x_1^2 - x_2^2 = 0 \\ 3x_1x_2^2 - x_1^3 = 1 \end{cases}$$

Para esto, tome como punto inicial $x^{(0)} = (1,1)^{\mathrm{T}}$ y realice dos iteraciones, es decir, calcule $x^{(1)}$ y $x^{(2)}$.

Solución. Tenemos el sistema de ecuaciones no lineales

$$\begin{cases} f(x,y) &= 3x^2 - y^2 = 0 \\ g(x,y) &= 3xy^2 - x^3 - 1 = 0\,. \end{cases}$$

Sea $F(x,y) = (f(x,y), g(x,y))$. El método de Newton aplicado a F es dado por

$$N_F(x,y) = (x,y) - (JF(x,y))^{-1}F(x,y)\,.$$

Ahora,

$$JF(x,y) = \begin{pmatrix} \dfrac{\partial f}{\partial x}(x,y) & \dfrac{\partial f}{\partial y}(x,y) \\ \dfrac{\partial g}{\partial x}(x,y) & \dfrac{\partial g}{\partial y}(x,y) \end{pmatrix} = \begin{pmatrix} 6x & -2y \\ 3y^2 - 3x^2 & 6xy \end{pmatrix}.$$

Luego, $\det JF(x,y) = 36x^2y + 2y(3y^2 - 3x^2) = 36x^2y + 6y^3 - 6x^2y = 30x^2y + 6y^3 = 6y(5x^2 + y^2)$. Por lo tanto $\det JF(x,y) \neq 0$ si $y \neq 0$ y $(x,y) \neq (0,0)$.

Como $\mathbf{x}^{(0)} = (1,1)$

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - (JF(x_0,y_0))^{-1}F(x_0,y_0)$$

y

$$JF(x_0,y_0) = JF(1,1) = \begin{pmatrix} 6 & -2 \\ 0 & 6 \end{pmatrix}$$

se tiene que $\det JF(x_0, y_0) = 36$ y

$$\begin{aligned}(JF(1,1))^{-1} &= \frac{1}{36}\begin{pmatrix} 6 & 2 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} \frac{1}{6} & \frac{1}{18} \\ 0 & \frac{1}{6} \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ &= \left(\frac{2}{6} + \frac{1}{18}, \frac{1}{6}\right) \\ &= \left(\frac{7}{18}, \frac{1}{6}\right)\end{aligned}$$

y

$$\mathbf{x}^{(1)} = (x_1, y_1) = (1,1) - \left(\frac{7}{18}, \frac{1}{6}\right) = \left(\frac{11}{18}, \frac{5}{6}\right)$$

Para $\mathbf{x}^{(2)}$, tenemos

$$JF(x_1, y_1) = JF\left(\frac{11}{18}, \frac{5}{6}\right) = \begin{pmatrix} \frac{11}{3} & -\frac{5}{3} \\ \frac{26}{27} & \frac{55}{18} \end{pmatrix}$$

luego $\det JF(x_1, y_1) = \dfrac{2075}{162}$.

Ahora,

$$(JF(x_1, y_1))^{-1} = \frac{162}{2075}\begin{pmatrix} \frac{55}{18} & \frac{5}{3} \\ -\frac{26}{27} & \frac{11}{3} \end{pmatrix} = \begin{pmatrix} \frac{99}{415} & \frac{54}{415} \\ -\frac{156}{2075} & \frac{594}{2075} \end{pmatrix}$$

Así

$$\begin{aligned}(JF(x_1, y_1))^{-1}F(x_1, y_1) &= \begin{pmatrix} \frac{99}{415} & \frac{54}{415} \\ -\frac{156}{2075} & \frac{594}{2075} \end{pmatrix}\begin{pmatrix} \frac{23}{54} \\ \frac{131}{2916} \end{pmatrix} \\ &= \left(\frac{1204}{11205}, -\frac{2147}{112050}\right) \\ &= (0.1074520303\,, -0.01916108880)\,.\end{aligned}$$

Luego,

$$\begin{aligned}\mathbf{x}^{(2)} = (x_2, y_2) &= \left(\frac{11}{18}, \frac{5}{5}\right) - \left(\frac{1204}{11205}, -\frac{2147}{112050}\right) \\ &= \left(\frac{11287}{22410}, \frac{47761}{56025}\right) \\ &= (0.5036590808\,,\, 0.8524944221)\end{aligned}$$

Una forma alternativa y más simple es resolver en cada paso un sistema de ecuaciones lineales, para ello recordemos que el método de Newton es dado

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix} + \begin{pmatrix} h_n \\ k_n \end{pmatrix},$$

donde $\begin{pmatrix} h_n \\ k_n \end{pmatrix}$ es la solución del sistema de ecuaciones lineales

$$JF(x_n, y_n) \begin{pmatrix} h_n \\ k_n \end{pmatrix} = -F(x_n, y_n)$$

Tenemos, $F(x, y) = (3x^2 - y^2, 3xy^2 - x^3 - 1)$

$$JF(x, y) = \begin{pmatrix} 6x & -2y \\ 3y^2 - 3x^2 & 6xy \end{pmatrix}$$

Como $(x_0, y_0) = (1, 1)$ nos queda $JF(x_0, y_0) = \begin{pmatrix} 6 & -2 \\ 0 & 6 \end{pmatrix}$ y $F(x_0, y_0) = (2, 1)$, obtenemos así el sistema de ecuaciones lineales

$$\begin{pmatrix} 6 & -2 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} h_0 \\ k_0 \end{pmatrix} = -\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

de donde $6k_0 = -1$, es decir, $k_0 = -\dfrac{1}{6}$; la otra ecuación es $6h_0 - 2k_0 = -2$, y de aquí $h_0 = -\dfrac{7}{18}$. Luego

$$\begin{pmatrix} x_1 \\ \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 \\ \\ 1 \end{pmatrix} + \begin{pmatrix} \dfrac{-7}{18} \\ \\ \dfrac{-1}{6} \end{pmatrix} = \begin{pmatrix} \dfrac{11}{18} \\ \\ \dfrac{5}{6} \end{pmatrix}$$

Para $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ debemos resolver el sistema

$$JF(x_1, y_1) \begin{pmatrix} h_1 \\ k_1 \end{pmatrix} = -F(x_1, y_1)$$

Este sistema es

$$\begin{pmatrix} \frac{11}{3} & -\frac{5}{3} \\ \frac{26}{27} & \frac{55}{18} \end{pmatrix} \begin{pmatrix} h_1 \\ k_1 \end{pmatrix} = - \begin{pmatrix} \frac{23}{54} \\ \frac{131}{2916} \end{pmatrix}$$

cuya solución es

$$(h_1, k_1) = \left(-\frac{1024}{11205}, \frac{2147}{112050} \right) = (-0.10745203030\,, 0.01916108880)\,.$$

Luego,

$$(x_2, y_2) = (x_1, y_1) + (h_1, k_1) = \left(\frac{11}{18} - \frac{1024}{11205}, \frac{5}{6} + \frac{2147}{112050} \right) = (0.5036590808\,, 0.8524944221)$$

2.11 Ejercicios

Problema 2.1 Use el manejo algebraico para demostrar que las siguientes funciones tienen un punto fijo en p exactamente cuando $f(p) = 0$, donde $f(x) = x^4 + 2x^2 - x - 3$.

1. $g_1(x) = \left(3 + x - 2x^2\right)^{1/4}$
2. $g_2(x) = \left(\frac{x+3-x^4}{2}\right)^{1/2}$
3. $g_3(x) = \left(\frac{x+3}{x^2+2}\right)^{1/2}$
4. $g_4(x) = \frac{3x^4+2x^2+3}{4x^3+x-1}$

Problema 2.2

1. Efectúe cuatro iteraciones, si es posible hacerlo, en las funciones definidas en el problema anterior, considerando $p_0 = 1$ y $p_{n+1} = g(p_n)\,, \quad n = 0,\ 1,\ 2,\ 3$.
2. ¿Cuál función, a su juicio, dará la mejor aproximación a la solución?

Problema 2.3 Se proponen los siguientes métodos para calcular $21^{1/3}$. Clasifique por orden, basándose para ello la rapidez de convergencia y suponiendo que $p_0 = 1$ en cada caso.

1. $p_n = \frac{20p_{n-1} + \frac{21}{p_{n-1}^2}}{21}$
2. $p_n = p_{n-1} - \frac{p_{n-1}^3 - 21}{3p_{n-1}^2}$
3. $p_n = p_{n-1} - \frac{p_{n-1}^4 - 21p_{n-1}}{p_{n-1}^2 - 21}$
4. $p_n = \left(\frac{21}{p_{n-1}}\right)^{1/2}$

Problema 2.4 Aplique el método de bisección para determinar c_3, para $f(x) = \sqrt{x} - \cos(x)$ en $[0, 1]$.

Problema 2.5 Sea $f(x) = 3(x+1)\left(x-\frac{1}{2}\right)(x-1)$. Aplique el método de bisección para determinar c_3 en los siguientes intervalos

1. $\left[-2, \frac{3}{2}\right]$
2. $\left[\frac{5}{4}, \frac{5}{2}\right]$

Problema 2.6 Aplique en los siguientes intervalos el método de bisección para determinar las aproxiamciones a las soluciones de $x^3 - 7x^2 + 14x - 6 = 0$ con una exactitud de 10^{-2}.

1. $[0, 1]$
2. $\left[1, \frac{16}{5}\right]$
3. $\left[\frac{16}{5}, 4\right]$

Problema 2.7 Aplique en los siguientes intervalos el método de bisección para determinar las aproximaciones a las soluciones de $x^4 - 2x^3 - 4x^2 + 4x + 4 = 0$ con una exactitud de 10^{-2}.

1. $[-2, -1]$
2. $[0, 2]$
3. $[2, 3]$
4. $[-1, 0]$

Problema 2.8 Aplique el método de bisección para determinar una aproximación a la solución de $\tan(x) = x$ con una exactitud de 10^{-3} en el intervalo $\left[4, \frac{9}{2}\right]$.

Problema 2.9 Aplique el método de bisección para determinar una aproximación a la solución de $2 + \cos(e^x - 2) - e^x = 0$ con una exactitud de 10^{-3} en el intervalo $\left[\frac{1}{2}, \frac{3}{2}\right]$

Problema 2.10 En cada caso aplique el método de bisección para determinar una aproximación a la solución con una exactitud de 10^{-5}.

1. $x - 2^{-x} = 0$ para $x \in [0, 1]$.
2. $e^{-x} - x^2 + 3x - 2 = 0$ para $x \in [0, 1]$.
3. $2x\cos(2x) - (x+1)^2 = 0$ para $x \in [-3, -2]$ y para $x \in [-1, 0]$.
4. $x\cos(x) - 2x^2 + 3x - 1 = 0$ para $x \in \left[\frac{1}{5}, \frac{3}{10}\right]$ y para $x \in \left[\frac{6}{5}, \frac{13}{10}\right]$.

Problema 2.11 Determine una aproximación de $\sqrt{3}$ con una exactitud de 10^{-4} usando el método de bisección.

Problema 2.12 Determine una cota del numero de iteraciones que se requiere para alcanzar una aproximación con una exactitud de 10^{-3} de la solución de $x^3 + x - 4 = 0$ en el intervalo $[1, 4]$.

Problema 2.13 Determine una cota del numero de iteraciones que se requiere para alcanzar una aproximación con una exactitud de 10^{-3} de la solución de $x^3 - x - 1 = 0$ en el intervalo $[1, 2]$.

Problema 2.14 Sea $f(x) = (x-1)^{10}$, $p = 1$ y $p_n = 1 + \frac{1}{n}$. Demuestre que $|f(p_n)| < 10^{-3}$ para todo $n > 1$, mientras que $|p - p_n| < 10^{-3}$ sólo si $n > 1000$.

Problema 2.15 Sea $(p_n)_{n\in\mathbb{N}}$ la sucesión dada por $p_n = \sum_{k=1}^{n} \frac{1}{k}$. Demuestre que $(p_n)_{n\in\mathbb{N}}$ diverge aún cuando $\lim_{n\to\infty} (p_n - p_{n-1}) = 0$.

Problema 2.16 Los cuatro siguientes métodos tienen por objeto calcular $7^{1/5}$. Clasifique por orden, basándose para ello la rapidez de convergencia y suponiendo que $p_0 = 1$ en cada caso.

1. $p_n = \left(1 + \frac{7 - p_{n-1}^3}{p_{n-1}^2}\right)^{1/2}$
2. $p_n = p_{n-1} - \frac{p_{n-1}^5 - 7}{p_{n-1}^2}$
3. $p_n = p_{n-1} - \frac{p_{n-1}^5 - 7}{5p_{n-1}^4}$
4. $p_n = p_{n-1} - \frac{p_{n-1}^5 - 7}{12}$
5. Aplique el método de iteración de punto fijo para determinar una solución con una exactitud de 10^{-2} para $x^4 - 3x^2 - 3 = 0$ en $[1, 2]$. Utilice $p_0 = 1$.

Problema 2.17 La ecuación $x^2 - x - 1 = 0$ tiene una raíz positiva $\alpha \approx 1.61803398\cdots$. Proponga un método iterativo convergente, no Newton, para aproximar dicha raíz.

Indicación. Tenemos $x^2 - x - = 0 \Leftrightarrow 2x^2 - x = x^2 + 1 \Leftrightarrow x(2x-1) = x^2 + 1 \Leftrightarrow x = \dfrac{x^2+1}{2x-1}$.

Proponemos $g(x) = \dfrac{x^2+1}{2x-1}$ como método iterativo.

Afirmación 1 $g : \left[\frac{3}{2}, 2\right] \to \left[\frac{3}{2}, 2\right]$

Afirmación 2 g es una contracción.

Problema 2.18 Considere la ecuación $e^x - x^2 + 3x - 2 = 0$.

1. Proponga un método de punto fijo, explicitando un intervalo $[a, b]$ contenido en $[0, 1]$ que contenga dicha raíz, para resolver esta ecuación. Justifique porqué el método por Ud. propuesto es convergente.
2. Suponga que realizamos iteraciones con el método propuesto en el item anterior hasta obtener un error absoluto no mayor que 10^{-5} respecto de la raíz exacta. Sin utilizar calculadora, encuentre este número de iteraciones, para ellos use el intervalo $[a, b]$ explicitado.

Problema 2.19 Sean $f(x) = x^2 - 6$ y $x_0 = 1$. Aplique el método de Newton para determinar x_4.

Problema 2.20 Sean $f(x) = -x^3 - \cos(x)$ y $x_0 = -1$. Aplique el método de Newton para determinar x_4 ¿Podríamos utilizar $x_0 = 0$?

Problema 2.21 Considere la ecuación no lineal

$$\cos\left(\frac{x^2+5}{x^4+1}\right) = 0$$

se sabe que esta ecuación tiene una única raíz en el intervalos $[0,1]$. Use los métodos de Newton y secante para aproximar la raíz. Para Newton use $x_0 = 0$ y para secante use $x_0 = 0$ y $x_1 = 1$.

Problema 2.22 Sean $f(x) = x^2 - 6$, $x_0 = 3$ y $x_1 = 2$ determine x_3.

1. Aplique el método de la secante.
2. Aplique el método de Regula Falsi.
3. ¿Que método da una mejor aproximación de $\sqrt{6}$?

Problema 2.23 Aplique el método de Newton para obtener soluciones con una exactitud de 10^{-5} para los siguiente problemas.

1. $x^3 - 2x^2 - 5 = 0$, $x \in [1,4]$.
2. $x - \cos x = 0$, $x \in \left[0, \frac{\pi}{2}\right]$.
3. $x^3 + 3x^2 - 1 = 0$, $x \in [-3,-2]$.
4. $x - 0.8 - 0.2\text{sen}(x) = 0$, $x \in \left[0, \frac{\pi}{2}\right]$.
5. $e^x + 2^{-x} + 2\cos(x) - 6 = 0$, $x \in [1,2]$.
6. $\ln(x-1) + \cos(x-1) = 0$, $x \in [1.3,2]$.
7. $2x\cos(x) - (x-2)^2 = 0$, $x \in [2,3]$.
8. $\text{sen}(x) - e^{-x} = 0$, $x \in [0,1]$.

Problema 2.24 Utilice el método de Newton para aproximar con exactitud de 10^{-5} el valor de x en la gráfica de $y = x^2$ más cercano del punto $(1,0)$.

Problema 2.25 Utilice el método de Newton para aproximar con exactitud de 10^{-5} el valor de x en la gráfica de $y = \frac{1}{x}$ más cercano del punto $(2,1)$.

Problema 2.26 Con una exactitud de 10^{-5} y utilizando el método de Newton resuelva la ecuación $\frac{1}{2} + \frac{1}{4}x^2 - x\text{sen}(x) - \frac{1}{2}\cos(2x) = 0$, con condición inicial $x_0 = \frac{\pi}{2}$. Explique por que el resultado parece poco usual para el método de Newton. También resuelva la ecuación con $x_0 = 5\pi$ y $x_0 = 10\pi$.

Problema 2.27 Considere la función $f : \mathbb{R} \longrightarrow \mathbb{R}$ definida por $f(x) = 3x^5 - 10x^3 + 23x$

a) Pruebe que $f(x) = 0$ tiene una solución en el intervalo $[-2,2]$.

b) Usando el método de Newton con la condición inicial $x_0 = 1.07$ y el criterio de parada $|f(x)| \leqslant 10^{-6}$, encuentre una aproximación a una raíz de $f(x)$.

c) Usando el método de Newton con la condición inicial $x_0 = 1.06$ realice iteraciones Qué ocurre en este caso? Explique.

d) Demuestre la convergencia de los iterados del método de Newton cerca de la raíz que encontró b).

e) Considere el siguiente método iterativo

$$Sf(x) = x - \frac{f(x)f'(x)}{(f(x))^2 - f(x)f''(x))}\,.$$

Considere las dos condiciones iniciales $x_0 = 1.07$ y $x_0 = 1.06$ y realice las iteraciones en ambos caso Qué ocurre?

Problema 2.28 Se sabe que la ecuación $2x^4 + 24x^3 + 61x^2 - 16x + 1 = 0$ tiene dos raices cerca de 0.

1. Usando el método de Newton encuentre estas dos raices, con un error absoluto de 10^{-6}.
2. Considere el método iterativo siguiente

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

encuentre las raíces, con un error absoluto de 10^{-6}.

3. Analizando el error absoluto, compare la rapidez de convergencia de ambos métodos.
4. Pruebe que este nuevo método iterativo tiene orden de convergencia 3 alrededor de cada raíz simple de una ecuación $f(x) = 0$, donde $f(x)$ es al menos 3 veces derivable, con $f'''(x)$ continua.

Problema 2.29 Considere la ecuación $e^x - 4x^2 = 0$.

(a) Usando el método de regula falsi encuentre una raíz positiva de la ecuación.

(b) Proponga un método de punto fijo, no Newton, y demuestre, sin iterar, que converge a la raíz encontrada en (a) de la ecuación.

(c) Usando el método de punto fijo propuesto en (b), encuentre la solución buscada con una precisión de $\varepsilon = 10^{-3}$, considerando como criterio de parada $|f(x_n)| \leqslant \varepsilon$.

Problema 2.30 Considere la ecuación $x^3 - x^2 - x - 1 = 0$, la cual posee una solución $\alpha \in [1, 2]$. Se propone el método iterativo $x_{n+1} = g(x_n)$, donde $g(x) = 1 + \dfrac{1}{x} + \dfrac{1}{x^2}$, para resolver la ecuación.

1. Verifique las condiciones para que el método iterativo propuesto sea convergente.

2. Si elige $x_0 \in [a,b] \subset [1,2]$ arbitrario para comenzar las iteraciones, estime el número de iteraciones que debe realizar para obtener x_k que satisface la condición $|x_{k+1} - x_k| \leqslant 5 \cdot 10^{-5}$.

Problema 2.31 Considere al ecuación $-x^3 + 2x^2 + 10x - 20 = 0$.

(a) Proponga un método iterativo de punto fijo, no Newton, el cual sea convergente a la solución $\alpha = -\sqrt{10}$. La demostración de la convergencia debe hacerse sin iteraciones.

(b) Usando el método que propuso en a), encuentre una solución aproximada a α con un error de no más de 10^{-2}

Problema 2.32 Considere la ecuación $x^3 - x^2 - x - 1 = 0$, la cual posee una solución $\alpha \in [1,2]$. Se propone el método iterativo $x_{n+1} = g(x_n)$, donde $g(x) = 1 + \dfrac{1}{x} + \dfrac{1}{x^2}$, para resolver la ecuación.

1. Verifique las condiciones para que el método iterativo propuesto sea convergente.
2. Si elige $x_0 \in [a,b] \subset [1,2]$ arbitrario para comenzar las iteraciones, estime el número de iteraciones que debe realizar para obtener x_k que satisface la condición $|x_{k+1} - x_k| \leqslant 5 \cdot 10^{-5}$.

Problema 2.33 Utilice el método de Newton para encontrar las soluciones de los siguientes problemas con una exactitud de 10^{-5}, usando uno de los criterios de paradas especificados arriba.

1. $x^2 - 2xe^{-x} + e^{-2x} = 0,\ x \in [0,1]$.
2. $\cos\left(x + \sqrt{2}\right) + x\left(\dfrac{x}{2} + \sqrt{2}\right) = 0,\ x \in [-2,-1]$.
3. $x^3 - 3x^2 2^{-x} + 2x 4^{-x} - 8^{-x} = 0,\ x \in [0,1]$.
4. $e^{6x} + 3\left(\ln 2\right)^2 e^{2x} - e^{4x} \ln 8 - \left(\ln 2\right)^3,\ x \in [-1,0]$.

Problema 2.34 Repita el ejercicio anterior, aplicando el método de Newton modificado ¿Mejora la rapidez o la exactitud en comparación con el ejercicio 1.?

Problema 2.35 Demuestre que las siguientes sucesiones convergen linealmente a $p = 0$ ¿Qué tan grande debe ser n para que $|p_n - p| \leqslant 5 \times 10^{-2}$?

a.) $p_n = \frac{1}{n},\ n \geqslant 1$, b.) $p_n = \frac{1}{n^2},\ n \geqslant 1$

Problema 2.36 Demuestre que la sucesión $p_n = 10^{-2^n}$ converge cuadráticamente a cero.

Problema 2.37 Demuestre que la sucesión $p_n = 10^{-n^c}$ no converge cuadráticamente a cero, sin importar el tamaño del exponente $c > 1$.

Problema 2.38 Demuestre que el algoritmo de bisección determina una sucesión con una cota de error que converge linealmente a cero.

Problema 2.39 El método iterativo para resolver $f(x) = 0$, dado por el método de punto fijo $g(x) = x$, donde

$$p_{n+1} = g(p_n) = p_n - \frac{f(p_n)}{f'(p_n)} - \frac{f''(p_n)}{2f'(p_n)} \left[\frac{f(p_n)}{f'(p_n)} \right]^2 ,$$

para $n = 1, 2, 3, \ldots$. Puebe que α es una raíz simple de f, entonces se tiene $g(\alpha) = \alpha$, $g'(\alpha) = g''(\alpha) = 0$, esto generalmente, producirá una convergencia cúbica $(\alpha = 3)$.

Problema 2.40 Considere el siguiente sistema de ecuaciones

$$\begin{aligned} x^2 + y^2 - 9 &= 0 \\ x^2 + y^2 - 8x + 12 &= 0 \end{aligned}$$

1. Explicite por componentes y simplifique al máximo, el esquema iterativo definido por el método de Newton.

2. Se sabe que el sistema tiene una solución (α, β) en el primer cuadrante. Demuestre que el método iterativo de Newton es convergente en una vecindad de la solución (α, β). La demostración debe hacerse sin iterar y sin usar argumentos teóricos de punto fijo.

3. Realice 3 iteraciones con el método de Newton, comenzando con el punto $(2.6, 0.7)$ y estime el error del resultado de la tercera iteración respecto a la solución exacta $(\alpha, \beta) = (\frac{21}{8}, \frac{\sqrt{135}}{8})$.

Problema 2.41 Considere el siguiente sistema de ecuaciones

$$\begin{cases} x^2 + y^2 - 16 &= 0 \\ x^2 + y^2 - 8x &= 0 . \end{cases}$$

1. Explicite por componentes y simplifique al máximo, el esquema iterativo definido por el método de Newton.

2. Se sabe que el sistema tiene una solución (α, β) en el primer cuadrante. Demuestre que el método iterativo de Newton es convergente en una vecindad de la solución (α, β). La demostración debe hacerse sin iterar y sin usar argumentos teóricos de punto fijo.

3. Realice 3 iteraciones con el método de Newton, comenzando con el punto $(2.2, 3.4)$ y estime el error del resultado de la tercera iteración respecto a la solución exacta $(\alpha, \beta) = (3, \sqrt{12})$.

Problema 2.42 Considere el sistema de ecuaciones no lineales

$$\begin{cases} -x_1^2 + 8x_1 - x_2^2 - 6 &= 0 \\ -x_1^2 x_2 - x_1 + 8x_2 - 6 &= 0 . \end{cases}$$

1. Determinar una región $D \subset \mathbb{R}^2$ que contiene una única solución $\alpha = (\alpha_1, \alpha_2)$ del sistema.

2. Proponga un método de punto fijo convergente para determinar la solución $\alpha = (\alpha_1, \alpha_2) \in D$ del sistema. Demuestre la convergencia sin iterar.

3. Demuestre (sin iterar) que que el método de Newton converge a alguna solución del sistema Qué puede decir acerca de la rapidez de convergencia de ambos métodos (el propuesto en 2) y Newton en este caso particular?

4. Realice 2 iteraciones con el método propuesto, comenzando con $(0,0)$ (especifique cual de los métodos está usando)

Problema 2.43 Considere al ecuación $-x^3 + 2x^2 + 10x - 20 = 0$.

1. Proponga un método iterativo de punto fijo, no Newton, el cual sea convergente a la solución $\alpha = \sqrt{10}$. La demostración de la convergencia debe hacerse sin iteraciones.

2. Usando el método que propuso en 1), encuentre una solución aproximada a α con un error de no más de 10^{-2}

Problema 2.44 Considere la función $f(x) = e^{1/x} - x$, definida para $x > 0$.

1. Proponga un método iterativo convergente, no Newton, para encontrar una solución de la ecuación $f(x) = 0$. Use el método propuesto por usted para encontrar una solución de $f(x) = 0$, con una precisión $\varepsilon = 10^{-3}$, con el criterio de parada $|x_{n+1} - x_n| \leqslant \varepsilon$, y comenzando con $x_0 = 1.76$. La demostración de la convergencia debe hacerla en forma teórica.

2. Demuestre que al menos uno de los métodos iterativos propuestos abajo para encontrar la solución de $f(x) = 0$ es convergente (ambos podrían ser convergentes). La demostración de la convergencia debe hacerse en forma teórica.

$$x_{n+1} = g_1(x_n) = x_n - x_n^2 \frac{x_n - e^{1/x_n}}{x_n^2 + e^{1/x_n}}$$
$$x_{n+1} = g_2(x_n) = x_n + \frac{1 - x_n \ln(x_n)}{1 + \ln(x_n)}.$$

   Use el métodos que demostró ser convergente para encontrar una solución de $f(x) = 0$, con una precisión $\varepsilon = 10^{-3}$, con el criterio de parada $|x_{n+1} - x_n| \leqslant \varepsilon$, y comenzando con $x_0 = 1.76$.

3. Use el método de la secante para encontrar una solución de $f(x) = 0$, con una precisión $\varepsilon = 10^{-3}$, con el criterio de parada $|x_{n+1} - x_n| \leqslant \varepsilon$, comenzando con $x_0 = 1.76$ y $x_1 = 1.763$.

4. En los casos anteriores Cuál de los métodos converge más rápido?

Problema 2.45 Considere el sistema de ecuaciones no lineales

$$\begin{cases} x_1^2 - 10x_1 + x_2^2 + 8 &= 0 \\ x_1 x_2^2 + x_1 - 10x_2 + 8 &= 0. \end{cases}$$

1. Determinar una región $D \subset \mathbb{R}^2$ que contiene una única solución $\alpha = (\alpha_1, \alpha_2)$ del sistema.

2. Proponga un método de punto fijo convergente para determinar la solución $\alpha = (\alpha_1, \alpha_2) \in D$ del sistema. Demuestre la convergencia sin iterar.

3. Demuestre (sin iterar) que que el método de Newton converge a alguna solución del sistema Qué puede decir acerca de la rapidez de convergencia de ambos métodos (el propuesto en 2) y Newton en este caso particular?

4. Realice 2 iteraciones con el método propuesto, comenzando con $(0,0)$ (especifique cual de los métodos está usando)

Problema 2.46 Considere el sistema de ecuaciones no lineales $\mathbf{u}(x,y) = 0$, donde

$$\begin{cases} u_1(x,y) &= x^2 + xy - 10 \\ u_2(x,y) &= y + 3xy^2 - 57 \end{cases}$$

1. Proponga un método iterativo de punto fijo convergente, no Newton, para encontrar una solución al sistema, comenzando con $(x_0, y_0) = (1.5, 3.5)$

2. Aplique el método de Newton con las mismas condiciones iniciales. Obtenga dos iteraciones hay convergencia en este caso?

3. Si ambos métodos (1) y (2) son convergentes Cuál de ellos converge más rápido a la solución buscada del sistema?

Problema 2.47 Dado el sistema de ecuaciones no lineales

$$\begin{cases} 5\cos(x^2y) - 5\cos(x^3y) &= x \\ \operatorname{sen}^2(xy) + \frac{\pi}{2}\cos(x^3y) &= y \end{cases}$$

1. Determine un dominio en el plano que contenga una única solución (α, β) del sistema. Justifique teóricamente.

2. Proponga un método de punto fijo (no Newton) para encontrar la solución (α, β) del sistema. La demostración de la convergencia debe hacerse sin iterar.

3. Si considera una precisión $\varepsilon = 10^{-3}$. Cuál es la solución aproximada que se obtiene? (use el criterio de parada $||(x_n, y_n) - (x_{n-1}, y_{n-1})||_\infty \leqslant \varepsilon$.)

Problema 2.48 Dado el sistema de ecuaciones no lineales

$$\begin{cases} \operatorname{sen}(3x^2y) &= x \\ \cos(x) - 3\cos(x^3y^3) &= y \end{cases}$$

1. Determine un dominio en el plano que contenga una única solución (α, β) del sistema. Justifique teóricamente.

2. Proponga un método de punto fijo (no Newton) para encontrar la solución (α, β) del sistema. La demostración de la convergencia debe hacerse teóricamente.

3. Si considera una precisión $\varepsilon = 10^2$ y la norma $||(x,y)||_M = \max\{|x|, |y|\}$ Obtenga la solución aproximada al sistema en este caso.

Problema 2.49 el sistema de ecuaciones no lineales

$$\begin{cases} x^3 + x + 4x^2y + 4xy^2 &= 0.7 \\ 4x^3 + 3x - 4x^2y + xy^2 &= 0.3 \end{cases}$$

1. Determine un dominio en el plano que contenga una única solución (α, β) del sistema. Justifique teóricamente.

2. Proponga un método de punto fijo (no Newton) para encontrar la solución (α, β) del sistema. La demostración de la convergencia debe hacerse sin iterar.

3. Si considera una precisión $\varepsilon = 10^3$ Cuál es la solución aproximada que se obtiene?

Problema 2.50 Dado el sistema de ecuaciones no lineales

$$\begin{cases} x^{4/3} + y^{4/3} &= b^{4/3} \\ y - ax &= 0 \end{cases}$$

donde $a = 1/4$ y $b = \left(1 + \left(\frac{1}{4}\right)^{4/3}\right)^{3/4}$

1. Proponga un método de punto fijo, no Newton, y demuestre teóricamente que converge a la raíz $\alpha = (1, 1/4)$,

2. Proponga un método casi-Newton y demuestre teóricamente que converge a la raíz $\alpha = (1, 1/4)$.

3. Partiendo del punto inicial $(x_0, y_0) = (0.95, 0.23)$ realice 2 iteraciones con ambos métodos. Compare los resultados. Cuál es su conclusión?

Problema 2.51 Considere el sistema de ecuaciones no lineales

$$\begin{cases} x &= 1 + h\,\dfrac{e^{-x^2}}{1+y^2} \\ y &= 0.5 + h\,\arctan(x^2 + y^2)\,. \end{cases}$$

Muestre que si h es elegido suficientemente pequeño y no cero, entonces el sistema tiene una única solución $\alpha = (\alpha_1, \alpha_2)$ en alguna región rectangular. Además, muestre que el método iterativo de punto fijo definido por el sistema es convergente a la solución α para cualquier elección (x_0, y_0) elegida en la región que determinó. (La demostración de la convergencia debe hacerse sin iterar.)

Problema 2.52 Considere el sistema de ecuaciones no lineales

$$\begin{cases} x^2 + y^2 - 2x - 2y + 1 &= 0 \\ x + y - 2xy &= 0\,. \end{cases}$$

1. Proponga un método de punto fijo, no Newton, convergente para encontrar una solución del sistema. La demostración de la convergencia debe hacerse sin iterar.

2. Demuestre sin iterar que el método de Newton es convergente en un región que contiene una solución del sistema, y encuentre una solución con una aproximación de 10^{-3} usando la norma $||\ ||_\infty$.

Problema 2.53 Suponga que p es una raíz de multiplicidad $m \geqslant 2$ de la ecuación $f(x) = 0$, donde $f'''(x)$ es continua en un intervalo abierto que contiene a p.

(a) Demuestre que el método iterativo de punto fijo $g(x) = x - m\dfrac{f(x)}{f'(x)}$ tiene orden de convergencia la menos 2.

(b) Use la parte a) para calcular la raíz positiva de $x^4 - 1 = 0$, redondeando cada operación a 4 dígitos y con una precisión de 10^{-3}, comenzando con $x_0 = 0.8$.

Problema 2.54 Considere la ecuación $\mathrm{e}^{1-x} - 1 = 0$.

(a) Demuestre que ella tiene una única raíz, la cual es positiva.

(b) Demuestre que el método de Newton asociado a esta ecuación es convergente para cualquier condición inicial $x_0 \in [0, 10]$

(c) Comenzando con la condición inicial $x_0 = 10$. Realice algunas iteraciones con el método de Newton Cuál es su conclusión?

(d) Considere el intervalo $[0, 10]$, use el método de bisección para encontrar una aproximación a la raíz, con una precisión de 10^{-3}. Compare su resultado con lo obtenido en el item c). Cuál es su conclusión?

Problema 2.55 Considere el polinomio $f(x) = 230x^4 + 18x^3 + 9x^2 - 221x - 9$. Se sabe que este polinomio tiene una raíz en el intervalo $[0, 1]$.

a) Use el método de bisección para aproximar la raíz de $f(x)$ con una presición de 10^{-3}.

b) Use el método de Newton para aproximar la raíz de $f(x)$ con una presición de 10^{-5}, comenzando con $x_0 = 0.3$.

c) Proponga un método iterativo de punto fijo convergente para aproximar la raíz de $f(x)$. La convergencia debe demostrarla sin iterar.

Problema 2.56 Considere el sistema de ecuaciones no lineales

$$\begin{cases} 3x - \cos(yz) - \frac{1}{2} &= 0 \\ x^2 - 81(y + 0.1)^2 + \operatorname{sen}(z) + 1.06 &= 0 \\ e^{-xy} + 20z + \frac{10\pi - 3}{3} &= 0 \end{cases}$$

a) Proponga un método iterativo de punto fijo convergente para encontrar la solución del sistema. La demostración de la convergencia debe hacerse sin iterar.

b) Explícite las componentes del método iterativo de Newton para este sistema, y demuestre la convergencia del método para este caso (teóricamente).

c) Usando el método iterativo convergente que propuso en a), comenzando con el punto $(0.1, 0.1, -0.1)$ realice iteraciones hasta obtener una aproximación (x_k, y_k, z_k) de la solución, la cual satisface la siguiente condición: $||(x_k, y_k, z_k) - (x_{k-1}, y_{k-1}, z_{k-1})||_\infty \leqslant 10^{-3}$.

Problema 2.57 Dado el sistema de ecuaciones no lineales

$$\begin{cases} x^2 + xy^3 &= 9 \\ 3x^2y - y^3 &= 4 \end{cases}$$

Estudie la convergencia del método de Newton para el sistema, suponiendo que su condicións iniciales son:

(a) $(x_0, y_0) = (1.3, 1.7)$,

(b) $(x_0, y_0) = (-1, -2)$,

(c) $(x_0, y_0) = (-3, 0.2)$

(d) Realizando 4 iteraciones encuentre una aproximación a la solución en el caso (b)

Problema 2.58 Considere el sistema

$$\begin{cases} x^2 + y^2 &= 1 \\ 2x^2 - y &= 0 \end{cases}$$

1. Demuestre, sin iterar, que el método iterativo de Newton es convergente para este caso.
2. Usando el método iterativo de Newton, encuentre la solución del sistema ubicada en el primer cuadrante.

Problema 2.59 Dado el sistema de ecuaciones no lineales

$$\begin{cases} x + x(x+y)^2 - 0.5 &= 0 \\ \\ y + y(y-x)^2 - 0.5 &= 0 \end{cases}$$

a) Determine un dominio en el plano que contenga una única solución (α, β) del sistema. Justifique teóricamente.

b) Proponga un método de punto fijo, no Newton, y demuestre, sin iterar, que converge a (α, β).

Problema 2.60 Dado el sistema de ecuaciones no lineales

$$\begin{cases} \operatorname{sen}(3x^2 y) &= x \\ \cos(x) - 3\cos(x^3 y^2) &= y \end{cases}$$

1. Determine un dominio en el plano que contenga una única solución (α, β) del sistema. Justifique teóricamente.
2. Proponga un método de punto fijo convergente (no Newton) para encontrar la solución (α, β) del sistema. La demostración de la convergencia debe hacerse teóricamente.
3. Considere una precisión $\varepsilon = 10^2$ y la norma $||(x, y)||_M = \max\{|x|, |y|\}$. Obtenga la solución aproximada al sistema en este caso.

Problema 2.61 Dado el sistema no lineal

$$\begin{cases} x_1^2 + x_2^2 + x_1 - x_2 - 1 = 0 \\ x_1^2 + 2x_2^2 + x_1 + x_2 - 1 = 0 \end{cases}$$

y considere la raíz $\alpha = (\alpha_1, \alpha_2)^T$, tal que $\alpha_1 > 0$.

1. Demuestre sin iterar, que el método de Newton converge a α, si la condición inicial $x^{(0)} = (x_1^{(0)}, x_2^{(0)})$ se elige suficientemente cerca de α.

2. Partiendo de $x^{(0)} = (0.6, 0.1)^T$ y utilizando la máxima capacidad de su calculadora, resuelva el sistema dado, para la raíz α mediante el método de Newton con una presición de 10^{-2} (utilizando el criterio de parada $||x^{(k+1)} - x^{(k)}|| \leqslant \varepsilon$)

Problema 2.62 Considere la ecuación $x^4 - 3x^3 + 4x^2 - 3x + 1 = 0$. Esta tiene una raíz $\alpha = 1$

1. Use el método de Newton para determinar dicha raíz α, partiendo desde $x_0 = 0.8$ y realizando 5 iteraciones.

2. Use el siguiente método iterativo

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

para determinar en forma aproximada la raíz α, comenzando con $x_0 = 0.8$.

3. Utilizando los resultados de las iteraciones en a) y en b), compare la rapidez de convergencia de ambos métodos, estimando el error relativo de la aproximación x_5, correspondiente.

Problema 2.63 Considere la ecuación

$$\cos(3x) + \cos^3(61x) = 0.$$

1. Proponga un método de punto fijo (no del tipo Newton) convergente a la raíz $\alpha = \frac{\pi}{2}$. La demostración de convergencia debe hacerse sin iterar.

2. Sin iterar, demuestre que el método de la secante converge a dicha raíz α.

Problema 2.64 Considere la función $f(x) = 2x^2 - x + 6\mathrm{e}^{-x} - 8$.

1. Usando el método iterativo de Newton, comenzando con $x_0 = 0.3$. Encuentre una raíz negativa de $f(x) = 0$ con precisión de 10^{-4}, utilizando la máxima capacidad de dígitos de su calculadora.

2. Usando el método iterativo de la Secante, comenzando con $x_0 = 0.3$ y $x_1 = 0$. Encuentre una raíz negativa de $f(x) = 0$ con precisión de 10^{-4}, utilizando la máxima capacidad de dígitos de su calculadora.

3. Considerando los resultados obtenidos en 1) y 2) Qué puede decir acerca de la convergencia de ambos métodos en este caso particular? Compare con los resultados teóricos relacionados.

Problema 2.65 Considere la ecuación $230x^4 + 18x^3 + 9x^2 - 221x - 9 = 0$.

1. Demuestre (teóricamente) que la ecuación tiene una raíz α en el intervalo $[0, 1]$.

2. Usando el método de bisección encuentre la raíz α, con una precisión de 10^{-3}. Cuántos iteraciones serían necesarias para obtener una aproximación de la raíz, con precisión de 10^{-3}, en el peor de los casos? (Use la fórmula de la cota del error).

3. Use el método de Newton para determinar dicha raíz α, partiendo desde $x_0 = 0.8$ con una precisión de 10^{-3}.

Problema 2.66 Considere el sistema de ecuaciones no lineales

$$\begin{cases} 4x^2 - 40x + \frac{1}{4}y + 8 &= 0 \\ \frac{1}{2}xy^2 + 2x - 10y + 8 &= 0\,. \end{cases}$$

1. Determinar una región $D \subset \mathbb{R}^2$ que contiene una única solución $\alpha = (\alpha_1, \alpha_2)$ del sistema.

2. Proponga un método de punto fijo convergente para determinar la solución $\alpha = (\alpha_1, \alpha_2) \in D$ del sistema. Demuestre la convergencia sin iterar. Realizando iteraciones comenzando con $(x_0, y_0) = (0, 0)$ y utilizando la máxima capacidad de su calculadora, obtenga la solución con una precisión de $\varepsilon = 0.05$, usando como criterio $\sqrt{(x_n - x_{n+1})^2 + (y_n - y_{n+1})^2} \leqslant \varepsilon$.

3. Demuestre (sin iterar) que que el método de Newton converge a alguna solución del sistema Qué puede decir acerca de la rapidez de convergencia de ambos métodos (el propuesto en 2) y Newton) en este caso particular?

Problema 2.67 Considere el sistema de ecuaciones no lineales

$$\begin{cases} 4x^2 - 20x + \frac{1}{4}y + 8 &= 0 \\ \frac{1}{2}xy^2 + 2x - 5y + 8 &= 0\,. \end{cases}$$

1. Determinar una región $D \subset \mathbb{R}^2$ que contiene una única solución $\alpha = (\alpha_1, \alpha_2)$ del sistema.

2. Proponga un método de punto fijo convergente para determinar la solución $\alpha = (\alpha_1, \alpha_2) \in D$ del sistema. Demuestre la convergencia sin iterar. Realizando iteraciones comenzando con $(x_0, y_0) = (0, 0)$ y utilizando la máxima capacidad de su calculadora, obtenga la solución con una precisión de $\varepsilon = 0.05$, usando como criterio $\sqrt{(x_n - x_{n+1})^2 + (y_n - y_{n+1})^2} \leqslant \varepsilon$.

3. Demuestre (sin iterar) que que el método de Newton converge a alguna solución del sistema Qué puede decir acerca de la rapidez de convergencia de ambos métodos (el propuesto en 2) y Newton) en este caso particular?

Problema 2.68 Considere la ecuación $f(x) = x^3 - 5x^2 + 3x - 7 = 0$. Se sabe que esta ecuación tiene una raíz α cerca de 4.6783.

1. Use el método de Newton para obtener una aproximación para α con una presición $\varepsilon = 10^{-8}$, utilizando como criterio de parada $|f(x_n)| < \varepsilon$, y como punto inicial a $x_0 = 4.67$. Justifique teoricamente la convergencia del método de Newton en este caso.

2. Considere el método iterativo siguiente

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

Use este método para obtener una aproximación a la raíz α de la ecuación anterior con una presición $\varepsilon = 10^{-8}$, utilizando como criterio de parada $|f(x_n)| < \varepsilon$, y como punto inicial a $x_0 = 4.67$.

3. Cuál de ellos converge más rápido?

Problema 2.69 Considere la función $f(x) = e^{1/x} - x$, definida para $x > 0$.

1. Demuestre que al menos uno de los métodos iterativos propuestos abajo para encontrar la solución de $f(x) = 0$ es convergente (ambos podrían ser convergentes). La demostración de la convergencia debe hacerse en forma teórica.

$$x_{n+1} = g_1(x_n) = x_n - x_n^2 \frac{x_n - e^{1/x_n}}{x_n^2 + e^{1/x_n}} \qquad \text{y} \qquad x_{n+1} = g_2(x_n) = x_n + \frac{1 - x_n \ln(x_n)}{1 + \ln(x_n)}$$

   Use el métodos que demostró que es convergente para encontrar una solución de $f(x) = 0$, con una precisión $\varepsilon = 10^{-3}$, con el criterio de parada $|x_{n+1} - x_n| \leqslant \varepsilon$, y comenzando con $x_0 = 1.76$.

2. Proponga un método iterativo convergente, no Newton, para encontrar una solución de la ecuación $f(x) = 0$. Use el método propuesto por usted para encontrar una solución de $f(x) = 0$, con una precisión $\varepsilon = 10^{-3}$, con el criterio de parada $|x_{n+1} - x_n| \leqslant \varepsilon$, y comenzando con $x_0 = 1.76$. La demostración de la convergencia debe hacerla en forma teórica.

3. Use el método de la secante para encontrar una solución de $f(x) = 0$, con una precisión $\varepsilon = 10^{-3}$, con el criterio de parada $|x_{n+1} - x_n| \leqslant \varepsilon$, comenzando con $x_0 = 1.76$ y $x_1 = 1.763$.

4. En los casos anteriores Cuál de los métodos converge más rápido?

Problema 2.70 Considere el sistema de ecuaciones no lineales

$$\begin{cases} -x_1^2 + 8x_1 - x_2^2 - 6 &= 0 \\ -x_1^2 x_2 - x_1 + 8x_2 - 6 &= 0\,. \end{cases}$$

1. Determinar una región $D \subset \mathbb{R}^2$ que contiene una única solución $\alpha = (\alpha_1, \alpha_2)$ del sistema.

2. Proponga un método de punto fijo convergente para determinar la solución $\alpha = (\alpha_1, \alpha_2) \in D$ del sistema. Demuestre la convergencia sin iterar.

3. Demuestre (sin iterar) que que el método de Newton converge a alguna solución del sistema . Qué puede decir acerca de la rapidez de convergencia de ambos métodos (el propuesto en 2) y Newton) en este caso particular?

4. Realice 2 iteraciones con el método propuesto en 2) o 3), comenzando con $(0, 0)$ (especifique cual de los métodos está usando)

Problema 2.71 Considere el siguiente sistema de ecuaciones

$$\begin{cases} x^2 + y^2 - 9 &= 0 \\ x^2 + y^2 - 8x + 12 &= 0 \end{cases}$$

1. Explicite por componentes y simplifique al máximo, el esquema iterativo definido por el método de Newton.

2. Se sabe que el sistema tiene una solución (α, β) en el primer cuadrante. Demuestre que el método iterativo de Newton es convergente en una vecindad de la solución (α, β). La demostración debe hacerse sin iterar y sin usar argumentos teóricos de punto fijo.

3. Realice 3 iteraciones con el método de Newton, comenzando con el punto $(2.6, 0.7)$ y estime el error del resultado de la tercera iteración respecto a la solución exacta $(\alpha, \beta) = (\frac{21}{8}, \frac{\sqrt{135}}{8})$.

Problema 2.72 Considere al ecuación $-x^3 + 2x^2 + 10x - 20 = 0$.

1. Proponga un método iterativo de punto fijo, no Newton, el cual sea convergente a la solución $\alpha = \sqrt{10}$. La demostración de la convergencia debe hacerse sin iteraciones.

2. Usando el método que propuso en a), encuentre una solución aproximada a α con un error de no más de 10^{-2}

Problema 2.73 Considere al ecuación $-x^3 + 2x^2 + 10x - 20 = 0$.

1. Proponga un método iterativo de punto fijo, no Newton, el cual sea convergente a la solución $\alpha = -\sqrt{10}$. La demostración de la convergencia debe hacerse sin iteraciones.

2. Usando el método que propuso en a), encuentre una solución aproximada a α con un error de no más de 10^{-2}

Problema 2.74 Considere el siguiente sistema de ecuaciones

$$\begin{cases} x^2 + y^2 - 16 &= 0 \\ x^2 + y^2 - 8x &= 0 \end{cases}$$

1. Explicite por componentes y simplifique al máximo, el esquema iterativo definido por el método de Newton.

2. Se sabe que el sistema tiene una solución (α, β) en el primer cuadrante. Demuestre que el método iterativo de Newton es convergente en una vecindad de la solución (α, β). La demostración debe hacerse sin iterar y sin usar argumentos teóricos de punto fijo.

3. Realice 3 iteraciones con el método de Newton, comenzando con el punto $(2.2, 3.4)$ y estime el error del resultado de la tercera iteración respecto a la solución exacta $(\alpha, \beta) = (3, \sqrt{12})$.

Problema 2.75 Considere el sistema de ecuaciones no lineales

$$\begin{cases} x_1^2 - 10x_1 + x_2^2 + 8 &= 0 \\ x_1 x_2^2 + x_1 - 10x_2 + 8 &= 0\,. \end{cases}$$

1. Determinar una región $D \subset \mathbb{R}^2$ que contiene una única solución $\alpha = (\alpha_1, \alpha_2)$ del sistema.

2. Proponga un método de punto fijo convergente para determinar la solución $\alpha = (\alpha_1, \alpha_2) \in D$ del sistema. Demuestre la convergencia sin iterar.

3. Demuestre (sin iterar) que que el método de Newton converge a alguna solución del sistema Qué puede decir acerca de la rapidez de convergencia de ambos métodos (el propuesto en 2) y Newton) en este caso particular?

4. Realice 2 iteraciones con el método propuesto en 2) o 3), comenzando con $(0,0)$ (especifique cual de los métodos está usando)

Problema 2.76 Sea $f(x) = (x-3)^5$.

(a) Determine la fórmula de Newton $x_{n+1} = g(x_n)$ para calcular el cero de la función f. Tomando como valor inicial $x_0 = 1$, calcule según la fórmula de iteración de Newton encontrada en la parte (a) los valores de x_1, x_2 y x_3.

(b) Calcule usando como valores iniciales x_0 y x_1 obtenidos en (a), los valores de x_2 y x_3 usando el método de la secante. ¿Podría usted usar el método de Regula-falsi?. Explique su respuesta en el caso negativo o partiendo de dos puntos iniciales apropiados x_0 y x_1, obtenga los valores de x_2 y x_3 en caso afirmativo.

(c) Demuestre que el método de Newton usado en la parte (a) tiene un orden de convergencia $p = 1$. De un método alternativo de iteración que garantice un orden de convergencia $p = 2$ y demuestre que efectívamente este es su orden de convergencia.

Problema 2.77 Aplique el método de la bisección y regula falsi para encontrar una aproximación a solución de ecuación $x = \tan(x)$ en [4,4.5]. Use una tolerancia de 10^{-3}.

Problema 2.78 Use el método de la bisección y regula falsi para encontrar una aproximación a $\sqrt{5}$ correcta en 5 cifras significativas. Sugerencia: considere $f(x) = x^2 - 5$.

Problema 2.79 Calcule el número de iteraciones que se requieren usando el método de bisección para alcanzar una aproximación con una exactitud de 10^3 a la solución de $x^3 + x - 4 = 0$ que se encuentra en el intervalo $[1,4]$. Obtenga una aproximación de la raíz con este grado de exactitud.

Problema 2.80 Sean $f(x) = (x-1)^{10}$, $\alpha = 1$ y $x_n = 1 + 1/n$. Pruebe que $|f(x_n)| < 10^{-3}$ siempre que $n > 1$ pero que $|\alpha - x_n| < 10^{-3}$ requiere que $n > 1000$. Este ejercicio muestra una función $f(x)$ tal que $|f(x_n)|$ se aproxima a cero, mientras que la sucesión de aproximaciones $(x_n)_{n\in\mathbb{N}}$ lo hace mucho más lento.

Problema 2.81 Sea $(x_n)_{n\in\mathbb{N}}$ la sucesión definida por $x_n = \sum^{k=1} 1/k$. Pruebe que $(x_n)_{n\in\mathbb{N}}$ diverge aún cuando $\lim_{n\to\infty}(x_n - x_{n-1}) = 0$. Este ejercicio muestra una sucesión $(x_n)_{n\in\mathbb{N}}$ con la propiedad de que las diferencias $x_n - x_{n-1}$ (en referencia al error relativo) convergen a cero, mientras que la sucesión $(x_n)_{n\in\mathbb{N}}$ diverge.

Problema 2.82 El método dede bisección se puede aplicar siempre que $f(a)f(b) < 0$. Si $f(x)$ tiene más de un cero en (a,b) se podrá saber de antemano cuál cero es el que se encuentra al aplicar el método de la bisección? ilustre su respuesta con ejemplos.

Problema 2.83 Las siguientes funciones cumplen con la condición $f(a) \cdot f(b) < 0$, donde $a = 0$ y $b = 1$. Si se aplica el método de bisección en el intervalo $[a, b]$ a cada una de esas funciones Qué punto se encuentra en cada caso? Es este punto un cero de f?

a. $f(x) = (3x - 1)^{-1}$ **b.** $f(x) = \cos(10x)$ **c.** $f(x) = \begin{cases} 1, & \text{para } x > 0 \\ -1, & \text{para } x \leqslant 0 \,. \end{cases}$

Problema 2.84 Verifique que se puede aplicar el método de bisección para aproximar el único cero de la función $f(x) = x^3 - x - 1$ en el intervalo $[1, 2]$ Cuántas iteraciones serán necesarias para que al aplicar el método de bisección en el intervalo [1,2] se logre una aproximación de por lo menos 3 cifras decimales exactas? Calcule tal aproximación.

Problema 2.85 Estudie la función $g(x) = \sqrt{1 + x^2}$ como una posible función de iteración de punto fijo Por qué no es convegente la iteración $x_n = g(x_{n-1})$, $n = 1, 2, \ldots$?.

Problema 2.86 Verifique que cada una de las siguientes funciones $g_i(x)$, $i = 1, 2, 3, 4$ es una función de iteración de punto fijo para la ecuación $x^4 + 2x^2 - x - 3 = 0$, es decir, $\alpha = g_i(\alpha)$, $i = 1, 2, 3, 4$ implica que $f(\alpha) = 0$, siendo $f(x) = x^4 + 2x^2 - x - 3$

a) $g_1(x) = (3 + x - 2x^2)^{1/4}$

b) $g_2(x) = \left(\dfrac{3 + x - x^4}{2}\right)^{1/2}$

c) $g_3(x) = \left(\dfrac{3 + x}{x^2 + 2}\right)^{1/2}$

d) $g_4(x) = \dfrac{3x^4 + 2x^2 + 3}{4x^3 + 4x - 1}$.

Efectúe 4 iteraciones, si es posible, con cada una de las funciones de iteración definidas arriba, tomando $x_0 = 1$ y $x_n = g_i(x_{n-1}), i = 1, 2, 3, 4$ Cuál función cree usted que da la mejor aproximación? Explique.

Problema 2.87 Demuestre que la ecuación $2 \sin(\pi x) + x = 0$ tiene una única raíz $\alpha \in [1/2, 3/2]$. Use un método de iteración de punto fijo para encontrar una aproximaión de α con una precisión de por lo menos tres cifras significativas.

Problema 2.88 Pruebe que la función $g(x) = 2 + x - \tan^{-1}(x)$ tiene la propiedad $|g'(x)| < 1$ para toda x. Pruebe que g no tiene un punto fijo. Explique porqué esto no contradice los teoremas sobre existencia de punto fijo.

Problema 2.89 Use un método iterativo de punto fijo para demostrar que la sucesión $(x_n)_{n \in \mathbb{N}}$ definida por

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right), \quad n = 0, 1, 2, \ldots$$

converge a $\sqrt{2}$ para $x_0 > 0$ escogido adecuadamente.

En general, si $R > 0$, entonces la sucesión $(x_n)_{n \in \mathbb{N}}$ definida por

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{R}{x_n}\right), \quad n = 0, 1, 2, \cdots$$

converge a $\sqrt{R}$ para $x_0 > 0$ escogido adecuadamente. Esta sucesión se usa con frecuencia en subrutinas para calcular raíces cuadradas.

Problema 2.90 Cuál es el valor de la siguiente expresión?

$$x = \sqrt{2 + \sqrt{2 + \sqrt{2+}}}$$

Note que esta expresión puede ser interpretada como significado $x = \lim_{n\to\infty} x_n$, donde $x_0 = \sqrt{2}, x_1 = \sqrt{2+\sqrt{2}} = \sqrt{2+x_0}$ y así sucesivamente. USe un método de punto fijo con una función de iteración $g(x)$ apropiada.

Problema 2.91 Resuelva la ecuación $4\cos(x) = e^x$ con una precisión de 5×10^{-5}, es decir, calcular las iteraciones x_n hasta que $|x_n - x_{n-1}| < 5 \times 10^{-5}$ usando

a) El método de Newton con $x_0 = 1$

b) El método de la secante con $x_0 = \pi/4$ y $x_1 = \pi/2$.

Problema 2.92 Use el método de Newton para resolver la ecuación

$$\left(\text{sen}(x) - \frac{x}{2}\right)^2 = 0, \quad \text{con } x_0 = \frac{\pi}{2}.$$

Itere hasta obtener una precisión de 5×10^{-5} para la raíz aproximada con $f(x) = (\text{sen}(x) - \frac{x}{2})^2$ e parecen los resultados fuera de lo común para el método de Newton? Resuelva también la ecuación con $x_0 = 5\pi$ y $x_0 = 10\pi$.

Problema 2.93 se el método de Newton modificado para encontrar una aproximación de la raíz de la ecuación

$$f(x) = x^2 + 2xe^x + e^{2x} = 0$$

empezando con $x_0 = 0$ y efectuando 10 iteraciones ¿Cúal es la multiplicidad de la raíz buscada?

Problema 2.94 Se desea resolver la ecuación no lineal $f(x) = 0$, donde $f : \mathbb{R} \longrightarrow \mathbb{R}$ tiene tantas derivadas cuantas sean necesarias Es posible elegir una función h de modo que la fórmula iterativa siguiente

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} + h(x_n)\left(\frac{f(x_n)}{f'(x_n)}\right)^2$$

tenga orden de convergencia 3 a una raíz simple de $f(x) = 0$?

Problema 2.95 Considere una firma que desea maximizar sus ganancias, eligiendo el precio de venta de su producto, denotado por y y la cantidad gastada en publicidad, denotada por z. Para esto, se supone que la función de beneficio B viene dada por

$$B = yx - (z + g_2(x)),$$

con $x = g_1(y,z) = a_1 + a_2y + a_3z + a_4yz + a_5z^2$, $g_2(x) = e_1 + e_2x$, y donde

$$\begin{aligned} B &= \text{Beneficio} \\ x &= \text{número de unidades vendidas} \\ y &= \text{precio de venta por unidad} \\ z &= \text{dinero gastado en publicidad} \\ g_1(y,z) &\quad \text{representa una predicción de unidades vendidas} \\ &\quad \text{cuando el precio es y y el gasto en publicidad es } z \\ g_2(x) &\quad \text{es el costo de producir } \times \text{ unidades} \end{aligned}$$

y los parámetros a_i y e_i son

$$\begin{aligned} a_1 = 50000, &\quad a_2 = -5000, \quad a_3 = 40, \quad a_4 = -1, \quad a_5 = -0.002 \\ e_1 = 100000, &\quad e_2 = 2. \end{aligned}$$

Se pide encontrar los valores de y y de z que maximicen el Beneficio B. Para esto, resuelva el sistema no lineal

$$\nabla B(y,z) = 0.$$

Problema 2.96 Sea $f(x) = e^x - 1 - \cos(\pi x)$. Muestre que la ecuación $f(x) = 0$ tiene una única raíz en $[0,1]$. Estudie que ocurre con el método de Newton comenzando con las condiciones iniciales $x_0 = -\varepsilon$ y $x_0 = 1 + \varepsilon$, $x_0 = 0$ y $x_0 = 1$.

Problema 2.97 Defina $f :]0,\infty[\longrightarrow \mathbb{R}$ definida como $f(x) = \frac{8x-1}{x} - e^x$. Considere los métodos iterativos

$$x_{n+1} = f_1(x_n) = \frac{1}{8}\left(1 + x_n e_n^x\right) \quad \text{y} \quad x_{n+1} = f_2(x_n) = \ln\left(\frac{8x_n - 1}{x_n}\right)$$

Cual de ellos es convergente a la raíz de $f(x) = 0$ más próxima de cero? Cual de ellos es convergente a la raíz de $f(x) = 0$ más próxima de 2?

Problema 2.98

Problema 2.99

2.12 Uso de métodos de integración para obtener fórmulas iterativas para resolver ecuaciones no lineales

[1]

[1] En una primera lectura, puede obviarse esta sección

Es común obtener la fórmula de iteración de Newton

$$N_f(x) = x - \frac{f(x)}{f'(x)}$$

mediante un argumento geométrico. En esta nota veremos que podemos obtener esa fórmula a partir del Teorema Fundamental del Cálculo.

Del teorema Fundamental del Cálculo tenemos que

$$f(x) = f(x_n) + \int_{x_n}^{x} f'(t)dt \tag{2.25}$$

aproximando el valor de la integral por el área del rectángulo de base x_n, x y altura $f'(x_n)$, obtenemos

$$\int_{x_n}^{x} f'(t)dt \approx f'(x_n)(x - x_n) \tag{2.26}$$

luego, reemplazando esto en (4.18) obtenemos

$$f(x) \approx f(x_n) + f'(x_n)(x - x_n) \tag{2.27}$$

si x_n es tal que $f(x_{n+1}) = 0$, tenemos

$$f(x_n) + f'(x_n)(x_{n+1} - x_n) \approx 0\,, \tag{2.28}$$

de donde

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{2.29}$$

que no es otra que la fórmula iterativa de Newton.

2.13 Otras fórmulas iterativas

Ahora si usamos la regla de Simpson para aproximar las integral (que estudiaremos en forma detallada en un capítulo mas adelante)

$$\int_{x_n}^{x} f'(t)dt\,,$$

obtenemos

$$\int_{x_n}^{x} f'(t)dt \approx \frac{x - x_n}{6}\left[f'(x) + 4f'(\frac{x + x_n}{2}) + f'(x_n)\right] \tag{2.30}$$

y tenemos la ecuación

$$\hat{M}(x) = f(x_n) + \frac{x - x_n}{6}\left[f'(x) + 4f'(\frac{x + x_n}{2}) + f'(x_n)\right]. \tag{2.31}$$

Sea x_{n+1} la solución de la ecuación $\hat{M}(x) = 0$, entonces obtenemos

$$f(x_n) + \frac{x - x_n}{6}\left[f'(x) + 4f'(\frac{x + x_n}{2}) + f'(x_n)\right] = 0 \tag{2.32}$$

de donde

$$x_{n+1} = x_n - \frac{6f(x_n)}{f'(x_{n+1}) + 4f'(\frac{x_n + x_{n+1}}{2}) + f'(x_n)} \tag{2.33}$$

si reemplazamos x_{n+1} por $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ en el lado derecho de la igualdad anterior, obtenemos la fórmula iterativa

$$x_{n+1} = x_n - \frac{6f(x_n)}{f'\left(x_n - \frac{f(x_n)}{f'(x_n)}\right) + 4f'\left(x_n - \frac{1}{2}\frac{f(x_n)}{f'(x_n)}\right) + f'(x_n)} \tag{2.34}$$

Este es un método de tercer orden de convergencia en las raíces simples de $f(x)$.

Tenemos también el método iterativo

$$x_{n+1} = x_n - \frac{2f(x_n)}{f'(x_n) + f'\left(x_n - \frac{f(x_n)}{f'(x_n)}\right)} \tag{2.35}$$

que también es de orden de convergencia 3 en las raíces simples de $f(x)$. Para obtener esta fórmula iterativa, aproximamos la integral

$$\int_{x_n}^{x} f'(t)dt$$

por la regla de los trapecios, es decir,

$$\int_{x_n}^{x} f'(t)dt \approx \frac{x - x_n}{2}\left(f('(x) + f'(x_n)\right). \tag{2.36}$$

Resolviendo la ecuación

$$M_f(x) = f(x_n) + \frac{(x - x_n)}{2}\left(f'(x) + f'(x_n)\right) = 0, \tag{2.37}$$

si x_{n+1} es su solución, obtenemos

$$f(x_n) + \frac{(x_{n+1} - x_n)}{2}\left(f'(x_n) + f'(x_{n+1})\right) = 0, \tag{2.38}$$

de donde

$$x_{n+1} = x_n - \frac{2f(x_n)}{f'(x_n) + f'(x_{n+1})} \tag{2.39}$$

reemplazando x_{n+1} por $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ en el lado derecho de esta última igualdad, tenemos

$$x_{n+1} = x_n - \frac{2f(x_n)}{f'(x_n) + f'\left(x_n - \frac{f(x_n)}{f'(x_n)}\right)}\,. \tag{2.40}$$

Capítulo 3

Sistemas de Ecuaciones Lineales

Sistemas de ecuaciones lineales se utilizan en muchos problemas de ingeniería y de las ciencias, así como en aplicaciones de las matemáticas a las ciencias sociales y al estudio cuantitativo de problemas de administración y economía.

En este capítulo se examinan métodos iterativos y métodos directos para resolver sistemas de ecuaciones lineales

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n\,. \end{cases} \tag{3.1}$$

Este sistema puede ser expresado en la forma matricial como

$$A\mathbf{x} = \mathbf{b}, \tag{3.2}$$

donde $A = (a_{ij})_{n\times n} \in M\,(n \times n, \mathbb{R})\,,\ \mathbf{b} = (b_1\,,\, b_2\,,\, \ldots\,,\, b_n)^T \in \mathbb{R}^n$ y $\mathbf{x} = (x_1\,,\, x_2\,,\, \ldots\,,\, x_n)^T \in \mathbb{R}^n$ es la incógnita.

En la siguiente sección estudiaremos algunos métodos directos para resolver sistemas de ecuaciones lineales, posteriormente abordaremos algunos métodos iterativos .

3.1 Normas matriciales

Para estudiar algunos métodos iterativos que nos permitan resolver sistemas de ecuaciones lineales, necesitamos de algunos conceptos básicos de Álgebra Lineal.

3.1.1 Normas vectoriales

Sea V un espacio vectorial de dimensión finita.

Definición 3.1 *Una norma en V es una función $||\ ||: V \longrightarrow \mathbb{R}$ que satisface*

N1*)* $\|x\| \geqslant 0$ *para todo* $x \in V$ *,*

N2*)* $\|ax\| = |a|\ \|x\|$ *para todo* $x \in V$ *y todo* $a \in \mathbb{R}$,

N3*)* $\|x+y\| \leqslant \|x\| + \|y\|$ *para todo* $x, y \in V$ *(desigualdad triangular).*

Ejemplo 44 En $\mathbb{R}^n$ podemos definir las siguientes normas

1. $\| (x_1, x_2, \ldots, x_n) \|_2 = (\sum_{i=1}^n x_i^2)^{1/2}$ (*norma euclideana*)
2. $\| (x_1, x_2, \ldots, x_n) \|_\infty = \max\{\ |x_i|\ :\ 1 \leqslant i \leqslant n\ \}$
3. $\| (x_1, x_2, \ldots, x_n) \|_1 = \sum_{i=1}^n |x_i|$
4. $||(x_1, x_2, \ldots, x_n)||_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ (*norma* p).

Ejemplo 45 Denotemos por $V = M(n \times n, \mathbb{R})$ el espacio vectorial real de las matrices de orden $n \times n$ con entradas reales. Definimos las siguientes normas en V

1. $||A||_2 = \sqrt{\text{traza}(AA^T)}$
2. $||A||_F = \left(\sum_{i,j=1}^n a_{ij}^2\right)^{1/2} = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2\right)^{1/2}$ (norma de Fröbenius)
3. $||A||_{\alpha,\beta} = \sup\{||A\mathbf{x}||_\beta\ :\ \mathbf{x} \in \mathbb{R}^n,\ \text{ con }\ \|\mathbf{x}\|_\alpha = 1\}$, donde $||\ ||_\beta$ y $||\ ||_\beta$, denotan normas cualesquiera norma sobre $\mathbb{R}^n$. Esta norma es llamada *norma subordinada* a las norma en $\mathbb{R}^n$.

   Notemos que también se tiene

$$\|A\|_{\alpha,\beta} = \sup\left\{\ \frac{\|A\mathbf{x}\|_\beta}{\|\mathbf{x}\|_\alpha}\ :\ \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\|_\alpha \neq 0\ \right\}.$$

   La verificación que lo anterior define una norma sobre el espacio vectorial $M(n \times n, \mathbb{R})$ no es difícil, por ejemplo N1 es inmediata, para verificar N2, usamos la propiedad siguiente del supremo de conjuntos en $\mathbb{R}$, $\sup(\lambda A) = \lambda \sup(A)$, para todo $\lambda \geqslant 0$, donde $\lambda A = \{\lambda x : x \in A\}$. Finalmente, para verificar N3, usamos la propiedad $\sup(A+B) \leqslant \sup(A) + \sup(B)$, donde $A + B = \{\ x + y : x \in A,\ y \in B\ \}$.

Observación.

1. Si $A \in M(n \times n, \mathbb{R})$ es no singular, esto es, $\det(A) \neq 0$, entonces

$$\inf\{||A\mathbf{x}||\ :\ ||\mathbf{x}|| = 1\} = \frac{1}{||A^{-1}||}.$$

2. $||A||_2 = \sqrt{\lambda_{\max}}$, donde $\lambda_{\max}$ es el mayor valor propio de AA^T.
3. Si $A \in M(n \times n, \mathbb{R})$ es no singular, entonces

$$||A^{-1}||_2 = \frac{1}{\inf\{||A\mathbf{x}||_2\ :\ ||\mathbf{x}||_2 = 1\}} = \frac{1}{\sqrt{\lambda_{\min}}},$$

   donde $\lambda_{\min}$ es el menor valor propio de AA^T.

4. $||A||_2 = ||A^T||_2$.

5. $$\left|\left|\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right|\right|_2 = \max\{||A||_2, ||B||_2\} .$$

Teorema 3.1 *Para* $A \in M(n \times n, \mathbb{R})$ *y* $\mathbf{x} \in \mathbb{R}^n$ *, se tiene que* $\|A\mathbf{x}\| \leqslant \|A\| \, \|\mathbf{x}\|$.

Demostración. Desde la definición de norma matricial se tiene que $||A|| \geqslant \frac{||A\mathbf{x}||}{||\mathbf{x}||}$, para todo $\mathbf{x} \in \mathbb{R}^n$, con $\mathbf{x} \neq 0$, de donde $||A\mathbf{x}|| \leqslant ||A|| \, ||\mathbf{x}||$, y como esta última desigualdad vale trivialmente para $\mathbf{x} = 0$, el resultado se sigue.

Ejemplo 46 Si en $\mathbb{R}^n$ elegimos la $|| \, ||_\infty$, entonces la norma matricial subordinada viene dada por

$$\begin{aligned} ||A||_\infty &= \sup\{\|A\mathbf{x}\|_\infty \ : \ \|\mathbf{x}\|_\infty = 1\} \\ &= \max\left\{\sum_{j=1}^n |a_{1j}|, \sum_{j=1}^n |a_{2j}|, \ldots, \sum_{j=1}^n |a_{ij}|, \ldots, \sum_{j=1}^n |a_{nj}|\right\} . \end{aligned}$$

Ejemplo 47 Si en $\mathbb{R}^n$ elegimos la $|| \, ||_1$, entonces la norma matricial subordinada viene dada por

$$\begin{aligned} ||A||_1 &= \sup\{\|A\mathbf{x}\|_1 \ : \ \|\mathbf{x}\|_1 = 1\} \\ &= \max\left\{\sum_{i=1}^n |a_{i1}|, \sum_{i=1}^n |a_{i2}|, \ldots, \sum_{i=1}^n |a_{ij}|, \ldots, \sum_{i=1}^n |a_{in}|\right\} . \end{aligned}$$

Una de las normas más usadas en Algebra Lineal es la *norma de Fröbenius* la cual es dada por

$$||A||_F = \left(\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2\right)^{1/2} , \quad A \in M(m \times n, \mathbb{R})$$

y también las p–normas ($p \geqslant 1$), las cuales son dadas por

$$||A||_p = \sup\left\{\frac{||A\mathbf{x}||_p}{||\mathbf{x}||_p} \ : \ \mathbf{x} \neq 0\right\} .$$

Teorema 3.2 *Para cualquier norma matricial subordinada se tiene*

1. $\| \, I \, \| = 1$.
2. $\|AB\| \leqslant \|A\| \cdot \|B\|$.

Demostración. Del teorema (3.1) para todo $\mathbf{x} \in \mathbb{R}^n$ se tiene que

$$||AB\mathbf{x}|| \leqslant ||A|| \cdot ||B\mathbf{x}|| \leqslant ||A|| \cdot ||B|| \, ||\mathbf{x}|| ,$$

de donde el resultado se sigue.

Ejemplo 48 No toda norma matricial es subordinada a alguna norma en $\mathbb{R}^n$. Para verlo, usamos la parte 2 del teorema anterior.

Definamos la norma matricial

$$||A||_\Delta = \max_{i,j=1,\ldots,n} |a_{i,j}| .$$

Afirmamos que esta es una norma matricial (se deja al lector la verificación) y no es subordinada a ninguna norma en $M(n \times n, \mathbb{R})$.

En efecto, tomemos las matrices

$$A = B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} .$$

Tenemos $||A||_\Delta = ||B||_\Delta = 1$ y como

$$AB = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

se tiene que $||AB||_\Delta = 2 > ||A||_\Delta ||B||_\Delta = 1$.

Teorema 3.3 *Para las normas matriciales definidas anteriormente valen las siguientes relaciones. Sea $A \in M(n \times n, \mathbb{R})$, entonces*

1. $||A||_2 \leqslant ||A||_F \leqslant \sqrt{n}\ ||A||_2$
2. $||A||_\Delta \leqslant ||A||_2 \leqslant n\ ||A||_\Delta$
3. $\frac{1}{\sqrt{n}}\ ||A||_\infty \leqslant ||A||_2 \leqslant \sqrt{n}\ ||A||_\infty$
4. $\frac{1}{\sqrt{n}}\ ||A||_1 \leqslant ||A||_2 \leqslant \sqrt{n}\ ||A||_1$.

Teorema 3.4 *Sea $A \in M(n \times n, \mathbb{R})$, con $\| A \| < 1$, entonces $I - A$ es una matriz que tiene inversa, la cual viene dada por $(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$, donde $A^0 = I$. Además, se tiene que $\left\|(I - A)^{-1}\right\| \leqslant \dfrac{1}{1 - \|A\|}$.*

Demostración. Es sólo cálculo y se deja a cargo del lector.

Teorema 3.5 *Sean $A,\ B \in M(n \times n, \mathbb{R})$ tales que $\|I - AB\| < 1$ entonces A y B tienen inversas, las cuales son dadas por $A^{-1} = B\left(\sum_{k=0}^{\infty} (I - AB)^k\right)$ y $B^{-1} = \left(\sum_{k=0}^{\infty} (I - AB)^k\right) A$, respectivamente,.*

Demostración. Directa desde el teorema anterior.

Definición 3.2 *Sea $A \in M(n \times n, \mathbb{R})$, decimos que $\lambda \in \mathbb{C}$ es un valor propio de A si la matriz $A - \lambda I$ no es invertible, en otras palabras, λ es solución de la ecuación polinomial*

$$p_A(\lambda) = \det(A - \lambda I) = 0,$$

donde $p(\lambda)$ *es el polinomio característico de* A. *Se define el espectro de* A *como el conjunto de sus valores propios, es decir,*

$$\sigma(A) = \{\lambda \in \mathbb{C} \ : \ \lambda \text{ valor propio de } A\}.$$

Observación. El polinomio característico de A, $p(\lambda)$, tiene grado menor o igual a n.

Definición 3.3 *Sea* $A \in M(n \times n, \mathbb{R})$, *definimos el radio espectral de* A *por*

$$\rho(A) = \max\{ \ |\lambda| \ : \ \lambda \in \sigma(A)\}.$$

Observación. Geométricamente $\rho(A)$ es el menor radio tal que el círculo centrado en el origen en el plano complejo con radio $\rho(A)$ contiene todos los valores propios de A.

Observación. Si $\lambda = a + ib \in \mathbb{C}$, su *valor absoluto* o *norma* es dado por $|\lambda| = \sqrt{a^2 + b^2}$.

Teorema 3.6 *Se tiene* $\rho(A) = \inf \|A\|$, *donde el ínfimo es tomado sobre todas las normas matriciales definidas sobre el espacio vectorial* $M(n \times n, \mathbb{R})$.

Observación. Desde la definición de las norma matriciales $||\ ||_\infty$ y $||\ ||_1$ vemos que calcularlas para una matriz dada es fácil. Para lo norma $||\ ||_2$ el cálculo no es tan sencillo, pero tenemos el siguiente resultado.

Teorema 3.7 *Sea* $A \in M(n \times n, \mathbb{R})$. *Entonces*

$$||A||_2 = \left(\max\left\{ \ |\lambda| \ : \ \lambda \in \sigma(A^T A)\right\}\right)^{1/2},$$

en otras palabras, $||A||_2$ *es la raíz cuadrada del mayor valor propio de la matriz* $A^T A$.

Corolario 3.1 *Si* $A \in M(n \times n, \mathbb{R})$ *es simétrica, entonces*

$$||A||_2 = \max\{|\lambda| \ : \ \lambda \in \sigma(A)\}.$$

3.2 Número de condición

Consideremos el sistema $A\mathbf{x} = \mathbf{b}$, donde A es una matriz invertible, por lo tanto tenemos que $\mathbf{x}_T = A^{-1}\mathbf{b}$ es la solución exacta.

Caso 1: Se perturba A^{-1} para obtener una nueva matriz B, la solución $\mathbf{x} = A^{-1}\mathbf{b}$ resulta perturbada y se obtiene un vector $\mathbf{x}_A = B\mathbf{b}$ que debería ser una solución aproximada al sistema original. Una pregunta natural ¿es que ocurre con $E(\mathbf{x}_A) = \|\mathbf{x}_T - \mathbf{x}_A\|$? Tenemos

$$\|\mathbf{x}_T - \mathbf{x}_A\| = \|\mathbf{x}_T - B\mathbf{b}\| = \|\mathbf{x}_T - BA\mathbf{x}_T\| = \|(I - BA)\,\mathbf{x}_T\| \leqslant \|I - BA\| \ \|\mathbf{x}_T\|$$

de donde obtenemos que

$$E(\mathbf{x}_A) = \|\mathbf{x}_T - \mathbf{x}_A\| \leqslant \|I - BA\| \ \|\mathbf{x}_T\|$$

y para el error relativo $E_R(\mathbf{x}_A)$, tenemos

$$E_R(\mathbf{x}_A) = \frac{\|\mathbf{x}_T - \mathbf{x}_A\|}{\|\mathbf{x}_T\|} \leqslant \|I - BA\|.$$

Caso 2: Si perturbamos $\mathbf{b}$ para obtener un nuevo vector $\mathbf{b}_A$. Si $\mathbf{x}_T$ satisface $A\mathbf{x} = \mathbf{b}$ y $\mathbf{x}_A$ satisface $A\mathbf{x} = \mathbf{b}_A$. Una pregunta natural es determinar cotas para $E(\mathbf{x}_A) = \|\mathbf{x}_T - \mathbf{x}_A\|$ y $E_R(\mathbf{x}_A) = \frac{\|\mathbf{x}_T - \mathbf{x}_A\|}{\|\mathbf{x}_T\|}$. Tenemos que

$$\|\mathbf{x}_T - \mathbf{x}_A\| = \left\|A^{-1}\mathbf{b} - A^{-1}\mathbf{b}_A\right\| = \left\|A^{-1}(\mathbf{b} - \mathbf{b}_A)\right\| \leqslant \left\|A^{-1}\right\| \ \|\mathbf{b} - \mathbf{b}_A\|,$$

por lo tanto

$$E(\mathbf{x}_A) \leqslant \|A^{-1}\| \ \|\mathbf{b} - \mathbf{b}_A\| \ = \ \|A^{-1}\| \ E(\mathbf{b}_A),$$

esto es,

$$E(\mathbf{x}_A) \leqslant \|\mathbf{b} - \mathbf{b}_A\| \ = \ \left\|A^{-1}\right\| \ E(\mathbf{b}_A), \tag{3.3}$$

y por otro lado se tiene

$$\|\mathbf{x}_T - \mathbf{x}_A\| \leqslant \left\|A^{-1}\right\| \ \|\mathbf{b} - \mathbf{b}_A\| = \left\|A^{-1}\right\| \ \|A\mathbf{x}_T\| \ \frac{\|\mathbf{b} - \mathbf{b}_A\|}{\|\mathbf{b}\|} \leqslant \left\|A^{-1}\right\| \ \|A\| \ \|\mathbf{x}_T\| \ \frac{\|\mathbf{b} - \mathbf{b}_A\|}{\|\mathbf{b}\|}$$

luego

$$E_R(\mathbf{x}_T) = \frac{\|\mathbf{x}_T - \mathbf{x}_A\|}{\|\mathbf{x}_T\|} \leqslant \left\|A^{-1}\right\| \ \|A\| \ \frac{\|\mathbf{b} - \mathbf{b}_A\|}{\|\mathbf{b}\|} = \left\|A^{-1}\right\| \ \|A\| \ \mathrm{E}_R(\mathbf{b}_A).$$

esto es,

$$E_R(\mathbf{x}_A) \leqslant ||A^{-1}||\,||A|| E_R(\mathbf{b}_A\,. \tag{3.4}$$

Definición 3.4 *Sea $A \in M(n \times n, \mathbb{R})$ una matriz invertible, el número condición de A es dado por*

$$\kappa(A) = \|A\| \ \left\|A^{-1}\right\|. \tag{3.5}$$

Observación. Note que $\kappa(A) \geqslant 1$, pues es evidente que $\kappa(I) = 1$ y que

$$1 = ||I|| \leqslant ||A||\ ||A^{-1}|| = \kappa(A).$$

Ejemplo 49 Considere

$$A = \begin{pmatrix} 1 & 1+\varepsilon \\ 1-\varepsilon & 1 \end{pmatrix},$$

con $\varepsilon > 0$, suficientemente pequeño. Tenemos que $\det(A) = \varepsilon^2 \approx 0$ es muy pequeño, lo cual significa que A es "casi singular". Como

$$A^{-1} = \varepsilon^{-2} \begin{pmatrix} 1 & -1-\varepsilon \\ -1+\varepsilon & 1 \end{pmatrix},$$

obtenemos que $\|A\|_\infty = 2+\varepsilon$, $\left\|A^{-1}\right\|_\infty = \varepsilon^{-2}(2+\varepsilon)$, por lo tanto

$$\kappa(A) = \frac{(2+\varepsilon)^2}{\varepsilon^2} \geqslant \frac{4}{\varepsilon}.$$

Observe que si $\varepsilon < 0.0001$, entonces $k(A) \geqslant 40.000$.

Ejemplo 50 Consideremos la matriz

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1.00000001 \end{pmatrix}$$

Tenemos

$$A^{-1} = \begin{pmatrix} 1.0 \times 10^8 & -1.0 \times 10^8 \\ -1.0 \times 10^8 & 1.0 \times 10^8 \end{pmatrix}$$

Luego, $||A||_1 = ||A||_2 \approx ||A||_\infty = 2$, $||A^{-1}||_1 = ||A^{-1}||_\infty \approx ||A^{-1}||_2 \approx 2 \times 10^8$, luego $\kappa_1(A) = \kappa_\infty(A) \approx 4 \times 10^8$.

Observación. Si $\kappa(A)$ es demasiado grande diremos que A está *mal condicionada*.

Ejemplo 51 Consideremos la matriz

$$A = \begin{pmatrix} 1 & -1 & -1 & \cdots & -1 \\ 0 & 1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

La matriz A tiene sólo unos en su diagonal, y sobre ella tiene sólo menos unos. Tenemos que $\det(A) = 1$, pero $\kappa_\infty(A) = n2^{n-1}$, esta matriz está mal condicionada.

Si resolvemos $A\mathbf{x} = b$ numéricamente no obtenemos una solución exacta, $\mathbf{x}_T$, si no una solución aproximada $\mathbf{x}_A$. Una pregunta natural al resolver numéricamente el sistema es ¿qué tan cerca esta $A\mathbf{x}_A$ de $\mathbf{b}$?

Para responder dicha interrogante definamos $\mathbf{r} = \mathbf{b} - A\mathbf{x}_A$ como el *vector residual* y $\mathbf{e} = \mathbf{x}_T - \mathbf{x}_A$ como el *vector de error*.

Observación. Tenemos que

1. $A\mathbf{e} = \mathbf{r}$.
2. $\mathbf{x}_A$ es solución exacta de $A\mathbf{x}_A = \mathbf{b}_A = \mathbf{b} - \mathbf{r}$.

¿Qué relación existe entre $\frac{\|\mathbf{x}_T-\mathbf{x}_A\|}{\|\mathbf{x}_T\|}$ y $\frac{\|\mathbf{b}-\mathbf{b}_A\|}{\|\mathbf{b}\|}=\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$?

Teorema 3.8 *Para toda matriz invertible* $A\in M(n\times n,\mathbb{R})$, *se tiene que*

$$\frac{1}{\kappa(A)}\frac{\|\mathbf{b}-\mathbf{b}_A\|}{\|\mathbf{b}\|}\leqslant\frac{\|\mathbf{x}_T-\mathbf{x}_A\|}{\|\mathbf{x}_T\|}\leqslant\kappa(A)\frac{\|\mathbf{b}-\mathbf{b}_A\|}{\|\mathbf{b}\|},\tag{3.6}$$

es decir,

$$\frac{1}{\kappa(A)}\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}\leqslant\frac{\|\mathbf{e}\|}{\|\mathbf{x}_T\|}\leqslant\kappa(A)\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}.\tag{3.7}$$

Observación. Si tenemos una matriz B escrita en la forma

$$B=A(I+E)$$

donde I denota la matriz identidad y E es una matriz de error, entonces se tienen las siguientes cotas para el error relativo

$$\frac{||\mathbf{x}_T-\mathbf{x}_A||}{||\mathbf{x}_T||}\leqslant\frac{\kappa(A)}{1-\kappa(A)\frac{||AE||}{||A}}\frac{||AE||}{||A||}\tag{3.8}$$

si $||AE||<1$ y

$$\frac{||\mathbf{x}_T-\mathbf{x}_A||}{||\mathbf{x}_T||}\leqslant\frac{||E||}{1-||E||}\tag{3.9}$$

si $||E||<1$.

Observación. Sea $A\in M(n\times n,\mathbb{R})$. Denotemos por $\lambda_{\max}$ y $\lambda_{\min}$, respectivamente, el máximo y el mínimo de los valores propios de AA^T. Entonces $\kappa_2(A)=\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$.

Ejemplo 52 Determine $||A||_2$, $||A^{-1}||_2$ y $\kappa_2(A)$, donde

$$A=\begin{pmatrix}\frac{3}{\sqrt{3}} & -\frac{1}{\sqrt{3}}\\ 0 & \frac{\sqrt{8}}{\sqrt{3}}\end{pmatrix}$$

Tenemos que

$$A^T=\begin{pmatrix}\frac{3}{\sqrt{3}} & 0\\ -\frac{1}{\sqrt{3}} & \frac{\sqrt{8}}{\sqrt{3}}\end{pmatrix}$$

Luego

$$AA^T-\lambda\mathrm{I}=\begin{pmatrix}\frac{10}{3}-\lambda & -\frac{\sqrt{8}}{3}\\ -\frac{\sqrt{8}}{3} & \frac{8}{3}-\lambda\end{pmatrix}.$$

Por lo tanto, $\det(AA^T - \lambda \mathrm{I}) = \lambda^2 - 6\lambda + 8$, de donde los valores propios de AA^T son $\lambda = 2$ y $\lambda = 4$, es decir, $\lambda_{\min} = 2$ y $\lambda_{\max} = 4$. Luego

$$||A||_2 = \sqrt{\lambda_{\max}} = 2 \quad \text{y} \quad ||A^{-1}||_2 = \frac{1}{\sqrt{\lambda_{\min}}} = \frac{1}{\sqrt{2}},$$

y se tiene que $\kappa_2(A) = \frac{2}{\sqrt{2}} = \sqrt{2}$.

3.3 Solución de sistemas de ecuaciones lineales: métodos directos

En esta parte estudiaremos métodos directos, es decir, algebraicos, para resolver sistemas de ecuaciones lineales. Herramienta fundamental aquí es el Álgebra Lineal.

3.3.1 Conceptos básicos

La idea es reducir un sistema de ecuaciones lineales a otro que sea más sencillo de resolver.

Definición 3.5 *Sean $A, B \in M(n \times n, \mathbb{R})$. Decimos que dos sistemas de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$ y $B\mathbf{x} = \mathbf{d}$ son equivalentes si poseen exactamente las mismas soluciones.*

Por lo tanto si podemos reducir un sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$ a un sistema equivalente y más simple $B\mathbf{x} = \mathbf{d}$, podemos resolver este último, y obtenemos de ese modo las soluciones del sistema original.

Definición 3.6 *Sea $A \in M(n \times n, \mathbb{R})$. Llamaremos operaciones elementales por filas sobre A a cada una de las siguientes operaciones*

1. *Intercambio de la fila i con la fila j, denotamos esta operación por $F_i \leftrightarrow F_j$.*
2. *Reemplazar la fila i, F_i, por un múltiplo no nulo λF_i de la fila i, denotamos esta operación por $F_i \mapsto \lambda F_i$*
3. *Reemplazar la fila i, F_i, por la suma de la fila i más un múltiplo no nulo λF_j de la fila j, $i \neq j$, denotamos esta operación por $F_i \mapsto F_i + \lambda F_j$*

Definición 3.7 *Decimos que una matriz A es una matriz elemental si ella se obtiene a partir de la matriz identidad a través de una única operación elemental. Una matriz elemental será denotada por E.*

Ejemplo 53 1. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \; F_3 \leftrightarrow F_2 \; \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$.

2. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \; F_2 \mapsto rF_2 \; \begin{pmatrix} 1 & 0 & 0 \\ 0 & r & 0 \\ 0 & 0 & 1 \end{pmatrix}$

3. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ $F_3 \mapsto F_3 + rF_2$ $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & r & 1 \end{pmatrix}$

Observación. Si a una matriz A se le aplican una serie de operaciones elementales $E_1, E_2, \ldots, E_k$ denotaremos el resultado por $E_k E_{k-1} \cdots E_2 E_1 A$.

Además, si $E_k E_{k-1} \cdots E_2 E_1 A = I$, se tiene que A posee inversa y $E_k E_{k-1} \cdots E_2 E_1 = A^{-1}$.

El teorema siguiente resume las condiciones para que una matriz tenga inversa.

Teorema 3.9 *Sea $A \in M(n \times n, \mathbb{R})$ entonces son equivalentes*

1. *A tiene inversa.*
2. $\det(A) \neq 0$.
3. *Las filas de A forman una base de $\mathbb{R}^n$.*
4. *Las columnas de A forman una base de $\mathbb{R}^n$.*
5. *La transformación lineal $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ dada por $T(\mathbf{x}) = A\mathbf{x}$, asociada a la matriz A, es inyectiva,*
6. *La transformación $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ dada por $T(\mathbf{x}) = A\mathbf{x}$, asociada a la matriz A, es sobreyectiva*
7. *El sistema $A\mathbf{x} = 0$, donde $\mathbf{x} \in \mathbb{R}^n$ posee solución única $\mathbf{x} = 0$.*
8. *Para cada $\mathbf{b} \in \mathbb{R}^n$ existe un único vector $\mathbf{x} \in \mathbb{R}^n$ tal que $A\mathbf{x} = \mathbf{b}$.*
9. *A es producto de matrices elementales.*
10. *Todos los valores propios de A son distinto de cero.*

3.4 Factorización de matrices

En esta sección estudiaremos las condiciones necesarias para que una matriz $A \in M(n \times n, \mathbb{R})$ tenga una factorización de la forma LU donde $L, U \in M(n \times n, \mathbb{R})$, con L una matriz triangular inferior y U una matriz triangular superior.

Observe que si $A \in M(n \times n, \mathbb{R})$ tiene una factorización de la forma LU como arriba, entonces el sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$, con $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$, puede resolverse de una manera más sencilla, para ello consideramos el cambio de variable $U\mathbf{x} = \mathbf{z}$, y resolvemos primero el sistema $L\mathbf{z} = \mathbf{b}$ primero y enseguida resolvemos el sistema $U\mathbf{x} = \mathbf{z}$, obteniendo así la solución del sistema original, es decir,

$$A\mathbf{x} = \mathbf{b} \Longleftrightarrow L\underbrace{U\mathbf{x}}_{\mathbf{z}} = \mathbf{b} \Longleftrightarrow \begin{cases} L\mathbf{z} = \mathbf{b} \\ U\mathbf{x} = \mathbf{z}. \end{cases}$$

Supongamos que $A \in M(n \times n, \mathbb{R})$ tiene una descomposición de la forma LU como arriba, con

$$L = \begin{pmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn-1} & l_{nn} \end{pmatrix} \text{ y } U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n-1} & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n-1} & u_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{nn} \end{pmatrix} \tag{3.10}$$

De esta descomposición tenemos

$$\det(A) = \det(L)\det(U) = \left(\prod_{i=1}^{n} l_{ii}\right)\left(\prod_{i=1}^{n} u_{ii}\right) \tag{3.11}$$

así, $\det(A) \neq 0$ si y sólo si $l_{ii} \neq 0$ y $u_{ii} \neq 0$ para todo $i = 1, \ldots, n$.

Para simplificar la notación, escribamos

$$L = (F_1\ F_2\ \ldots\ F_n)^T \quad \text{y} \quad U = (C_1 C_2\ \cdots\ C_n)\ ,$$

donde $F_i = (l_{i1}\ l_{i2}\ \cdots\ l_{ii}\ 0\ \cdots\ 0)$ y $C_i = (u_{1i}\ u_{2i}\ \cdots\ u_{ii}\ 0\ \cdots\ 0)^T$, son las filas de L y las columnas de U, respectivamente.

Ahora, al desarrollar la multiplicación $A = LU$,

$$\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix} \ (C_1 C_2\ \cdots\ C_n) = \begin{pmatrix} F_1C_1 & F_1C_2 & \cdots & F_1C_n \\ F_2C_1 & F_2C_2 & \cdots & F_2C_n \\ \vdots & \vdots & \ddots & \vdots \\ F_nC_1 & F_nC_2 & \cdots & F_nC_n \end{pmatrix}$$

donde F_iC_j es considerado como el producto de las matrices $F_i = (l_{i1}\ l_{i2}\ \cdots l_{ii}\ 0\ \cdots 0)$ y $C_i = (u_{1i}\ u_{2i}\ \cdots\ u_{ii}\ 0\ \cdots\ 0)^T$. Tenemos entonces

$$\begin{cases} a_{11} &= F_1C_1 = l_{11}u_{11} \\ a_{12} &= F_1C_2 = l_{11}u_{12} \\ \quad\vdots & \vdots\ \ \vdots \\ a_{1n} &= F_1C_n = l_{11}u_{1n} \end{cases} \tag{3.12}$$

de aquí, vemos que si fijamos $l_{11} \neq 0$, podemos resolver las ecuaciones anteriores para u_{1i}, $i = 1, \ldots, n$. Enseguida, para la segunda fila de A tenemos

$$\begin{cases} a_{21} &= F_2C_1 = l_{21}u_{11} \\ a_{22} &= F_2C_2 = l_{21}u_{12} + l_{22}u_{22} \\ \quad\vdots & \vdots\ \ \vdots \\ a_{2n} &= F_2C_n = l_{21}u_{1n} + l_{22}u_{2n} \end{cases} \tag{3.13}$$

como u_{11} ya lo conocemos desde las ecuaciones (3.12), podemos encontrar el valor l_{21}, conocido este valor, fijando un valor no cero para l_{22}, y dado que conocemos los valores u_{1i} para $i = 2, \ldots, n$, podemos encontrar los valores de u_{2i} para $i = 2, \ldots, n$. Siguiendo este método vemos que es posible obtener la descomposición de A en la forma LU. Note que también

podemos comparar las columnas de A con aquellas obtenidas desde el producto LU y proceder a resolver las ecuaciones resultantes. El problema es ahora saber bajo que condiciones dada una matriz A ella puede descomponerse en la forma LU, es decir, siempre y cuando podamos hacer las elecciones anteriores.

Ejemplo 54 Sea $A = \begin{pmatrix} 10 & -3 & 6 \\ 1 & 8 & -2 \\ -2 & 4 & -9 \end{pmatrix}$. Encontremos una descomposición LU para A. Escribamos

$$\begin{pmatrix} 10 & -3 & 6 \\ 1 & 8 & -2 \\ -2 & 4 & -9 \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

desarrollando, obtenemos que

$$\begin{cases} l_{11}u_{11} &= 10 \\ l_{11}u_{12} &= -3 \\ l_{11}u_{13} &= 6 \end{cases}$$

de aquí fijando un valor no cero para l_{11}, digamos $l_{11} = 2$, obtenemos que $u_{11} = 5$, $u_{12} = -\frac{3}{2}$, y $u_{13} = 3$. Enseguida tenemos las ecuaciones

$$\begin{cases} l_{21}u_{11} &= 1 \\ l_{21}u_{12} + l_{22}u_{22} &= 8 \\ l_{21}u_{13} + l_{22}u_{23} &= -2 \end{cases}$$

como $u_{11} = 5$ se tiene que $l_{21} = \frac{1}{5}$. Ahora fijamos el valor de $l_{22} = 1$ y obtenemos los valores $u_{22} = \frac{83}{10}$, y $u_{23} = -\frac{13}{5}$. Finalmente, para la última fila de A, nos queda

$$\begin{cases} l_{31}u_{11} &= -2 \\ l_{31}u_{12} + l_{32}u_{22} &= 4 \\ l_{31}u_{13} + l_{32}u_{23} + l_{33}u_{33} &= -2 \end{cases}$$

como $u_{11} = 5$ de la primera ecuación $l_{31} = -\frac{2}{5}$, reemplazando los valores de $l_{31} = -\frac{2}{5}$, $u_{12} = -\frac{3}{2}$ y $u_{22} = \frac{83}{10}$ en la segunda ecuación encontramos que $l_{32} = \frac{34}{83}$. Ahora, reemplazando los valores ya obtenidos en la tercera ecuación obtenemos que $l_{33}u_{33} = -\frac{1694}{415}$, para encontrar el valor de u_{33} podemos fijar el valor de l_{33}, digamos $l_{33} = -\frac{1}{415}$, y obtenemos que $u_{33} = 1604$. Tenenmos así una descomposición LU para la matriz A, es decir,

$$\begin{pmatrix} 10 & -3 & 6 \\ 1 & 8 & -2 \\ -2 & 4 & -9 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ \frac{1}{5} & 1 & 0 \\ -\frac{3}{5} & \frac{34}{83} & -\frac{1}{415} \end{pmatrix} \begin{pmatrix} 5 & -\frac{3}{2} & 3 \\ 0 & \frac{83}{10} & -\frac{13}{5} \\ 0 & 0 & 1604 \end{pmatrix}.$$

No siempre es posible encontrar una descomposición de la forma LU para una matriz A dada, esto se muestra en el siguiente ejemplo.

Ejemplo 55 La matriz $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ no tiene descomposición de la forma LU. Notemos que $\det(A) \neq 0$.

Para verlo supongamos que podemos escribir $A = LU$, donde

$$L = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \quad \text{y} \quad U = \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix}.$$

Tenemos entonces que

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix}$$

de donde $l_{11}u_{11} = 0$, por lo tanto

$$l_{11} = 0 \ \ (*) \quad \text{o} \quad u_{11} = 0 \ \ (**).$$

Como $l_{11}u_{12} = 1$, no es posible que (*) sea verdadero, es decir, debemos tener que $l_{11} \neq 0$, en consecuencia se debe tener que $u_{11} = 0$, pero además se tiene que $u_{12} \neq 0$. Ahoram, como $l_{21}u_{11} = 1$, llegamos a una contradicción. Luego tal descomposición para A no puede existir.

En el algoritmo que vimos para buscar la descomposición de una matriz A en la forma LU, en los sucesivos paso fue necesario fijar valores no cero para los coeficientes l_{ii} de L, de modo a poder resolver las ecuaciones resultantes. También, si procedemos con el algoritmo a comparar las columnas de la matriz producto LU con las columnas de A, vemos que será necesario fijar en paso sucesivos valores no cero para los coeficientes u_{ii} de U, de modo a poder resolver las ecuaciones resultantes. Como es evidente existe una manera simple de hacer esto, por ejemplo si en el primer algoritmo fijamos los valores $l_{ii} = 1$ para $i = 1, \ldots, n$, decimos que tenemos una *descomposición LU de Doolitle* para A, y si en el segundo algoritmo fijamos los valores $u_{ii} = 1$ para $i = 1, \ldots, n$, decimos que tenemos una *descomposición LU de Crout* para A, finalmente si fijamos $U = L^T$ (transpuesta de L) se tiene que $l_{ii} = u_{ii}$, y decimos que tenemos una *descomposición LU de Cholesky* para A. Note que para la existencia de descomposición de Cholesky es necesario que A sea simétrica, pues si $A = LL^T$ entonces $A^T = A$, esto significa que es una condición necesario, pero una no es una condición suficiente.

Sobre la existencia de descomposición LU para una matriz A veremos a continuación algunos teoremas. Primero recordemos que si

$$A = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}_{n \times n}$$

entonces los menores principales de A son las submatrices A_k, $k = 1, \ldots, n$, definidas por

$$A_1 = (a_{11})_{1\times 1}\,,\; A_2 = \begin{pmatrix} a_{11} & a_{21} \\ a_{21} & a_{22} \end{pmatrix}_{2\times 2}\,,\ldots,\; A_k = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}_{k\times k},$$

$\ldots\,,\; A_n = A\,.$

Teorema 3.10 *Si los n menores principales de una matriz A tienen determinante distinto de cero, entonces A tiene una descomposición LU .*

Para tener descomposición de Cholesky, recordemos algunos conceptos y resultados.

Definición 3.8 *Sea $A \in M(n\times n,\mathbb{R})$. Decimos que A es definida positiva si $\langle A\mathbf{x},\mathbf{x}\rangle > 0$ para todo $\mathbf{x}\in\mathbb{R}^n$, con $\mathbf{x}\neq 0$.*

Ejemplo 56 La matriz

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

es definida positiva. En efecto considere $\mathbf{x} = (x\ y\ z)^T$, tenemos

$$\langle \mathbf{x}, A\mathbf{x}\rangle = (x,y,z)\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2x^2 - 2xy + 2y^2 - 2yz + 2z^2.$$

Luego,

$$\begin{aligned} \langle \mathbf{x}, A\mathbf{x}\rangle &= 2x^2 - 2xy + 2y^2 - 2yz + 2z^2 \\ &= x^2 + (x-y)^2 + (y-z)^2 + z^2 > 0, \end{aligned}$$

para todo $(x,y,z)\neq(0,0,0)$.

Teorema 3.11 *Si $A \in M(n\times n,\mathbb{R})$ es definida positiva entonces todos sus valores propios son reales y positivos.*

Observación. Si $A \in M(n\times n,\mathbb{R})$ entonces podemos descomponer A de manera única como la suma de una matriz simétrica, $A_0 = \frac{1}{2}\left(A + A^T\right)$, y una matriz antisimétrica, $A_1 = \frac{1}{2}\left(A - A^T\right)$. Desde la definición de matriz positiva definida se tiene que $\langle A\mathbf{x},\mathbf{x}\rangle = \langle A_0\mathbf{x},\mathbf{x}\rangle$, por lo cual sólo necesitamos considerar matrices simétricas.

Teorema 3.12 *Sea $A \in M(n\times n,\mathbb{R})$, entonces A es positiva definida si y sólo si todos los menores principales de A tiene determinante positivo.*

Teorema 3.13 *Sea $A \in M(n\times n,\mathbb{R})$, entonces A simétrica y positiva definida si y sólo si A tiene una descomposición de Cholesky.*

Ejemplo 57 La matriz

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

es definida positiva, pues se tiene que $\det(A_1) = 2 > 0$, $\det(A_2) = \det \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} = 3 > 0$ y $\det(A_3) = \det(A) = 4 > 0$. Luego tiene descomposición de Cholesky.

Ahora, para encontrar la descomposición de Cholesky para esta matriz procedemos como sigue.

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix}$$

Luego, $l_{11}^2 = 2$, es decir, $l_{11} = \sqrt{2}$; $-1 = l_{11}l_{21}$, de donde $l_{21} = -\frac{\sqrt{2}}{2}$; $0 = l_{11}l_{31}$, de donde $l_{31} = 0$; $l_{21}^2 + l_{22}^2 = 2$, de donde $l_{22} = \frac{\sqrt{5}}{\sqrt{2}}$; $l_{21}l_{31} + l_{22}l_{32} = -1$, de donde $l_{32} = -\frac{\sqrt{2}}{\sqrt{5}}$; $l_{31}^2 + l_{32}^2 + l_{33}^2 = 2$, de donde $l_{33} = \frac{2\sqrt{2}}{\sqrt{5}}$.

Por lo tanto

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{5}}{\sqrt{2}} & 0 \\ 0 & -\frac{\sqrt{2}}{\sqrt{5}} & \frac{2\sqrt{2}}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & -\frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{5}}{\sqrt{2}} & -\frac{\sqrt{2}}{\sqrt{5}} \\ 0 & 0 & \frac{2\sqrt{2}}{\sqrt{5}} \end{pmatrix}.$$

3.5 Método de eliminación gaussiana

En esta sección estudiaremos el método de eliminación gaussiana para resolver sistemas de ecuaciones lineales

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

Colocamos primero esta información en la forma

$$\left(\begin{array}{c|c} A & b \end{array} \right) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array} \right) \begin{matrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{matrix}$$

esto es, consideramos la matriz aumentada del sistema. En la notación anterior, los símbolos F_i denotan las filas de la nueva matriz.

Si $a_{11} \neq 0$, el primer paso de la eliminación gaussiana consiste en realizar las operaciones elementales

$$(F_j - m_{j1}F_1) \to F_j, \text{ donde } m_{j1} = \frac{a_{j1}}{a_{11}}, \quad j = 2, 3, ..., n. \tag{3.14}$$

Estas operaciones transforman el sistema en otro en el cual todos los componentes de la primera columna situada bajo la diagonal son cero.

Podemos ver desde otro punto de vista el sistema de operaciones. Esto lo logramos simultáneamente al multiplicar por la izquierda la matriz original A por la matriz

$$M^{(1)} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -m_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -m_{n1} & 0 & \cdots & 1 \end{pmatrix} \tag{3.15}$$

Esta matriz recibe el nombre de primera matriz gaussiana de transformación. El producto de $M^{(1)}$ con la matriz $A^{(1)} = A$ lo denotamos por $A^{(2)}$ y el producto de $M^{(1)}$ con la matriz $b^{(1)} = b$ lo denotamos por $b^{(2)}$, con lo cual obtenemos

$$A^{(2)}\mathbf{x} = M^{(1)}A\mathbf{x} = M^{(1)}b = b^{(2)} \tag{3.16}$$

De manera análoga podemos construir la matriz $M^{(2)}$, en la cual si $a_{22}^{(2)} \neq 0$ los elementos situados bajo de la diagonal en la segunda columna con los elementos negativos de los multiplicadores

$$m_{j2} = \frac{a_{j2}^{(2)}}{a_{22}^{(2)}}, \quad j = 3, 4, \ldots, n \tag{3.17}$$

Así obtenemos

$$A^{(3)}\mathbf{x} = M^{(2)}A^{(2)}\mathbf{x} = M^{(2)}M^{(1)}A\mathbf{x} = M^{(2)}M^{(1)}b = b^{(3)} \tag{3.18}$$

En general, con $A^{(k)}\mathbf{x} = b^{(k)}$ ya formada, multiplicamos por la k–ésima matriz de transformación gaussiana

$$M^{(k)} = \begin{pmatrix} 1 & 0 & & \cdots & & \cdots & 0 \\ 0 & 1 & & & & & \\ & \ddots & 1 & \ddots & & & \vdots \\ & & 0 & & & & \\ \vdots & & & -m_{k+1\,k} & \ddots & & \vdots \\ & & \vdots & \vdots & 0 & & 0 \\ 0 & \cdots & 0 & -m_{n\,k} & 0 & 0 & 1 \end{pmatrix}$$

para obtener

$$A^{(k+1)}\mathbf{x} = M^{(k)}A^{(k)}\mathbf{x} = M^{(k)}\cdots M^{(1)}A\mathbf{x} = M^{(k)}b^{(k)} = b^{(k+1)} = M^{(k)}\cdots M^{(1)}b \qquad (3.19)$$

Este proceso termina con la formación de un sistema $A^{(n)}\mathbf{x} = b^{(n)}$, donde $A^{(n)}$ es una matriz triangular superior

$$A^{(n)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ \vdots & \vdots & \ddots & a_{n-1\ n}^{(n-1)} \\ 0 & \cdots & 0 & a_{nn}^{(n)} \end{pmatrix}$$

dada por

$$A^{(n)} = M^{(n-1)}M^{(n-2)}M^{(n-3)}\cdots M^{(1)}A. \qquad (3.20)$$

El proceso que hemos realizado sólo forma la mitad de la factorización matricial $A = LU$, donde con U denotamos la matriz triangular superior $A^{(n)}$. Si queremos determinar la matriz triangular inferior L, primero debemos recordar la multiplicación de $A^{(k)}\mathbf{x} = b^{(k)}$ mediante la transformación gaussiana $M^{(k)}$ con que obtuvimos

$$A^{(k+1)}\mathbf{x} = M^{(k)}A^{(k)}\mathbf{x} = M^{(k)}b^{(k)} = b^{(k+1)}$$

para poder revertir los efectos de esta transformación debemos considerar primero la matriz $L^{(k)} = \left(M^{(k)}\right)^{-1}$, la cual esta dada por

$$L^{(k)} = \left(M^{(k)}\right)^{-1} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & & \cdots & 0 \\ 0 & 1 & & & & & \vdots \\ \vdots & & 1 & & \vdots & & \\ & \vdots & 0 & \ddots & & & \vdots \\ \vdots & & & m_{k+1\ k} & \ddots & & \\ & \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & m_{n\ k} & 0 & 0 & 1 \end{pmatrix}.$$

Por lo tanto si denotamos

$$L = L^{(1)}L^{(2)}\cdots L^{(n-1)}$$

obtenemos

$$LU = L^{(1)}L^{(2)}\cdots L^{(n-1)}M^{(n-1)}M^{(n-2)}M^{(n-3)}\cdots M^{(1)} = A.$$

Con lo cual obtenemos el siguiente resultado.

Teorema 3.14 *Si podemos efectuar la eliminación gaussiana en el sistema lineal* $A\mathbf{x} = b$ *sin intercambio de filas, entonces podemos factorizar la matriz* A *como el producto de una matriz triangular inferior* L *y una matriz triangular superior* U*, donde*

$$U = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \ddots & \vdots \\ \vdots & \vdots & \ddots & a_{n-1\ n}^{(n-1)} \\ 0 & \cdots & 0 & a_{nn}^{(n)} \end{pmatrix} \quad y \quad L = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ m_{21} & 1 & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ m_{n1} & \cdots & m_{n\ n-1} & 1 \end{pmatrix}.$$

Ejemplo 58 Consideremos el siguiente sistema lineal

$$\begin{aligned} x + y + 3w &= 4 \\ 2x + y - z + w &= 1 \\ 3x - y - z + w &= -3 \\ -x + 2y + 3z - w &= 4 \end{aligned}$$

Si realizamos las siguientes operaciones elementales $F_2 \mapsto (FE_2 - 2F_1)$, $F_3 \mapsto (F_3 - 3F_1)$, $F_4 \mapsto (F_4 - (-1)\,F_1)$, $F_3 \mapsto (F_3 - 4F_2)$, $F_4 \mapsto (F_4 - (-3)\,F_2)$ con lo que obtenemos el sistema triangular

$$\begin{aligned} x + y + 3w &= 4 \\ -y - z - 5w &= -7 \\ 3z + 13w &= 13 \\ -13w &= -13 \end{aligned}$$

Por lo tanto la factorización LU está dada por

$$A = \begin{pmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix}}_{U} = LU.$$

Esta factorización nos permite resolver fácilmente todo sistema que posee como matriz asociada la matriz A. Así, por ejemplo, para resolver el sistema

$$A\mathbf{x} = LU\mathbf{x} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \\ 14 \\ -7 \end{pmatrix}$$

primero realizaremos la sustitución $\mathbf{y} = U\mathbf{x}$. Luego resolvemos $L\mathbf{y} = \mathbf{b}$, es decir,

$$LU\mathbf{x} = L\mathbf{y} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \\ 14 \\ -7 \end{pmatrix}.$$

Este sistema se resuelve para $\mathbf{y}$ mediante un simple proceso de susbtitución hacia adelante, con lo cual obtenemos

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 8 \\ -9 \\ 26 \\ -26 \end{pmatrix}.$$

Por ultimo resolvemos $U\mathbf{x} = \mathbf{y}$, es decir, resolvemos el sistema por un proceso de susbtitución hacia atrás

$$U\mathbf{x} = \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 8 \\ -9 \\ 26 \\ -26 \end{pmatrix}$$

el cual tiene por solución

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 0 \\ 2 \end{pmatrix}.$$

que es la solución buscada.

3.6 Eliminación gaussiana con pivoteo

Al resolver un sistema lineal $A\mathbf{x} = \mathbf{b}$, a veces es necesario intercambiar las ecuaciones, ya sea porque la elimninación gaussiana no puede efectuarse o porque las soluciones que se obtienen no corresponden a as soluciones exactas del sistema. Este proceso es llamado *pivoteo* y nos lleva a la descomposición $PA = LU$ del sistema. Así el sistema se transforma en $LU\mathbf{x} = P\mathbf{b}$ y resolvemos entonces los sistemas

$$\begin{cases} U\mathbf{x} &= \mathbf{z} \\ L\mathbf{z} &= P\mathbf{b} \end{cases}$$

donde P es una *matriz de permutación*, es decir, es una matriz obtenida a partir de la matriz identidad intercambiando algunas de sus filas. Más precisamente.

Definición 3.9 *Una matriz de permutación $P \in M(n \times n, \mathbb{R})$ es una matriz obtenida desde la matriz identidad por permutación de sus filas.*

Observación. Las matrices de permutación poseen dos propiedades de gran utilidad que se relacionan con la eliminación gaussiana. La primera de ellas es que al multiplicar por la izquierda una A por una matriz de permutación P, es decir, al realizar el producto PA se permutan las filas de A. La segunda propiedad establece que si P es una matriz de permutación, entonces $P^{-1} = P^T$.

Si $A \in M(n \times n, \mathbb{R})$ es una matriz invertible entonces podemos resolver el sistema lineal $A\mathbf{x} = \mathbf{b}$ vía eliminación gaussiana, sin excluir el intercambio de filas. Si conociéramos los intercambios que se requieren para resolver el sistema mediante eliminación gaussiana, podríamos arreglar las ecuaciones originales de manera que nos garantice que no se requieren intercambios de filas. Por lo tanto, existe un arreglo de ecuaciones en el sistema que nos permite resolver el sistema con eliminación gaussiana sin intercambio de filas. Con lo que podemos concluir que si $A \in M(n \times n, \mathbb{R})$ es una matriz invertible entonces existe una matriz de permutación P para la cual podemos resolver el sistema $PA\mathbf{x} = P\mathbf{b}$, sin hacer intercambios de filas. Pero podemos factorizar la matriz PA en $PA = LU$, donde L es una triangular inferior y U triangular superior, y P es una matriz de permutación. Dado que $P^{-1} = P^T$ obtenemos que $A = \left(P^T L\right) U$. Sin embargo, la matriz $P^T L$ no es triangular inferior, salvo que $P = I$.

Supongamos que tenemos el sistema

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{3.21}$$

Se define la *escala* de cada fila como

$$s_i = \max\{|a_{i1}|, |a_{i2}|, \ldots, |a_{in}|\}, \qquad 1 \leqslant i \leqslant n. \tag{3.22}$$

Elección de la fila pivote. Elegimos como *fila povote* la fila para la cual se tiene que

$$\frac{|a_{i_0 1}|}{s_{i_0}} \tag{3.23}$$

es el mayor.

Intercambiamos en A la fila 1 con la fila i_0, esto es, realizamos la operación elemental $F_1 \leftrightarrow F_{i_0}$. Esto nos produce la primera permutación. Procedemos ahora a producir ceros bajo la primera columna de la matriz permutada, tal como se hizo en la eliminación gaussiana. Denotamos el índice que hemos elegido por p_1, este se convierte en la primera componente de la permutación.

Más especificamente, comenzamos por fijar

$$p = (p_1\, p_2\, \ldots\, p_n) = (1\, 2\, \ldots\, n) \tag{3.24}$$

como la permutación antes de comenzar el pivoteo. Así la primera permutación que hemos hecho es

$$\begin{pmatrix} 1 & 2 & \ldots & i_0 & \ldots & n \\ i_0 & 2 & \ldots & 1 & \ldots & n \end{pmatrix}.$$

Para fijar las ideas, comenzamos con la permutación $(p_1\, p_2\, \ldots\, p_n) = (1\ 2\ \ldots\ n)$. Primero seleccionamos un índice j para el cual se tiene

$$\frac{|a_{p_j j}|}{s_{p_j}}, \quad 2 \leqslant j \leqslant n \tag{3.25}$$

es el mayor, y se intercambia p_1 con p_j en la permutación p. Procedemos a la elmincación gaussiana, restando

$$\frac{a_{p_i 1}}{a_{p_1 1}} \times F_{p_1}$$

a las filas p_i, $2 \leqslant i \leqslant n$.

En general, supongamos que tenemos que producir ceros en la columna k. Impeccionamos los números

$$\frac{|a_{p_i k}|}{s_{p_i}}, \quad k \leqslant i \leqslant n \tag{3.26}$$

y buscamos el mayor de ellos. Si j es el índice del primer cuociente más grande, intercambiamos p_k con p_j en la permutación p y las filas F_{p_k} con la fila F_{p_j} en la matriz que tenemos en esta etapa de la eliminación gaussiana–pivoteo. Enseguida producimos ceros en la columna p_k bajo el elemento $a_{p_k k}$, para ellos restamos

$$\frac{a_{p_i k}}{a_{p_k k}} \times F_{p_k}$$

de la fila F_{p_i}, $k+1 \leqslant i \leqslant n$, y continuamos el proceso hasta obtener una matriz triangular superior.

Veamos con un ejemplo específico como podemos ir guardando toda la información del proceso que vamos realizando en cada etapa del proceso de eliminación gaussiana–pivoteo

Ejemplo 59 Resolver el sistema de ecuaciones lineales

$$\begin{pmatrix} 2 & 3 & -6 \\ 1 & -6 & 8 \\ 3 & -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Solución. Primero calculemos la escala de las filas. Tenemos

$$\begin{aligned} s_1 &= \max\{|2|, |3|, |-6|\} = 6 \\ s_2 &= \max\{|1|, |-6|, |8|\} = 8 \\ s_3 &= \max\{|3|, |-2|, |1|\} = 3 \end{aligned}$$

permutación inicial $p = (1\ 2\ 3)$. Denotemos por s el vector de la escala de las filas, es decir, $s = (6\ 8\ 3)$.

Elección de la primera fila pivote:

$$\begin{aligned} \frac{|a_{11}|}{s_1} &= \frac{2}{6} = \frac{1}{3} \\ \frac{|a_{21}|}{s_2} &= \frac{1}{8} \\ \frac{|a_{31}|}{s_3} &= 1\,, \end{aligned}$$

luego $j = 3$. Antes de proceder a intercambiar la fila F_1 con la fila F_3 guardamos la información en la forma siguiente

$$\left(\begin{array}{ccc|c||c|c} 2 & 3 & -6 & 1 & 6 & 1 \\ 1 & -6 & 8 & 1 & 8 & 2 \\ 3 & -2 & 1 & 1 & 3 & 3 \end{array}\right)$$

donde la primera parte de esta matriz representa a la matriz A, la segunda al vector b, la tercera al vector s y la última a la permutación inicial $p = (1\,2\,3)$. Como $j = 3$ intercambianos las filas $F_1 \leftrightarrow F_3$ y obtenemos

$$\left(\begin{array}{ccc|c||c|c} 3 & -2 & 1 & 1 & 3 & 3 \\ 1 & -6 & 8 & 1 & 8 & 2 \\ 2 & 3 & -6 & 1 & 6 & 1 \end{array}\right)$$

en la parte $(A|\mathbf{b})$ de la matriz arriba procedemos a hacer eliminación gaussiana. Los multiplicadores son $m_{21} = \frac{1}{3}$ y $m_{31} = \frac{2}{3}$, obtenemos así

$$\left(\begin{array}{ccc|c||c|c} 3 & -2 & 1 & 1 & 3 & 3 \\ 0 & -\frac{16}{3} & \frac{23}{3} & \frac{2}{3} & 8 & 2 \\ 0 & \frac{13}{3} & -\frac{20}{3} & \frac{1}{3} & 6 & 1 \end{array}\right)$$

Agregamos la información de los multiplicadores a esta estructura como sigue

$$\left(\begin{array}{ccc|c||c|c|c} 3 & -2 & 1 & 1 & 3 & 3 & \\ 0 & -\frac{16}{3} & \frac{23}{3} & \frac{2}{3} & 8 & 2 & \frac{1}{3} \\ 0 & \frac{13}{3} & -\frac{20}{3} & \frac{1}{3} & 6 & 1 & \frac{2}{3} \end{array}\right)$$

Ahora determinamos la nueva fila pivote. Recuerde que la primera fila de esta nueva matriz permanece inalterada, y sólo debemos trabajar con las filas restantes. Para determinar la nueva fila pivote, calculamos

$$\begin{aligned}\frac{|a_{p_2}2|}{s_{p_2}} &= \frac{\frac{16}{3}}{8} = \frac{16}{24} = \frac{2}{3}\\ \frac{|a_{p_3}2|}{s_{p_3}} &= \frac{\frac{13}{3}}{6} = \frac{13}{18}\end{aligned}$$

como $\frac{13}{18} > \frac{2}{3}$ la nueva fila pivote es la fila 3. Intercambiando $F_2 \leftrightarrow F_3$ nos queda

$$\left(\begin{array}{ccc|c||c|c|c|}
3 & -2 & 1 & 1 & 3 & 3 & \\
0 & \frac{13}{3} & -\frac{20}{3} & \frac{1}{3} & 6 & 1 & \frac{2}{3} \\
0 & -\frac{16}{3} & \frac{23}{3} & \frac{2}{3} & 8 & 2 & \frac{1}{3}
\end{array}\right)$$

El multiplicador es $m_{32} = \frac{-\frac{16}{3}}{\frac{13}{3}} = -\frac{16}{13}$. Agregando la información del nuevo multiplicador y aplicando eliminación gaussiana, nos queda

$$\left(\begin{array}{ccc|c||c|c|c|}
3 & -2 & 1 & 1 & 3 & 3 & \\
0 & \frac{13}{3} & -\frac{20}{3} & \frac{1}{3} & 6 & 1 & \frac{2}{3} \\
0 & 0 & -\frac{7}{13} & \frac{94}{117} & 8 & 2 & \frac{1}{3} & -\frac{16}{13}
\end{array}\right)$$

De esto, el sistema a resolver es

$$\underbrace{\begin{pmatrix} 3 & -2 & 1 \\ 0 & \frac{13}{3} & -\frac{20}{3} \\ 0 & 0 & -\frac{7}{13} \end{pmatrix}}_{U} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{3} \\ \frac{94}{117} \end{pmatrix}$$

La matriz L es

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{2}{3} & 1 & 0 \\ \frac{1}{3} & -\frac{16}{13} & 1 \end{pmatrix}$$

y la matriz de permutación P es

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Ahora es fácil verificar que $PA = LU$.

Ejemplo 60 Consideremos la matriz

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

puesto que $a_{11} = 0$ la matriz no posee una factorización LU. Pero si realizamos el intercambio de filas $F_1 \leftrightarrow F_2$, seguido de $F_3 \mapsto (F_3 - F_1)$ y de $F_4 \mapsto (F_4 - F_1)$ obtenemos

$$\begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Realizando las operaciones elementales $F_2 \leftrightarrow F_4$ y $F_4 \mapsto (F_4 + F_3)$, obtenemos

$$U = \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

La matriz de permutación asociada al cambio de filas $F_1 \leftrightarrow F_2$ seguida del intercambio de filas $F_2 \leftrightarrow F_4$ es

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

La eliminación gaussiana en PA se puede realizar sin intercambio de filas usando las operaciones $F_2 \mapsto (F_2 - F_1)$, $F_3 \mapsto (F_3 - F_1)$ y $(F_4 + F_3) \mapsto F_4$, esto produce la factorización LU de PA

$$PA = \begin{pmatrix} 1 & 1 & -1 & 2 \\ 1 & 2 & -1 & 3 \\ 1 & 1 & 0 & 3 \\ 0 & 0 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} = LU.$$

Al multiplicar por $P^{-1} = P^T$ obtenemos la factorización

$$A = \left(P^T L\right) U = \begin{pmatrix} 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

3.7 Matrices especiales

En esta sección nos ocuparemos de clases de matrices en las cuales podemos realizar la eliminación gaussiana sin intercambio de filas.

Decimos que una matriz $A \in M(n \times n, \mathbb{R})$ es *diagonal dominante* si

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$$

para todo $i = 1,\ 2, \ldots,\ n$.

Ejemplo 61 Consideremos las matrices

$$A = \begin{pmatrix} 7 & 2 & 0 \\ 3 & 5 & -1 \\ 0 & 5 & -6 \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} 6 & 4 & -3 \\ 4 & -2 & 0 \\ -3 & 0 & 1 \end{pmatrix}.$$

La matriz A es diagonal dominante y la matriz B no es diagonal dominante, pues por ejemplo $|6| < |4| + |-3| = 7$.

Teorema 3.15 *Toda matriz $A \in M(n \times n, \mathbb{R})$ diagonal dominante tiene inversa, más aun, en este caso podemos realizar la eliminación gaussiana de cualquier sistema lineal de la forma $A\mathbf{x} = \mathbf{b}$ para obtener su solución única sin intercambio de filas, y los cálculos son estables respecto al crecimiento de los errores de redondeo.*

Teorema 3.16 *Si $A \in M(n \times n, \mathbb{R})$ es una matriz definida positiva entonces*

1. *A tiene inversa.*
2. *$a_{ii} > 0$, para todo $i = 1, 2, \ldots, n$.*
3. *$\max_{1 \leqslant k,\, j \leqslant n} |a_{kj}| \leqslant \max_{1 \leqslant i \leqslant n} |a_{ii}|$.*
4. *$(a_{ij})^2 < a_{ii}\, a_{jj}$, para todo $i \neq j$*

Teorema 3.17 *Una matriz simétrica A es definida positiva si y sólo si la eliminación gaussiana sin intercambio de filas puede efectuarse en el sistema $A\mathbf{x} = \mathbf{b}$ con todos los elementos pivotes positivos. Además, en este caso los cálculos son estables respecto al crecimiento de los errores de redondeo.*

Corolario 3.2 *Una matriz simétrica A es definida positiva si y sólo si A puede factorizarse en la forma LDL^T, donde L es una matriz triangular inferior con unos en su diagonal y D es una matriz diagonal con elementos positivos a lo largo de la diagonal.*

Corolario 3.3 *Una matriz simétrica A es definida positiva si y sólo si A puede factorizarse de la forma LL^T, donde L es una matriz triangular inferior con elementos distintos de cero en su diagonal.*

Corolario 3.4 *Sea $A \in M(n \times n, \mathbb{R})$ una matriz simétrica a la cual puede aplicarse la eliminación gaussiana sin intercambio de filas. Entonces, A puede factorizarse en LDL^T, donde L es una matriz triangular inferior con unos en su diagonal y donde D es una matriz diagonal con $a_{11}^{(1)}, \ldots, a_{nn}^{(n)}$ en su diagonal.*

Ejemplo 62 La matriz

$$A = \begin{pmatrix} 4 & -1 & 1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{pmatrix}$$

es definida positiva. La factorización LDL^T de la matriz A es

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0.25 & 0.75 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -0.25 & 0.25 \\ 0 & 1 & 0.75 \\ 0 & 0 & 1 \end{pmatrix}$$

mientras que su descomposición de Cholesky es

$$A = \begin{pmatrix} 2 & 0 & 0 \\ -0.5 & 2 & 0 \\ 0.5 & 1.5 & 1 \end{pmatrix} \begin{pmatrix} 2 & -0.5 & 0.5 \\ 0 & 2 & 1.5 \\ 0 & 0 & 1 \end{pmatrix}.$$

Definición 3.10 *Una matriz $A \in M(n \times n, \mathbb{R})$ es llamada matriz banda si existen enteros p y q con $1 < p, q < n$, tales que $a_{ij} = 0$ siempre que $i + p \leqslant j$ o $j + q \leqslant i$. El ancho de la banda de este tipo de matrices esta dado por $w = p + q - 1$.*

Ejemplo 63 La matriz $A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$ es una matriz de banda con $p = q = 2$ y con ancho de banda 3.

Definición 3.11 *Una matriz $A \in M(n \times n, \mathbb{R})$ se denomina tridiagonal si es una matriz de banda con $p = q = 2$.*

Observación. Los algoritmos de factorización pueden simplificarse considerablemente en el caso de las matrices de banda.

Para ilustrar lo anterior, supongamos que podemos factorizar una matriz tridiagonal A en las matrices triangulares L y U. Ya que A tiene sólo $3n-2$ elementos distintos de cero, habrá apenas $3n-2$ condiciones aplicables para determinar los elementos de L y U, naturalmente a condición de que también se obtengan los elementos cero de A. Supongamos que podemos encontrar las matrices de la forma

$$L = \begin{pmatrix} l_{11} & 0 & \cdots & \cdots & 0 \\ l_{21} & l_{22} & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & l_{n\,n-1} & l_{nn} \end{pmatrix} \text{ y } U = \begin{pmatrix} 1 & u_{12} & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & u_{n-1\,n} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

De la multiplicación que incluye $A = LU$, obtenemos

$$\begin{aligned}
&a_{11} \quad = l_{11};\\
&a_{i\ i-1} = l_{i\ i-1} \ \text{ para cada } \ i = 2, 3, \ldots, n;\\
&a_{ii} \quad = l_{i\ i-1}u_{i-1\ i} \ + \ l_{ii} \ \text{ para cada } \ i = 2, 3, \ldots, n;\\
&a_{i\ i+1} \ = l_{ii}u_{i\ i+1} \ \text{ para cada } \ i = 1, 2, 3, \ldots, n-1;
\end{aligned}$$

3.8 Solución de sistemas de ecuaciones lineales: métodos iterativos

En general, puede resultar muy engorroso encontrar la solución exacta a un sistema de ecuaciones lineales, y por otra parte, a veces nos basta con tener buenas aproximaciones a dicha solución. Para obtener estas últimas usamos métodos iterativos de punto fijo, que en este caso particular resultan ser más analizar su convergencia.

Teorema 3.18 *Consideremos un método iterativo*

$$\mathbf{x}^{(k+1)} = G\mathbf{x}^{(k)} + \mathbf{c} \tag{3.27}$$

donde $G \in M(n\times n, \mathbb{R})$, $\mathbf{x}, \mathbf{c} \in \mathbb{R}^n$. *Entonces el método iterativo es convergente, para cualquier condición inicial* $\mathbf{x}^{(0)}$ *elegida arbitrariamente si y sólo si* $\rho(G) < 1$. *Además, si existe una norma* $|||\ ||$ *en* $M(n \times n, \mathbb{R})$, *en la cual se tiene* $||G|| < 1$, *entonces el método iterativo converge cualesquiera que sea la condición inicial* $\mathbf{x}^{(0)} \in \mathbb{R}^n$ *dada. Si tomamos* $\mathbf{x}^{(k)}$ *como una aproximación al punto fijo* $\mathbf{x}_T$ *del método iterativo (3.27), entonces valen las desigualdades siguientes,*

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leqslant \frac{\rho(G)^k}{1-\rho(G)} \left\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\right\|. \tag{3.28}$$

y

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leqslant \frac{\lambda^k}{1-\lambda} \left\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\right\|, \tag{3.29}$$

donde $\lambda = ||G||$.

Observación. Si el método iterativo $\mathbf{x}^{(k+1)} = G\mathbf{x}^{(k)} + \mathbf{c}$ es convergente y tiene por límite a $\mathbf{x} \in \mathbb{R}^n$, entonces

$$\mathbf{x} = \lim_{k\to\infty} \mathbf{x}^{(k+1)} = G\left(\lim_{k\to\infty} \mathbf{x}^{(k)}\right) + \mathbf{c} = G\mathbf{x} + \mathbf{c},$$

de donde $\mathbf{x} = (I - G)^{-1}\mathbf{c}$.

Regresemos al problema inicial de resolver el sistema de ecuaciones lineales

$$A\mathbf{x} = \mathbf{b},$$

donde $A = (a_{ij}) \in M(n \times n, \mathbb{R})$, $\mathbf{b} = (b_1, b_2, \ldots, b_n)^T \in \mathbb{R}^n$ y $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ es la incógnita.

Consideremos una matriz $Q \in M(n \times n, \mathbb{R})$, la cual suponemos tiene inversa. Tenemos que $A\mathbf{x} = \mathbf{b}$ es equivalente a escribir $0 = -A\mathbf{x} + \mathbf{b}$, sumando a ambos miembros de esta última igualdad el término $Q\mathbf{x}$, no queda la ecuación $Q\mathbf{x} = Q\mathbf{x} - A\mathbf{x} + \mathbf{b}$, multiplicando esta ecuación por la izquierda por Q^{-1} obtenemos $\mathbf{x} = (I - Q^{-1}A)\mathbf{x} + Q^{-1}\mathbf{b}$, lo que nos sugiere proponer el siguiente método iterativo para resolver la ecuación original $A\mathbf{x} = \mathbf{b}$,

$$\mathbf{x}^{(k+1)} = (I - Q^{-1}A)\mathbf{x}^{(k)} + Q^{-1}\mathbf{b}, \tag{3.30}$$

donde $\mathbf{x}^{(0)} \in \mathbb{R}^n$ es arbitrario. Sobre la convergencia del método propuesto, tenemos el siguiente corolario.

Corolario 3.5 *La fórmula de iteración* $\mathbf{x}^{(k+1)} = \left(I - Q^{-1}A\right)\mathbf{x}^{(k)} + Q^{-1}\mathbf{b}$ *define una sucesión que converge a la solución de* $A\mathbf{x} = \mathbf{b}$, *para cualquier condición inicial* $\mathbf{x}^{(0)} \in \mathbb{R}^n$ *si y sólo si* $\rho\left(I - Q^{-1}A\right) < 1$. *Además, si existe una norma* $|||\ ||$ *en* $M(n \times n, \mathbb{R})$, *en la cual se tiene* $||I - Q^{-1}A|| < 1$, *entonces el método iterativo converge cualesquiera que sea la condición inicial* $\mathbf{x}^{(0)} \in \mathbb{R}^n$ *dada.*

Observación La fórmula de iteración

$$\mathbf{x}^{(k+1)} = \left(I - Q^{-1}A\right)\mathbf{x}^{(k)} + Q^{-1}\mathbf{b}$$

define un método iterativo de punto fijo. En efecto, consideremos la función $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ definida por $F(\mathbf{x}) = \left(I - Q^{-1}A\right)\mathbf{x} + Q^{-1}\mathbf{b}$. Observemos que si $\mathbf{x}$ es un punto fijo de F entonces $\mathbf{x}$ es una solución de la ecuación $A\mathbf{x} = \mathbf{b}$. Además, si $\mathbf{x} = \left(I - Q^{-1}A\right)\mathbf{x} + Q^{-1}\mathbf{b}$ entonces se tiene que

$$\mathbf{x}^{(k)} - \mathbf{x} = \left(I - Q^{-1}A\right)\left(\mathbf{x}^{(k-1)} - \mathbf{x}\right),$$

por lo tanto

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leqslant \left\|I - Q^{-1}A\right\| \ \left\|\mathbf{x}^{(k-1)} - \mathbf{x}\right\|,$$

de donde, iterando esta desigualdad obtenemos

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leqslant \left\|I - Q^{-1}A\right\|^k \ \left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|,$$

luego si $\left\|I - Q^{-1}A\right\| < 1$, se tiene de inmediato que $\lim_{k\to\infty} \left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| = 0$ para cualquier $\mathbf{x}^{(0)} \in \mathbb{R}^n$ dado.

Observación. La condición $\left\|I - Q^{-1}A\right\| < 1$ implica que las matrices $Q^{-1}A$ y A tienen inversas.

Teorema 3.19 *Si* $\left\|I - Q^{-1}A\right\| < 1$ *para alguna norma matricial subordinada en* $M(n \times n, \mathbb{R})$, *entonces el método iterativo* $\mathbf{x}^{(k+1)} = \left(I - Q^{-1}A\right)\mathbf{x}^{(k)} + Q^{-1}\mathbf{b}$ *es convergente a una solución de* $A\mathbf{x} = \mathbf{b}$, *para cualquier vector inicial* $\mathbf{x}^{(0)} \in \mathbb{R}^n$. *Además, si* $\lambda = \left\|I - Q^{-1}A\right\| < 1$, *podemos usar el criterio de parada* $\left\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\right\| < \varepsilon$ *y se tiene que*

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leqslant \frac{\lambda}{1-\lambda}\left\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\right\| \leqslant \cdots \leqslant \frac{\lambda^k}{1-\lambda}\left\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\right\|.$$

, es decir,

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leqslant \frac{\lambda^k}{1-\lambda}\left\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\right\|. \tag{3.31}$$

También vale la desigualdad siguiente

$$\left\|\mathbf{x}^{(k)}-\mathbf{x}\right\| \leqslant \frac{\rho(I-Q^{-1}A)^k}{1-\rho(I-Q^{-1}A)}\left\|\mathbf{x}^{(1)}-\mathbf{x}^{(0)}\right\|. \tag{3.32}$$

Observación. Desde la definición de radio espectral de una matriz, se tiene que $\rho(A)<1$ implica que existe una norma matricial $||\ ||$ en $M(n\times n,\mathbb{R})$, tal que $||A||<1$. Por otra parte, si $||A||<1$ para alguna norma matricial en $M(n\times n,\mathbb{R})$ entonces se tiene que $\rho(A)<1$. Notemos que si existe una norma matricial $||\ ||$ en $M(n\times n,\mathbb{R})$, tal que $||A||\geqslant 1$ no necesariamente se tiene que $\rho(A)\geqslant 1$.

Observación. Si M_1 y M_2 son dos métodos iterativos convergentes para resolver un sistema de ecuaciones lineales $A\mathbf{x}=\mathbf{b}$, digamos $M_1:\ \mathbf{x}^{(k+1)}=G_1\mathbf{x}^{(k)}+\mathbf{c}_1$ y $M_2:\ \mathbf{x}^{(k+1)}=G_2\mathbf{x}^{(k)}+\mathbf{c}_2$. Si $\rho(G_1))\leqslant\rho(G_2)$, entonces M_1 converge más rápido que M_2 a la solución de la ecuación. Además, si existe una norma $||\cdot||$ en $M(n\times n,\mathbb{R})$ tal que $||G_1||\leqslant||G_2||$, entonces M_1 converge más rápido que M_2 a la solución de la ecuación.

3.9 Método de Richardson

Para este método tomamos $Q_R=I$, luego el método iterativo $\mathbf{x}^{(k+1)}=\left(I-Q^{-1}A\right)\mathbf{x}^{(k)}+Q^{-1}\mathbf{b}$ se transforma en

$$\mathbf{x}^{(k+1)}=\underbrace{(I-A)}_{T_R}\mathbf{x}^{(k)}+\mathbf{b}, \tag{3.33}$$

el cual converge si y sólo si $\rho(T_R)=\rho(I-A)<1$.

Si existe una norma matricial $\|\cdot\|$ para la cual se tiene que $\|T_R\|=||I-A||<1$, entonces el método iterativo de Richardson es convergente.

Ejemplo 64 Resolver el sistema de ecuaciones lineales usando el método de Richardson.

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} \frac{11}{18} \\ \frac{11}{18} \\ \frac{11}{18} \end{pmatrix}$$

Solución. Tenemos

$$T_R=I-A=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}-\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} & 1 \end{pmatrix}=\begin{pmatrix} 0 & -\frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{3} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{3} & 0 \end{pmatrix}$$

Llamando $\mathbf{x}^{(j)} = (x_j \ y_j \ z_j)^T$, nos queda

$$\mathbf{x}^{(k+1)} = T_R\mathbf{x}^{(k)} + \mathbf{b} = \begin{pmatrix} 0 & -\frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{3} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{3} & 0 \end{pmatrix} \begin{pmatrix} x_k \\ y_k \\ z_k \end{pmatrix} + \begin{pmatrix} \frac{11}{18} \\ \frac{11}{18} \\ \frac{11}{18} \end{pmatrix}$$

se tiene $||T_R||_\infty = \frac{5}{6} <$, luego el método iterativo de Richardson para este sistema converge.

3.10 Método de Jacobi

En este caso tomamos

$$Q_J = \text{diag}\,(A) = \begin{pmatrix} a_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{pmatrix}$$

donde los elementos fuera de la diagonal en Q_J son todos iguales a 0. Notemos que $\det(Q_J) = a_{11}a_{22}\cdots a_{nn}$, luego $\det(Q_J) \neq 0$ si y sólo si $a_{ii} \neq 0$ para todo $i = 1, \ldots, n$, es decir, Q_J tiene inversa si y sólo si $a_{ii} \neq 0$ para todo $i = 1, \ldots, n$.

Como antes, colocando esta matriz $Q_J = \text{diag}(A)$ en la fórmula iterativa definida por

$$\mathbf{x}^{(k+1)} = \underbrace{\left(I - \text{diag}(A)^{-1}A\right)}_{T_J}\mathbf{x}^{(k)} + \text{diag}(A)^{-1}\mathbf{b}\,, \tag{3.34}$$

obtenemos un método iterativo convergente si y sólo si $\rho(T_J) = \rho(I - \text{diag}(A)^{-1}A) < 1$. Si existe una norma matricial $\|\cdot\|$ para la cual se tiene que $\|T_J\| = ||I - \text{diag}(A)^{-1}A|| < 1$, entonces el método iterativo de Jacobi es convergente.

Examinemos un poco más la fórmula iterativa del método de Jacobi. Tenemos, en este caso, que un elemento genérico de $Q^{-1}A$ es de la forma $\frac{a_{ij}}{a_{ii}}$ y todos los elementos de la diagonal de $Q^{-1}A$ son iguales a 1. Luego, tomando la norma $\|\ \|_\infty$ en $M(n \times n, \mathbb{R})$ tenemos que

$$\left\|I - \text{diag}(A)^{-1}A\right\|_\infty = \max_{1 \leqslant i \leqslant n} \left\{ \sum_{j=1\,,j\neq i}^{n} \frac{|a_{ij}|}{|a_{ii}|} \right\}.$$

Recordemos que una matriz A es *diagonal dominante* si

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

es decir, $\sum\limits_{j=1\,,\,j\neq i}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| < 1$, por lo tanto tenemos el siguiente teorema.

Teorema 3.20 *Si A es una matriz diagonal dominante, entonces el método iterativo de Jacobi es convergente para cualquiera que sea la condición inicial $\mathbf{x}^{(0)}$ elegida.*

3.11 Método de Gauss–Seidel

En esta caso, consideremos la matriz Q como la matriz triangular obtenida considerando la parte triangular inferior de la matriz A incluyendo su diagonal, es decir,

$$Q_{\text{G-S}} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

Se tiene que $\det(Q_{\text{G-S}}) = a_{11}a_{22}\cdots a_{nn}$. Luego, $\det(Q_{\text{G-S}}) \neq 0$ si y sólo si $a_{ii} \neq 0$ ($i = 1, \ldots, n$). Usando la matriz $Q_{\text{G-S}}$, obtenemos el método iterativo

$$\mathbf{x}^{(k+1)} = \underbrace{\left(I - Q_{\text{G-S}}^{-1}A\right)}_{T_{\text{G-S}}}\mathbf{x}^{(k)} + Q_{\text{G-S}}^{-1}\mathbf{b}\,. \tag{3.35}$$

Teorema 3.21 *Sea $A \in M(n \times n, \mathbb{R})$. Entonces el método iterativo de Gauss–Seidel es convergente si y sólo si $\rho(T_{G\text{-}S}) = \rho(I - Q_{G\text{-}S}^{-1}A) < 1$, donde $Q_{G\text{-}S}$ es la matriz triangular inferior definida arriba a partir de A.*

Como en el caso del método iterativo de Jacobi, tenemos el siguiente teorema.

Teorema 3.22 *Si $A \in M(n \times n, \mathbb{R})$ es una matriz es diagonal dominante entonces el método de Gauss–Seidel converge para cualquier elección inicial $\mathbf{x}^{(0)}$.*

Ejemplo 65 Considere el sistema de ecuaciones lineales

$$\begin{pmatrix} 4 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 6 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 7 \end{pmatrix}$$

Explicitaremos los métodos iterativos de Jacobi y Gauss–Seidel para este sistema. Estudiaremos la convergencia de ellos sin iterar. Finalmente, sabiendo que la solución exacta del sistema es $(x_T,\, y_T,\, z_T) = (1,\, 0,\, 1)$ y usando el punto de partida $(1, 1, 1)$ nos podemos preguntar ¿Cuál de ellos converge más rápido?, para las comparaciones usaremos la norma $\|\quad\|_\infty$ en $M(n \times n, \mathbb{R})$.

Primero que nada tenemos que los métodos iterativos de Jacobi y Gauss–Seidel convergen, pues la matriz asociada al sistema es diagonal dominante.

Explicitaremos el método de Jacobi. En este caso tenemos

$$Q_J = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

por lo tanto

$$Q_J^{-1} = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{6} \end{pmatrix}$$

luego

$$Q_J^{-1}A = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{6} \end{pmatrix} \begin{pmatrix} 4 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 6 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{2}{5} & 1 & \frac{2}{5} \\ \frac{1}{6} & \frac{1}{3} & 1 \end{pmatrix}$$

de donde

$$T_J I - Q_J^{-1}A = \begin{pmatrix} 0 & -\frac{1}{2} & -\frac{1}{4} \\ -\frac{2}{5} & 0 & -\frac{2}{5} \\ -\frac{1}{6} & -\frac{1}{3} & 0 \end{pmatrix} \text{ y } Q_J^{-1}\mathbf{b} = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{6} \end{pmatrix} \begin{pmatrix} 5 \\ 4 \\ 7 \end{pmatrix} = \begin{pmatrix} \frac{5}{4} \\ \frac{4}{5} \\ \frac{7}{6} \end{pmatrix},$$

por lo tanto

$$\begin{cases} x_{k+1} = \dfrac{-2y_k - z_k + 5}{4} \\ y_{k+1} = \dfrac{-2x_k - 2z_k + 4}{5} \\ z_{k+1} = \dfrac{-x_k - 2y_k + 7}{6} \end{cases}$$

Como $||T_J||_\infty = \max\{|\frac{1}{2}|+|\frac{-1}{4}|, |\frac{-2}{5}|+|\frac{-2}{5}|, |\frac{-1}{6}|+|\frac{-1}{3}|\} = \max\{\frac{3}{4}, \frac{4}{5}, \frac{1}{2}\} = \{0.75, 0.8, 0.5\} = 0.8 < 1$, el método de Jacobi converge para cualquier condición inicial $\mathbf{x}^{(0)} \in \mathbb{R}^n$ dada. Ahora explicitemos el método de Gauss–Seidel. En este caso,

$$Q_{\text{G-S}} = \begin{pmatrix} 4 & 0 & 0 \\ 2 & 5 & 0 \\ 1 & 2 & 6 \end{pmatrix} \qquad \text{por lo tanto} \qquad Q_{\text{G-S}}^{-1} = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ -\frac{1}{10} & \frac{2}{10} & 0 \\ -\frac{1}{120} & -\frac{8}{120} & \frac{20}{120} \end{pmatrix}$$

así obtenemos que

$$Q_{\text{G-S}}^{-1}A = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ -\frac{1}{10} & \frac{2}{10} & 0 \\ -\frac{1}{120} & -\frac{8}{120} & \frac{20}{120} \end{pmatrix} \begin{pmatrix} 4 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 6 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{8}{10} & \frac{3}{10} \\ 0 & -\frac{2}{120} & \frac{103}{120} \end{pmatrix}$$

luego $I - Q_{\text{G-S}}^{-1}A = \begin{pmatrix} 0 & -\frac{1}{2} & -\frac{1}{4} \\ 0 & \frac{2}{10} & -\frac{3}{10} \\ 0 & \frac{2}{120} & \frac{17}{120} \end{pmatrix}$ y $Q_{\text{G-S}}^{-1}\mathbf{b} = \begin{pmatrix} \frac{5}{4} \\ \frac{3}{10} \\ \frac{103}{120} \end{pmatrix}$, por lo tanto

$$\begin{cases} x_{k+1} = \dfrac{-2y_k - z_k + 5}{4} \\ y_{k+1} = \dfrac{2y_k - 3z_k + 3}{10} \\ z_{k+1} = \dfrac{2y_k - 17z_k + 103}{120} \end{cases}$$

Como $||T_{\text{G-S}}||_\infty = \max\{|\frac{-1}{4}| + |\frac{-1}{4}|, |\frac{2}{10}| + |\frac{-3}{10}|, |\frac{2}{120}| + |\frac{17}{120}\} = \max\{\frac{3}{4}, \frac{5}{10}, \frac{19}{120}\} = 0.75 < 1$. Luego, el método de Gauss–Seidel converge para cualquier condición inicial $\mathbf{x}^{(0)} \in \mathbb{R}^n$ dada.

Notemos que cómo $||T_{\text{G-S}}||_\infty < ||T_J||_\infty$, vemos que el método de Gauss–Seidel converge más rápido que el método de Jacobi.

Ejercicio. Realizar las iteraciones en ambos caso, Jacobi y Gauss-Seidel para este ejemplo.

3.12 Método SOR (successive overrelaxation)

Este método, también llamado sobre-relajación sucesiva.

Supongamos que escogemos la matriz Q como $Q_{\text{SOR}} = \frac{1}{\omega} D - C$, donde $\omega \in \mathbb{R}$, $\omega \neq 0$, es un parámetro, D es una matriz simétrica, positiva definida y C satisface $C + C^T = D - A$, tenemos entonces el método iterativo

$$\mathbf{x}^{(k+1)} = \underbrace{\left(I - Q_{\text{SOR}}^{-1} A\right)}_{T_{\text{SOR}}} \mathbf{x}^{(k)} + Q_{\text{SOR}}^{-1} \mathbf{b}\,, \tag{3.36}$$

Teorema 3.23 *Si A es simétrica y positiva definida, Q_{SOR} es no singular y $0 < \omega < 2$, entonces el método iterativo SOR converge para todo valor inicial dado.*

Recordemos que una matriz $A \in M(n \times n, \mathbb{R})$ de la forma

$$A = \begin{pmatrix} a_1 & b_1 & 0 & 0 & \cdots & 0 \\ c_2 & a_2 & b_2 & 0 & \cdots & 0 \\ 0 & c_3 & a_3 & b_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & \cdots & 0 & 0 & c_n & a_n \end{pmatrix}$$

es llamada *tridiagonal*.

Teorema 3.24 *Si A es simétirca, positiva definida y tridiagonal, entonces, en caso se tiene $\rho(T_{G\text{-}S}) = \rho(T_J)^2$, donde $T_J = D^{-1}(C_L + C_U)$ es la matriz de iteración en el método de Jacobi y $T_{G\text{-}S} = (D - C_L)^{-1} C_U$ es la matriz de iteración en el método de Gauss–Seidel. Además, la elección óptima de ω en el método SOR es*

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho(T_J)^2}} = \frac{2}{1 + \sqrt{1 - \rho(T_{G\text{-}S})}} \tag{3.37}$$

Por otra parte, también se tiene

$$\rho(T_{SOR,\omega_{opt}}) = \omega_{opt} - 1\,. \tag{3.38}$$

3.13 Otra forma de expresar los métodos iterativos para sistemas lineales

Sea

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

podemos escribir A en la forma

$$\begin{aligned} A &= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}}_{D} - \underbrace{\begin{pmatrix} 0 & 0 & \cdots & 0 \\ -a_{21} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -a_{n1} & \cdots & -a_{nn-1} & 0 \end{pmatrix}}_{C_L} - \underbrace{\begin{pmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & -a_{n-1n} \\ 0 & 0 & \cdots & 0 \end{pmatrix}}_{C_U} \end{aligned}$$

Notemos primero que $A\mathbf{x} = \mathbf{b}$ si y sólo si $(D - C_L - C_U)\mathbf{x} = b$ si y sólo si $D\mathbf{x} = (C_L + C_U)\mathbf{x} + \mathbf{b}$.

Veamos como quedan los métodos anteriores con esta notación.

Método de Jacobi. Si $\det(D) \neq 0$, entonces como $(D - C_L - C_U)\mathbf{x} = b$ se sigue que

$$\mathbf{x} = \underbrace{D^{-1}(C_L + C_U)}_{T_j}\mathbf{x} + \underbrace{D^{-1}b}_{C_J} \tag{3.39}$$

luego el método de Jacobi se puede escribir como

$$\mathbf{x}^{(k+1)} = \underbrace{D^{-1}(C_L + C_U)}_{T_j}\mathbf{x}^{(k)} + \underbrace{D^{-1}b}_{C_J}$$

Método de Gauss–Seidel. Este se puede escribir como

$$(D - C_L)\mathbf{x}^{(k+1)} = C_U\mathbf{x}^{(k)} + \mathbf{b} \tag{3.40}$$

de donde, si $\det(D - C_L) \neq 0$ nos queda

$$\mathbf{x}^{(k+1)} = \underbrace{(D - C_L)^{-1}C_U}_{T_{\text{G-S}}}\mathbf{x}^{(k)} + \underbrace{(D - C_L)^{-1}}_{C_{\text{G-S}}}b$$

Resumen. Supongamos que la matriz A se escribe en la forma $A = D - C_L - C_U$, donde $D = diag(A)$, C_L es el negativo de la parte triangular inferior estricta de A y C_U es el negativo de la parte triangular superior estricta de A. Entonces podemos rescribir los métodos iterativos vistos antes como sigue:

Richardson

$$\begin{cases} Q = I \\ G = I - A \end{cases}$$

$$\mathbf{x}^{(k+1)} = (I - A)\mathbf{x}^{(k)} + \mathbf{b}$$

Jacobi

$$\begin{cases} Q = D \\ G = D^{-1}(C_L + C_U) \end{cases}$$

$$D\mathbf{x}^{(k+1)} = (C_L + C_U)\mathbf{x}^{(k)} + \mathbf{b}$$

Gauss-Seidel

$$\begin{cases} Q = D - C_L \\ G = (D - C_L)^{-1}C_U \end{cases}$$

$$(D - C_L)\mathbf{x}^{(k+1)} = C_U\mathbf{x}^{(k)} + \mathbf{b}$$

SOR

$$\begin{cases} Q = \omega^{-1}(D - \omega C_L) \\ G = (D - \omega C_L)^{-1}(\omega C_U + (1 - \omega)D) \end{cases}$$

$$(D - \omega C_L)\mathbf{x}^{(k+1)} = \omega(C_U\mathbf{x}^{(k)} + \mathbf{b}) + (1 - \omega)D\mathbf{x}^{(k)}$$

$$\omega = \tfrac{1}{\alpha},\ 0 < \omega < 2.$$

3.14 Ejemplos resueltos

Problema 3.1 Considere el sistema de ecuaciones lineales

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{4} \\ \\ \frac{1}{4} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} x_1 \\ \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{7}{12} \\ \\ 0.45 \end{pmatrix}.$$

(a) Para realizar los cálculos con decimales, se considera el vector $\hat{\mathbf{b}} = \begin{pmatrix} 0.58 \\ 0.45 \end{pmatrix}$. Obtenga una cota para el error relativo de la solución del sistema con respecto a la solución del sistema $A\mathbf{x} = \hat{\mathbf{b}}$.

(b) Ahora considere la matriz perturbada

$$\hat{A} = \begin{pmatrix} 0.33 & 0.25 \\ 0.25 & 0.20 \end{pmatrix}.$$

Obtenga una cota para el error relativo de la solución del sistema original con respecto a la solución del sistema $\hat{A}\mathbf{x} = \mathbf{b}$.

Solución. Tenemos el sistema de ecuaciones lineales

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{7}{12} \\ 0.45 \end{pmatrix}.$$

a) El vector $\hat{\mathbf{b}} = \begin{pmatrix} 0.58 \\ 0.45 \end{pmatrix}$ es una perturbación del vector $\mathbf{b} = \begin{pmatrix} \frac{7}{12} \\ 0.45 \end{pmatrix}$. Sean $\mathbf{x}_T$ la solución exacta del sistema $A\mathbf{x} = \mathbf{b}$ y $\mathbf{x}_A$ la solución exacta del sistema $A\mathbf{x} = \hat{\mathbf{b}}$. Se tiene entonces que $\mathbf{x}_A$ es una aproximación a $\mathbf{x}_T$. Para simplificar los cálculos trabajaremos con la norma subordinada $||\cdot||_\infty$. Como $\hat{\mathbf{b}}$ es una perturbación de $\mathbf{b}$, tenemos la fórmula

$$E_R(\mathbf{x}_A) = \frac{||\mathbf{x}_T - \mathbf{x}_A||_\infty}{||\mathbf{x}_T||_\infty} \leqslant ||A||_\infty ||A^{-1}||_\infty \frac{||\mathbf{b} - \hat{\mathbf{b}}||_\infty}{||\mathbf{b}||_\infty}.$$

Ahora,

$$\mathbf{b} - \hat{\mathbf{b}} = \begin{pmatrix} 3.33333333 \times 10^{-3} \\ 0 \end{pmatrix}.$$

Recordemos que si $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, entonces

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

En nuestro caso, $\det \begin{pmatrix} \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{5} \end{pmatrix} = \frac{1}{240}$. Luego,

$$A^{-1} = \frac{1}{\frac{1}{240}} \begin{pmatrix} \frac{1}{5} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{3} \end{pmatrix} = 240 \begin{pmatrix} \frac{1}{5} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 48 & -60 \\ -60 & 80 \end{pmatrix}.$$

De esto, tenemos

$$\begin{aligned} ||A||_\infty &= \max\left\{\frac{1}{3} + \frac{1}{4}, \frac{1}{4} + \frac{1}{5}\right\} = \max\left\{\frac{7}{12}, \frac{9}{20}\right\} = \frac{7}{12} \\ ||A^{-1}||_\infty &= \max\{|48| + |-60|, |-60| + |80|\} = 140 \\ ||\mathbf{b}||_\infty &= \max\left\{\frac{7}{12}, 0.45\right\} = \frac{7}{12} \\ ||\hat{\mathbf{b}}||_\infty &= \max\{0.58, 0.45\} = 0.58 \\ ||\mathbf{b} - \hat{\mathbf{b}}||_\infty &= \max\{3.33333333 \times 10^{-3}, 0\} = 3.33333333 \times 10^{-3}. \end{aligned}$$

Reemplazando nos queda

$$E_R(\mathbf{x}_A) \leqslant \frac{7}{12}\, 140\, \frac{3.3333333 \times 10^{-3}}{\frac{7}{12}} = 0.4666666662\,,$$

esto es, $E_R(\mathbf{x}_A) \leqslant 0.4666666662$.

b) Consideremos ahora la matriz perturbada

$$\hat{A} = \begin{pmatrix} 0.33 & 0.25 \\ 0.25 & 0.20 \end{pmatrix}.$$

Podemos escribir $\hat{A}$ en la forma $\hat{A} = A(I+E)$, donde I es la matriz identidad 2×2 y E es la matriz de error, esto es, $.E = \hat{A} - A$. Como $\hat{A} = A(I+E)$, se tiene que $\hat{A} - A = AE$. Luego,

$$AE = \hat{A} - A = \begin{pmatrix} 0.33 & 0.25 \\ 0.25 & 0.20 \end{pmatrix} - \begin{pmatrix} \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{5} \end{pmatrix} = \begin{pmatrix} -3.33333333 \times 10^{-3} & 0 \\ 0 & 0 \end{pmatrix},$$

de donde $||AE||_\infty = 3.33333333 \times 10^{-3}$.

Veamos si podemos usar la fórmula

$$\frac{||\mathbf{x}_T - \mathbf{x}_A||}{||\mathbf{x}_T||} \leqslant \frac{k(A)}{1 - k(A)\,\frac{||AE||}{||A||}} \cdot \frac{||AE||}{||A||}\,,$$

para ello debemos verificar que

$$||AE|| < \frac{1}{||A^{-1}||}\,.$$

Como $||A||_\infty = 140$ y $||AE||_\infty = 3.33333333 \times 10^{-3}$, se cumple que

$$||AE||_\infty < \frac{1}{140} = 7.14285714 \times 10^{-3}\,,$$

y podemos aplicar la fórmula anterior. Reemplazando nos queda

$$\begin{aligned} \frac{||\mathbf{x}_T - \mathbf{x}_A||_\infty}{||\mathbf{x}_T||_\infty} &\leqslant \frac{81.66666667}{1 - 81.66666667 \times \frac{3.33333333 \times 10^{-3}}{0.5833333333}} \times \frac{3.3333333 \times 10^{-3}}{0.5833333333} \\ &= \frac{81.66666667}{1 - 81.66666667 \times 5.7142857 \times 10^{-3}} \times 5.7142857 \times 10^{-3} \\ &= \frac{0.466666667}{0.5333333349} = 0.8749999945 \end{aligned}$$

Luego $E_R(\mathbf{x}_A) \leqslant 0.8749999945$.

Problema 3.2 Dada la matriz

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix} \quad \text{se tiene que} \quad A^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 5/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{bmatrix}$$

(a) Obtenga la descomposición de Cholesky de A. Utilice lo anterior para resolver el sistema $A\mathbf{x} = \mathbf{b}$ para el vector $b = (1, 1, 2)^T$.

(b) Estime a priori la magnitud del error relativo si la matriz A se perturba quedando finalmente

$$\tilde{A} = \begin{bmatrix} 2 & -0.99 & -0.98 \\ -1 & 1 & 0.99 \\ -1 & 1.02 & 2.99 \end{bmatrix}$$

Solución.

(a) Tenemos que

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix}$$

es simétrica. Ahora, $A_1 = [2]$, luego $\det(A_1) = 2 > 0$, $A_2 = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ y $\det(A_2) = 1 > 0$, $A_3 = A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix}$ y $\det(A_3) = 2 > 0$. Por lo tanto A es positiva definida, y consecuentemente tiene descomposición de Cholesky.

Para obtener la descompsición de Cholesky de A escribamos

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{bmatrix} \begin{bmatrix} \ell_{11} & \ell_{21} & \ell_{31} \\ 0 & \ell_{22} & \ell_{32} \\ 0 & 0 & \ell_{33} \end{bmatrix}$$

De aquí, $\ell_{11}^2 = 2$, de donde $\ell_{11} = \sqrt{2}$; $\ell_{11}\ell_{21} = -1$, de donde $\ell_{21} = -\frac{\sqrt{2}}{2}$, $\ell_{11}\ell_{31} = -1$; $\ell_{31} = -\frac{\sqrt{2}}{2}$; $\ell_{21}^2 + \ell_{22}^2 = 1$, de donde $\ell_{22} = \frac{\sqrt{2}}{2}$; $\ell_{21}\ell_{31} + \ell_{22}\ell_{32} = 1$, de donde $\ell_{32} = \frac{\sqrt{2}}{2}$; finalmente $\ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2 = 3$, de donde $\ell_{33} = \sqrt{2}$. Por lo tanto,

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & 0 & \sqrt{2} \end{bmatrix}$$

Ahora, resolver el sistema $A\mathbf{x} = \mathbf{b}$, con $\mathbf{b}^T = (1, 1, 2)$ es equivalente a resolver los sistemas

$$\begin{cases} L\mathbf{y} = \mathbf{b} \\ L^T\mathbf{x} = \mathbf{y} \end{cases}$$

El sistema $L\mathbf{y} = \mathbf{b}$

$$\begin{bmatrix} \sqrt{2} & 0 & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

De aquí $\sqrt{2}\,\alpha = 1$, de donde $\alpha = \frac{\sqrt{2}}{2}$; $-\frac{\sqrt{2}}{2}\alpha + \frac{\sqrt{2}}{2}\beta = 1$, reemplazando y despejando nos da que $\beta = \frac{3}{2}\sqrt{2}$; finalmente $-\frac{\sqrt{2}}{2}\alpha + \frac{\sqrt{2}}{2}\beta + \sqrt{2}\gamma = 2$, reemplazando y despejando nos da que $\gamma = \frac{\sqrt{2}}{2}$. Debemos resolver ahora el sistema $L^T\mathbf{x} = \mathbf{y}$, es decir,

$$\begin{bmatrix} \sqrt{2} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{3\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

de donde $z = \frac{1}{2}$, $y = \frac{5}{2}$, $x = 2$.

(b) Sean $\mathbf{x}_T$ y $\mathbf{x}_A$ las soluciones exactas de los sistemas $A\mathbf{x} = \mathbf{b}$ y $\tilde{A}\mathbf{x} = \mathbf{b}$, respectivamente, es decir, $\mathbf{x}_T = A^{-1}\mathbf{b}$. Entonces

$$\begin{aligned} E(\mathbf{x}_A = ||\mathbf{x}_A - \mathbf{x}_T|| &= ||\mathbf{x}_A - A^{-1}\mathbf{b}|| \\ &= ||\mathbf{x}_A - A^{-1}\tilde{A}\mathbf{x}_A|| = ||(I - A^{-1}\tilde{A})\mathbf{x}_A|| \\ &\leqslant ||I - A^{-1}\tilde{A}||\,||\mathbf{x}_A|| \end{aligned}$$

de donde

$$E_R(\mathbf{x}_A) = \frac{||\mathbf{x}_A - \mathbf{x}_T||}{||\mathbf{x}_A||} \leqslant ||I - A^{-1}\tilde{A}||\,.$$

Para simplificar los cálculos usamos la norma subordinada $||\ ||_\infty$. Ahora, tenemos

$$A^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & \frac{5}{2} & \frac{-1}{2} \\ 0 & \frac{-1}{2} & \frac{1}{2} \end{bmatrix}$$

$$\tilde{A} = \begin{bmatrix} 2 & -0.99 & -0.98 \\ -1 & 1 & 0.99 \\ -1 & 1.02 & 2.99 \end{bmatrix}$$

luego,

$$A^{-1}\tilde{A} = \begin{bmatrix} 1 & 0.01 & 0.01 \\ 0 & 1.0 & 0 \\ 0 & 0.01 & 1.0 \end{bmatrix}$$

Además, como

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

se tiene que

$$I - A^{-1}\tilde{A} = \begin{bmatrix} 0 & -0.01 & -0.01 \\ 0 & 0 & 0 \\ 0 & -0.01 & 0 \end{bmatrix}$$

luego, $||I-A^{-1}\tilde{A}||_\infty = \max\{0.02, 0, 0.01\} = 0.02$, de donde $\frac{||\mathbf{x}_A-\mathbf{x}_T||_\infty}{||\mathbf{x}_A||_\infty} \leqslant ||I-A^{-1}\tilde{A}||_\infty = 0.02$.

Otra solución. Escribimos $\tilde{A}$ de la forma $\tilde{A} = A(I+E)$. Entonces se tiene las fórmulas

1.
$$E_R(\mathbf{x}_A) = \frac{||\mathbf{x}_T - \mathbf{x}_A||}{||\mathbf{x}_T||} \leqslant \frac{||E||}{1-||E||}$$
si $||E|| < 1$.

2.
$$E_R(\mathbf{x}_A) = \frac{||\mathbf{x}_T - \mathbf{x}_A||}{||\mathbf{x}_T||} \leqslant \frac{k(A)}{1-k(A)\frac{||AE||}{||A||}} \frac{||AE||}{||A||}$$
si $||AE|| < \frac{1}{||A^{-1}||}$.

Usemos por ejemplo la primera de ellas. Debemos calcular la matriz E . De la ecuación $\tilde{A} = A(I+E)$ se tiene que $A^{-1}\tilde{A} = I + E$, de donde, $A^{-1}\tilde{A} - I = E$. Realizando los cálculos, obtenemos que

$$E = \begin{bmatrix} 0 & 0.01 & 0.01 \\ 0 & 0 & 0 \\ 0 & 0.01 & 0 \end{bmatrix}$$

Por simplicidad de los cálculos, usaremos la norma subordinada $||\ ||_\infty$. Luego, $||E||_\infty = 0.02 < 1$, por lo tanto, reemplazando nos queda,

$$\frac{||\mathbf{x}_T - \mathbf{x}_A||}{||\mathbf{x}_T||} \leqslant \frac{||E||}{1-||E||} = \frac{0.02}{1-0.02} = \frac{0.02}{0.98} = 0.0204016...$$

Ahora usamos la segunda de las fórmulas. Tenemos que calcular $k(A) = ||A||\ ||A^{-1}||$. Sabemos que

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix} \quad \text{luego} \quad ||A||_\infty = 5$$

y

$$A^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & \frac{5}{2} & \frac{-1}{2} \\ 0 & \frac{-1}{2} & \frac{1}{2} \end{bmatrix} \quad \text{luego} \quad ||A^{-1}||_\infty = 4\,,$$

de donde $k_\infty(A) = 5 \times 4 = 20$. Para aplicar la fórmula debemos verificar que $||AE||_\infty < \frac{1}{||A^{-1}||_\infty}$. Tenemos que

$$AE = \begin{bmatrix} 0 & 0.01 & 0.01 \\ 0 & 0.005 & 0.01 \\ 0 & 0.005 & 0 \end{bmatrix}$$

luego, $||AE||_\infty = 0.02$ y $||A^{-1}||_\infty = 4$, de aquí se tiene que $\frac{1}{||A^{-1}||_\infty} = \frac{1}{4} = 0.25$ y es claro entonces que se satisface la condición $||AE||_\infty < \frac{1}{||A^{-1}||_\infty}$. Reemplazando los valores obtenidos, nos queda

$$\begin{aligned} \frac{||\mathbf{x}_T - \mathbf{x}_A||_\infty}{||\mathbf{x}_T||_\infty} &\leqslant \frac{20}{1 - 20 \cdot \frac{0.02}{5}} \cdot \frac{0.02}{5} = \frac{20 \times 0.004}{1 - 20 \times 0.004} \\ &= \frac{0.08}{1 - 0.08} = \frac{0.08}{0.92} = 0.08695652... \end{aligned}$$

esto es,

$$\frac{||\mathbf{x}_T - \mathbf{x}_A||_\infty}{||\mathbf{x}_T||_\infty} \leqslant 0.08695652...$$

Problema 3.3 Sean

$$A = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad b = \begin{pmatrix} 5.00 \\ 1.02 \\ 1.04 \\ 1.10 \end{pmatrix}$$

(a) Calcule una solución aproximada $\mathbf{x}_A$ del sistema $A\mathbf{x} = \mathbf{b}$, primero aproximando cada entrada del vector $\mathbf{b}$ al entero más próximo, obteniendo un vector $\tilde{\mathbf{b}}$ y luego resolviendo el sistema $A\mathbf{x} = \tilde{\mathbf{b}}$.

(b) Calcule $||\mathbf{r}||_\infty$ y $k_\infty(A)$, donde $\mathbf{r}$ es el vector residual, es decir $\mathbf{r} = \mathbf{b} - A\mathbf{x}_A$ y $k_\infty(A)$ el número de condicionamiento de la matriz A.

(c) Estime una cota para el error relativo de la solución aproximada, respecto a la solución exacta (no calcule esta última explícitamente).

Solución.

(a) El vector perturbado es $(5\ 1\ 1\ 1)^T$. Luego la solución del sistema perturbado es

$$\begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

de donde $(x, y, z, w) = (12, 4, 2, 1) = \mathbf{x}_A$

(b) Tenemos $||b||_\infty = ||\tilde{b}||_\infty = 5$. Por otra parte,

$$\begin{aligned} r = b - A\mathbf{x}_A &= \begin{pmatrix} 5.00 \\ 1.02 \\ 1.04 \\ 1.10 \end{pmatrix} - \begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 12 \\ 4 \\ 2 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 5.00 \\ 1.02 \\ 1.04 \\ 1.10 \end{pmatrix} - \begin{pmatrix} 5 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.02 \\ 0.04 \\ 0.10 \end{pmatrix} \end{aligned}$$

Luego, $||r||_\infty = 0.10$.

Para encontrar la inversa de la matriz A basta resolver el sistema

$$\begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

de donde

$$A^{-1} = \begin{pmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Luego, $||A^{-1}||_\infty = 8$. Tenemos que $||A||_\infty = 4$, luego $k_\infty(A) = ||A||_\infty ||A^{-1}||_\infty = 4 \times 8 = 32$.

(c) Usamos la fórmula

$$\frac{1}{k(A)} \frac{||r||}{||b||} \leqslant \frac{||\mathbf{x}_T - \mathbf{x}_A||}{||\mathbf{x}_T||} \leqslant k(A) \frac{||r||}{||b||}$$

Reemplazando los datos nos queda

$$\frac{1}{32} \times \frac{0.10}{5} \leqslant \frac{||\mathbf{x}_T - \mathbf{x}_A||}{||\mathbf{x}_T||} \leqslant 32 \times \frac{0.10}{5}$$

es decir,

$$0.625 \times 10^{-3} \leqslant \frac{||\mathbf{x}_T - \mathbf{x}_A||}{||\mathbf{x}_T||} \leqslant 0.64\,.$$

Problema 3.4 Encuentre una descomposición del tipo LU para la matriz A siguiente

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

Use la descomposión anterior para solucionar el sistema $A\mathbf{x} = (1 \;\; 1 \;\; -1 \;\; -1)^T$.

Solución. La descomposición de Doolittle de A es dada por

$$A = L \cdot U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{3}{4} & 1 & 0 & 0 \\ \frac{1}{2} & \frac{6}{7} & 1 & 0 \\ \frac{1}{4} & \frac{5}{7} & \frac{5}{6} & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 3 & 2 & 1 \\ 0 & \frac{7}{4} & \frac{3}{2} & \frac{5}{4} \\ 0 & 0 & \frac{12}{7} & \frac{10}{7} \\ 0 & 0 & 0 & \frac{5}{3} \end{pmatrix}$$

Ahora resolver el sistemas $A\mathbf{x} = b$ es equivalente a resolver los sistemas $L\mathbf{y} = b$ y $U\mathbf{x} = y$. Tenemos que $b = (1\ 1\ -1\ -1)^T$. Llamando $\mathbf{y} = (y_1\ y_2\ y_3\ y_4)^T$, de la forma del sistema $L\mathbf{y} = b$, obtenemos que $y_1 = 1$, $y_2 = \frac{1}{4}$, $y_3 = -\frac{12}{7}$, y $y_4 = 0$. Ahora al resolver el sistema $U\mathbf{x} = b$ obtenemos $x_4 = 0$, $x_3 = -1$, $x_2 = 1$ y $x_1 = 0$.

Problema 3.5 Considere el sistema de ecuaciones lineales $A\mathbf{x} = b$, donde

$$A = \begin{pmatrix} 2 & -3 & 8 & 1 \\ 4 & 0 & 1 & -10 \\ 16 & 4 & -2 & 1 \\ 0 & 7 & -1 & 5 \end{pmatrix} \text{ y } \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

a) Usando aritmética exacta y el método de eliminación gaussiana con pivoteo resuelva el sistema de ecuaciones lineales $A\mathbf{x} = b$.

b) Denote por B a la matriz PA obtenida en el item anterior Demuestre que el método iterativo de Jacobi aplicado a la matriz B es convergente. Usando como punto inicial (0.0377,0.21819,0.20545,-0.06439) y la máxima capacidad de su calculadora, realice iteraciones con el método de Jacobi para obtener una solución aproximada del sistema $B\mathbf{x} = b$. Use como criterio de parada $\|(x_{n+1}, y_{n+1}, z_{n+1}, w_{n+1}) - (x_n, y_n, z_n, w_n)\| \leqslant 10^{-5}$.

Solución. a) Tenemos que

$$A = \begin{pmatrix} 2 & -3 & 8 & 1 \\ 4 & 0 & 1 & -10 \\ 16 & 4 & -2 & 1 \\ 0 & 7 & -1 & 5 \end{pmatrix} \text{ y } b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Escribiendo la matrix ampliada del sistema, nos queda

$$A = \left(\begin{array}{cccc|c} 2 & -3 & 8 & 1 & 1 \\ 4 & 0 & 1 & -10 & 1 \\ 16 & 4 & -2 & 1 & 1 \\ 0 & 7 & -1 & 5 & 1 \end{array}\right),$$

de donde

$$s = \begin{pmatrix} 8 \\ 10 \\ 16 \\ 7 \end{pmatrix} \text{ y } P = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Luego

$$m_1 = \max\left\{\frac{2}{8}, \frac{4}{10}, \frac{16}{16}, 0\right\} = 1 = m_{31}$$

por lo tanto debemos intercambiar la fila 1 con la fila 3. Realizando este intercambio de filas y la eliminación gaussiana, nos queda

$$\left(\begin{array}{cccc|c} 16 & 4 & -2 & 1 & 1 \\ 4 & 0 & 1 & -10 & 1 \\ 2 & -3 & 8 & 1 & 1 \\ 0 & 7 & -1 & 5 & 1 \end{array}\right) \begin{array}{c} F_2 - \frac{1}{4}F_1 \\ \rightarrow \\ \\ F_3 - \frac{1}{8}F_1 \end{array} \left(\begin{array}{cccc|c} 16 & 4 & -2 & 1 & 1 \\ 0 & -1 & 3/2 & -41/4 & 3/4 \\ 0 & -7/2 & 33/4 & 7/8 & 7/8 \\ 0 & 7 & -1 & 5 & 1 \end{array}\right)$$

de donde

$$L_1 = \begin{pmatrix} 1 \\ 1/4 \\ 1/8 \\ 0 \end{pmatrix}, \quad s_1 = \begin{pmatrix} 16 \\ 10 \\ 8 \\ 7 \end{pmatrix}, \; P_1 = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 4 \end{pmatrix}$$

Para la elección del segundo pivote tenemos

$$m_2 = \max\left\{\frac{1}{10}, \frac{7}{16}, \frac{7}{7}\right\} = 1 = m_{42}$$

por lo tanto debemos intercambiar la fila 2 con la fila 4

$$\begin{pmatrix} 16 & 4 & -2 & 1 & 1 \\ 0 & 7 & -1 & 5 & 1 \\ 0 & -7/2 & 33/4 & 7/8 & 7/8 \\ 0 & -1 & 3/2 & -41/4 & 3/4 \end{pmatrix} \begin{array}{c} F_3 - \frac{-1}{2}F_2 \\ \rightarrow \\ \\ F_4 - \frac{-1}{7}F_2 \end{array} \left(\begin{array}{cccc|c} 16 & 4 & -2 & 1 & 1 \\ 0 & 7 & -1 & 5 & 1 \\ 0 & 0 & 31/4 & 27/8 & 11/8 \\ 0 & 0 & 19/14 & -227/28 & 25/28 \end{array}\right)$$

luego,

$$L_1^1 = \begin{pmatrix} 1 \\ 0 \\ 1/8 \\ 1/4 \end{pmatrix}, \; L_2 = \begin{pmatrix} 0 \\ 1 \\ -1/2 \\ -1/7 \end{pmatrix}, \; P_2 = \begin{pmatrix} 3 \\ 4 \\ 1 \\ 2 \end{pmatrix}, \; S_2 = \begin{pmatrix} 16 \\ 7 \\ 8 \\ 10 \end{pmatrix}$$

Para la elección del tercer pivote tenemos que

$$m_3 = \max\left\{\frac{31}{32}, \frac{19}{140}\right\} = \frac{31}{32} = m_{33}$$

y obtenemos

$$\left(\begin{array}{cccc|c} 16 & 4 & -2 & 1 & \\ 0 & 7 & -1 & 5 & 1 \\ 0 & 0 & 31/4 & 27/8 & 11/8 \\ 0 & 0 & 19/14 & -227/28 & 25/8 \end{array}\right) \begin{array}{c} \\ \rightarrow \\ \\ F_4 - \frac{38}{217}F_3 \end{array} \left(\begin{array}{cccc|c} 16 & 4 & -2 & 1 & 1 \\ 0 & 7 & -1 & 5 & 1 \\ 0 & 0 & 31/4 & 27/8 & 11/8 \\ 0 & 0 & 0 & -4395/434 & 283/434 \end{array}\right)$$

luego,

$$L_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 38/217 \end{pmatrix}$$

Por lo tanto

$$PA = LU = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1/8 & -1/2 & 1 & 0 \\ 1/4 & -1/7 & 38/217 & 1 \end{pmatrix} \begin{pmatrix} 16 & 4 & -2 & 1 \\ 0 & 7 & -1 & 5 \\ 0 & 0 & 31/4 & 27/8 \\ 0 & 0 & 0 & -4395/434 \end{pmatrix} \tag{3.41}$$

donde

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Por lo tanto

$$PA = \begin{pmatrix} 16 & 4 & -2 & 1 \\ 0 & 7 & -1 & 5 \\ 2 & -3 & 8 & 1 \\ 4 & 0 & 1 & -10 \end{pmatrix}$$

Resolviendo la ecuación se obtiene que

$$w = -\frac{283}{4395},\ z = \frac{301}{1465},\ y = \frac{959}{4395},\ x = 331/8790 \tag{3.42}$$

Ahora, como

$$B = PA = \begin{pmatrix} 16 & 4 & -2 & 1 \\ 0 & 7 & -1 & 5 \\ 2 & -3 & 8 & 1 \\ 4 & 0 & 1 & -10 \end{pmatrix}$$

tenemos que la matriz B es diagonal dominante, por lo tanto el método de Jacobi converge. Por otra parte, la matriz de Jacobi es dada por

$$\begin{aligned} J &= I - Q^{-1}B \\ &= I_4 - \begin{pmatrix} 1/16 & 0 & 0 & 0 \\ 0 & 1/7 & 0 & 0 \\ 0 & 0 & 1/8 & 0 \\ 0 & 0 & 0 & -1/10 \end{pmatrix} \begin{pmatrix} 16 & 4 & -2 & 1 \\ 0 & 7 & -1 & 5 \\ 2 & -3 & 8 & 1 \\ 4 & 0 & 1 & -10 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -1/4 & 1/8 & -1/16 \\ 0 & 0 & 1/7 & -5/7 \\ -1/4 & 3/8 & 0 & -1/8 \\ 2/5 & 0 & 1/10 & 0 \end{pmatrix} \end{aligned}$$

Para determinar si el método de Jacobi converge determinaremos los valores propios de J. Tenemos

$$\det(J - \lambda I_4) = \det \begin{pmatrix} -\lambda & -1/4 & 1/8 & -1/16 \\ 0 & -\lambda & 1/7 & -5/7 \\ -1/4 & 3/8 & -\lambda & -1/8 \\ 2/5 & 0 & 1/10 & -\lambda \end{pmatrix} = \lambda^4 + \frac{17}{1120}\lambda^2 - \frac{219}{4480}\lambda + \frac{33}{2240} \quad (3.43)$$

de donde se tiene que los valores propios son

$$\begin{aligned} \lambda_{1,2} &= -0.2483754458 \pm 0.3442140503i \\ \lambda_{3,4} &= -0.2483754458 \pm 0.1416897424i \end{aligned} \quad (3.44)$$

Por lo tanto, se tiene que el radio espectral es $\rho = 0.4244686967 < 1$ y el método de Jacobi es convergente.

Realizando las iteraciones, obtenemos la siguinete tabla

n	x_n	y_n	z_n	w_n	E
0					0.000041875
1	0.037658125	0.2182	0.205445	-0.064375	0.00001725
2	0.0376540625	0.218188571428	0.20545734375	-0.064392640625	0.000014084822
3	0.0376595407368	0.21820265625	0.20545622991	-0.064392640625	0.3961078E-5
4	0.0376559047154	0.218202776148	0.205460190988	-0.0643905607143	

Problema 3.6 Para cada $n \in \mathbb{N}$, se definen

$$A_n = \begin{pmatrix} 1 & 2 \\ 2 & 4 + \frac{1}{n^2} \end{pmatrix} \text{ y } b_n = \begin{pmatrix} 1 \\ 2 - \frac{1}{n^2} \end{pmatrix}.$$

Se desea resolver el sistema $A_n x = b_n$. Un estudiante obtiene como resultado $\tilde{x} = (1, 0)^{\mathrm{T}}$.

1. Se define el vector residual asociado a esta solución como

$$r_n = A_n \tilde{x} - b_n$$

Calcule r_n ¿ Podemos decir que para n grande, la solución es razonablemente confiable? Justifique

2. Resolver $A_n x = b_n$ en forma exacta. Denote por $\bar{x}_n$ la solución exacta y cálcule el error relativo de la aproximación $\tilde{x} = (1,0)^{\mathrm{T}}$.

3. Lo razonablemente esperado es que $\bar{x}_n$ converja a $\tilde{x}$ cuando n tiende a infinito. Para este ejemplo, ¿ se tiene dicha afirmación?. Calcule $K_\infty(A_n)$ y explique el resultado obtenido.

Solución. Para cada $n \in \mathbb{N}$, un resultado para una solución de $A_n x = b_n$ es $\tilde{x} = (1,0)$.

1. El vector residual asociado a esta solución es

$$r_n = A_n \tilde{x} - b_n \, .$$

Tenemos

$$\begin{aligned} r_n &= \begin{pmatrix} 1 & 2 \\ 2 & 4+\frac{1}{n^2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 2-\frac{1}{n^2} \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2-\frac{1}{n^2} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{n^2} \end{pmatrix} \end{aligned}$$

Ahora, la solución exacta del sistema es

$$\begin{pmatrix} 1 & 2 \\ 2 & 4+\frac{1}{n^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2-\frac{1}{n^2} \end{pmatrix}$$

esto del par de ecuaciones lineales

$$\begin{aligned} x + 2y &= 1 \\ 2x + \left(4 + \frac{1}{n^2}\right) y &= 2 - \frac{1}{n^2} \, . \end{aligned}$$

De la primera ecuación obtenemos $x = 1 - 2y$, reemplazando en la segunda ecuación nos queda

$$2 - 4y + 4y + \frac{1}{n^2} y = 2 - \frac{1}{n^2} \, ,$$

es decir, $\dfrac{1}{n^2} y = -\dfrac{1}{n^2}$, de donde $y = -1$, por lo tanto $x = 3$. Note que la solución es exactamente la misma para cada $n \in \mathbb{N}$, es decir, la solución no depende de $n \in \mathbb{N}$.

Por lo tanto la solución dada no es razonablemente confiable, pues no se aproxima en nada a la solución exacta $x_T = (3,-1)$ del sistema $A_n x = b_n$.

2. Ya calculamos la solución exacta $\overline{x}_n$ de $A_n x = b_n$, la cual es $\overline{x}_n = (3,-1) = x_T$ para todo $n \in \mathbb{N}$.

Usando la norma $||\cdot||_\infty$ para simplificar los cálculos.

$$\begin{aligned} E_R(\tilde{x}) &= \frac{||x_T - \tilde{x}||_\infty}{||x_T||_\infty} \\ &= \frac{||(3,-1)-(1,0)||_\infty}{||(3,-1)||_\infty} \\ &= \frac{||(2,-1)||_\infty}{||(3,-1)||_\infty} \\ &= \frac{\max\{|2|,|-1|\}}{\max\{|3|,|-1|\}} \\ &= \frac{2}{3}. \end{aligned}$$

3. No, pues $x_n = (3,-1) = x_T$ es un vector constante y no se aproxima (obviamente) a $\tilde{x} = (1,0)$.

   Para calcular $k_\infty(A_n)$, tenemos

$$||A_n||_\infty = \max\{|1|+|2|, |2| + \left|4 + \frac{1}{n^2}\right|\} = 6 + \frac{1}{n^2}$$

   Ahora, $\det(A_n) = 4 + \frac{1}{n^2} - 4 = \frac{1}{n^2}$. Luego

$$\begin{aligned} A_n^{-1} &= \frac{1}{\frac{1}{n^2}} \begin{pmatrix} 4+\frac{1}{n^2} & -2 \\ -2 & 1 \end{pmatrix} \\ &= n^2 \begin{pmatrix} 4+\frac{1}{n^2} & -2 \\ -2 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 4n^2+1 & -2n^2 \\ -2n^2 & n^2 \end{pmatrix} \end{aligned}$$

   y $||A_n^{-1}||_\infty = \max\{|14n^2+1| + |-2n^2| + |n^2|\} = \max\{6n^2+1, 3n^2\} = 6n^2+1$. Por lo tanto

$$\begin{aligned} k_\infty(A_n) &= \left(6 + \frac{1}{n^2}\right)(6n^2+1) \\ &= 36n^2 + 6 + 6 + \frac{1}{n^2} \\ &= 36n^2 + 12 + \frac{1}{n^2} \xrightarrow[n\to\infty]{} \infty, \end{aligned}$$

lo cual significa que la matriz A_n está muy mal condicionada para valores grandes de n (es numéricamente no invertible para n grande), lo cual explica la mala aproximación $\tilde{x}$ de x_n para n grande.

3.15 Ejercicios Propuestos

Problema 3.1 Para resolver un sistema de ecuaciones lineales $Ax = b$, donde $x, b \in \mathbb{R}^2$, se propone el siguiente método iterativo

$$x^{(k+1)} = Bx^{(k)} + b\,, \qquad k \geqslant 0$$

donde

$$B = \begin{pmatrix} \lambda & c \\ 0 & -\lambda \end{pmatrix}, \qquad \lambda, c \in \mathbb{R}$$

1. ¿Para qué valores de λ y c el método iterativo propuesto es convergente?
2. Sea $\tilde{x}$ el punto fijo de la iteración. Calcule $||\tilde{x} - x^{(k)}||_\infty$ y $||x^{(k+1)} - x^{(k)}||_\infty$ cuando $k \longrightarrow \infty$ ¿Es la convergencia la punto fijo independiente de c ? Justificque.

Problema 3.2 Dado un método iterativo para resolver un sistema de ecuaciones lineales

$$x^{(k+1)} = Bx^{(k)} + c\,,$$

Si $\det(B) = 0$ ¿Puede el método propuesto ser convergente?

Problema 3.3 Si un método iterativo de punto fijo $x_{n+1} = Ax_n$, donde $A \in \mathbb{M}(n \times n, \mathbb{R})$ tiene un punto fijo $\bar{x} \neq 0$. Pruebe que $||A|| \geqslant 1$ para cualquier norma ssubordinada en $\mathbb{M}(n \times n, \mathbb{R})$.

Problema 3.4 Dada una matriz

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

y un vector $\mathbf{b} \in \mathbb{R}^3$, se quiere resolver el sistema $A\mathbf{x} = \mathbf{b}$. Para ello, se propone el método iterativo siguiente:

$$\mathbf{x}^{(k+1)} = -\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}^{-1} \cdot \begin{pmatrix} 0 & 0 & a_{13} \\ 0 & 0 & a_{23} \\ a_{31} & a_{32} & 0 \end{pmatrix} \mathbf{x}^{(k)} + \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}^{-1} \mathbf{b}$$

(a) Pruebe que el método propuesto resulta convergente cuando se aplica a las matriz A y al vector $\mathbf{b}$ siguientes

$$A = \begin{pmatrix} 8 & 2 & -3 \\ -3 & 9 & 4 \\ 3 & -1 & 7 \end{pmatrix} \qquad \text{y} \qquad \mathbf{b} = \begin{pmatrix} -20 \\ 62 \\ 0 \end{pmatrix}.$$

Indicación: Use que $\begin{pmatrix} 8 & 2 & 0 \\ -3 & 9 & 0 \\ 0 & 0 & 7 \end{pmatrix}^{-1} = \begin{pmatrix} 3/26 & -1/39 & 0 \\ 1/26 & 4/39 & 0 \\ 0 & 0 & 1/7 \end{pmatrix}$.

(b) Considere el vector $\mathbf{x}^{(0)} = (0\,0\,0)^T$. Encuentre el número mínimo de iteraciones necesarias $k \in \mathbb{N}$, de modo de tener una precisión $||\mathbf{x}^{(k)} - \mathbf{x}||_\infty \leqslant 10^{-4}$.

(c) Compare el método anterior con el método de Jacobi en cuanto a velocidad de convergencia. Justifique.

Problema 3.5 Encuentre la inversa A^{-1} de la matriz A, las normas $||A||_1$, $||A^{-1}||_1$, $||A||_2$, $||A^{-1}||_2$, $||A||_\infty$, $||A^{-1}||_\infty$ y los números de condición $\kappa_1(A)$, $\kappa_2(A)$ y $\kappa_\infty(A)$ si

1. $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

2. $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

3. $A = \begin{pmatrix} -2 & -1 & 2 & 1 \\ 1 & 2 & 1 & -2 \\ 2 & -1 & 2 & 1 \\ 0 & 2 & 0 & 1 \end{pmatrix}$

Problema 3.6 Sean $A, B \in \mathbb{M}(n\times n, \mathbb{R})$, matrices invertibles. Pruebe que $\kappa(AB) \leqslant \kappa(A)\kappa(B)$, donde el número de condición es calculado usando cualquier norma subordinada en $\mathbb{M}(n\times n, \mathbb{R})$.

Problema 3.7 Sean $A, B \in \mathbb{M}(3 \times 3, \mathbb{R})$ las matrices

$$A = \begin{pmatrix} a & c & 0 \\ c & a & c \\ 0 & c & a \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & b & 0 \\ b & 0 & b \\ 0 & b & 0 \end{pmatrix}$$

1. Probar que $\lim_{n\to\infty} B^n = 0$ si y sólo si $|b| < \sqrt{2}/2$.

2. Dar condiciones necesarias y suficientes sobre $a, c \in \mathbb{R}$ para la convergencia de los métodos de Jacobi y de Gauss–Seidel aplicados a resolver el sistema de ecuacion es lineales $A\mathbf{x} = \mathbf{b}$.

Problema 3.8 Considere el sistema de ecuaciones lineales siguiente

$$\begin{cases} 3x + y + z &= 5 \\ x + 3y - z &= 3 \\ 3x + y - 5z &= -1 \end{cases}$$

1. Explicite el método iterativo de Jacobi para encontrar solución a la ecuación. Justifique porqué este método converge (sin usar calculadora). Comenzando con $(0, 0, 0)$, realice 10 iteraciones ¿Cuál es el error absoluto respecto de la solución exacta?

2. Explicite el método iterativo de Gauss–Seidel y justifique la convergencia o no de este método para la ecuación dada.

Problema 3.9 Sean

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & -2 \\ -2 & 1 & 1 \end{pmatrix} \quad \text{y} \quad b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Proponga un método iterativo convergente para resolver $Ax = b$. La demostración de convergencia debe hacerse sin iterar. Resuelva la ecuación utilizando el método iterativo propuesto. Use como critero de parada $||(x_{n+1}, y_{n+1}, z_{n+1}) - (x_n, y_n, z_n)||_2 \leqslant 10^{-3}$.

Problema 3.10 Considere el sistema de ecuaciones lineales

$$\begin{cases} 3x - 2y + z &= 1 \\ x - \frac{2}{3}y + 2z &= 2 \\ -x + 2y - z &= 0 \end{cases}$$

1. Resolver el sistema usando descomposición LU y aritmética exacta.
2. Resolver el sistema usando descomposición LU y aritmética decimal con 4 dígitos y redondeo.
3. Resolver el sistema usando aritmética exacta sin pivoteo.
4. Resolver el sistema usando aritmética decimal con 4 dígitos y redondeo, y sin pivoteo.
5. Resolver el sistema usando aritmética exacta con pivoteo.
6. Resolver el sistema usando aritmética decimal con 4 dígitos y redondeo, y con pivoteo.
7. ¿cuál es su conclusión?

Problema 3.11

1. Obtener la descomposición LU para matrices tridiagonales.
2. Describir el proceso de eliminación gaussiana sin pivoteo para resolver un sistema con matriz tridiagonal.
3. Suponiendo que tenemos una matriz tridiagonal no singular. Explicitar los algoritmos iterativos de Richardson, Jacobi, y Gauss–Seidel.
4. Ilustre todo lo anterior con el sistema

$$\begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \\ 10 \end{pmatrix}$$

Problema 3.12 Considere los siguientes sistemas de ecuaciones lineales

$$\begin{pmatrix} 3 & -2 & 1 \\ 1 & -\frac{2}{3} & 2 \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \tag{3.45}$$

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 2 & 1 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \\ -5 \end{pmatrix} \tag{3.46}$$

$$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 3 & 1 & -5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \\ -1 \end{pmatrix} \tag{3.47}$$

1. Estudie la convergencia de los métodos iterativos de Richardson, Jacobi, Gauss–Seidel y de SOR para cada uno de ellos.

2. Caso algunos de los métodos anteriores sea convergente, encontrar la solución del sistema, con precisión $\varepsilon = 10^{-3}$ y usando como criterio de parada $||(x_{n+1}, y_{n+1}, z_{n+1}) - (x_n, y_n, z_n)||_\infty \leqslant \varepsilon$, comenzando con $(x_0, y_0, z_0) = (0, 0, 0)$.

Problema 3.13 Considere el sistema de ecuaciones lineales

$$\begin{pmatrix} -10^{-5} & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

1. Resuelva el sistema usando eliminación gaussiana sin pivoteo y con aritmética de redondeo a 4 dígitos.

2. Resuelva el sistema usando eliminación gaussiana con pivoteo y con aritmética de redondeo a 4 dígitos.

3. Cuál de las soluciones obtenidas en a) y b) es la más aceptable?

4. Calcule el número de condición para la matriz en el sistema inicial y para la matriz del sistema despues de hacer pivoteo Cuál es su conclusión?

Problema 3.14 Dado el sistema lineal de ecuaciones estructurado

$$\begin{pmatrix} a_{11} & a_{12} & & \cdots & a_{1n-1} & a_{1\,n} \\ 0 & a_{22} & & \cdots & a_{2n-1} & a_{2\,n} \\ 0 & 0 & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & & & & \\ 0 & \cdots & 0 & a_{n-2\,n-2} & a_{n-2\,n-1} & a_{n-2\,n} \\ a_{n-1\,1} & a_{n-1\,2} & \cdots & a_{n-1\,n-2} & a_{n-1\,n-1} & a_{n-1\,n} \\ a_{n\,1} & a_{n\,2} & \cdots & a_{n\,n-2} & a_{n\,n-1} & a_{n\,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

a) Diseñe un algoritmo basado en el método de eliminación gaussiana que permita resolver dicho sistema adaptado a la estructura particular de este, sin hacer ceros donde ya existen.

b) Realice un conteo de las operaciones (multiplicación, división, suma y resta) involucrados sólo en la triangulación del sistema en el algoritmo de la parte a).

Problema 3.15 Dado el sistema lineal $Ax = h$ con

$$A = \begin{pmatrix} a_1 & c_1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ b_2 & a_2 & c_2 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & b_3 & a_3 & c_3 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & b_4 & a_4 & c_4 & 0 & \cdots & 0 & 0 & 0 \\ & & & & & & & & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & & & & \\ 0 & 0 & 0 & 0 & & \cdots & & b_{n-1} & a_{n-1} & c_{n-1} \\ 0 & 0 & f_3 & f_4 & & \cdots & & f_{n-2} & b_n & a_n \end{pmatrix},$$

y $h \in \mathbb{R}^n$ un vector columna.

Diseñe un algoritmo en pseudo lenguaje basado en el método de eliminación gaussiana, sin elección de pivote, que utilice el mínimo número de localizaciones de memoria, realice el mínimo número de operaciones (es decir, que obvíe los elementos nulos de la matriz A (sin hacer ceros donde ya existen)) para resolver el sistema $Ax = h$.

Problema 3.16 Resuelva mediante el método de eliminación gaussiana con aritmética exacta el sistema

$$\begin{array}{rcrcrcr} 5x_1 & - & x_2 & & & = & 4 \\ -x_1 & + & 5x_2 & - & x_3 & = & -2 \\ & - & x_2 & + & 5x_3 & = & 1 \end{array}$$

Problema 3.17 Dado el sistema lineal

$$\begin{cases} x &= 0.9x + y + 17 \\ y &= 0.9y - 13 \end{cases}$$

1. Proponga un método iterativo convergente (en $\mathbb{R}^2$). La demostración de convergencia debe hacerse sin iterar.

2. Resuelva el sistema por su método propuesto, con aritmética decimal de redondeo a 3 dígitos y con una precisión de 10^{-3} en la norma $||\ ||_\infty$, eligiendo como punto de partida $(0, 0)$. (el redondeo debe hacerse en cada operación).

3. Analice la convergencia del método de Jacobi y del método de Gauss–Seidel. Sin iterar.

Problema 3.18 Dado el sistema lineal

$$\begin{cases} 3x + y + z &= 5 \\ 3x + y - 5z &= -1 \\ x + 3y - z &= 1 \end{cases}$$

1. Resuelva el sistema usando eliminación gaussiana, con o sin pivoteo, con aritmética decimal de redondeo a 3 dígitos. (el redondeo debe hacerse en cada operación).

2. Analice la convergencia del método de Jacobi, sin iterar.

3. Analice la convergencia del método de Gauss–Seidel, sin iterar.

Problema 3.19 Considere el sistema de ecuaciones lineal

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & 1 & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} .$$

1. Resuelva el sistema por el método de eliminación de Gauss (sin elección de pivote) redondeando cada operación a 4 dígitos en la mantisa.
2. Resuelva el sistema por el método de Jacobi redondeando cada operación a 4 dígitos en la mantisa, y con una presición de 10^{-3}, comenzando con $x_0 = y_0 = z_0 = 0$.
3. En este caso particular ¿Cuál de los dos métodos es más conveniente? Justifique.

Problema 3.20 Resuelva el sistema $Ax = b$, donde

$$A = \begin{pmatrix} -3 & 2 & 3 & -1 \\ 6 & -2 & -6 & 0 \\ -9 & 4 & 10 & 3 \\ 12 & -4 & -13 & -5 \end{pmatrix} \quad \text{y} \quad b = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

usando descomposición LU con pivoteo.

Problema 3.21 Considere la matriz

$$A = \begin{pmatrix} 0.0001 & 1 \\ 1 & 1 \end{pmatrix}$$

1. Trabajando con aritmética decimal de tres dígitos y redondeo ¿Es posible encontrar una descomposición LU para A?
2. Considere ahora aritmética decimal de cinco dígitos y redondeo. Encuentre una descomposición LU para A, y compare los números de condición de A y de LU ¿Cuál es su conclusión?
3. Ahora considere la matriz
$$A = \begin{pmatrix} 1 & 1 \\ 0.0001 & 1 \end{pmatrix}$$
Es posible ahora encontrar una descomposición LU, con aritmética decimal de redondeo a tres dígitos para A?

Problema 3.22 Considere el siguiente sistema de ecuaciones lineales

$$\begin{pmatrix} 10 & -3 & 6 \\ 1 & -8 & -2 \\ -2 & 4 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 25 \\ -9 \\ -50 \end{pmatrix}$$

1. Resuelva el sistema usando descomposición LU de Doolittle.

2. Cuáles de los métodos iterativos: Richardson, Jacobi, Gauss–Seidel convergen?

3. Use un método iterativo convergente para encontrar la solución del sistema con una precisión de $\varepsilon = 10^{-3}$, con criterio de parada $||(x_n, y_n, z_n) - (x_T, y_T, z_T)||_\infty \leqslant \varepsilon$, donde (x_T, y_T, z_T) es la solución exacta del sistema.

Problema 3.23 Considere la matriz

$$A = \begin{pmatrix} 4 & -1 & -1 \\ -1 & 4 & 1 \\ -1 & 1 & 4 \end{pmatrix}$$

a) Demuestre que existe la descomposión de Cholesky (sin calcularla) de A.

b) Determine la descompisición de Cholesky de A.

Problema 3.24 a) Demuestre que la matriz

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2 & 1 \\ 2 & 3 & 1 \end{pmatrix}$$

no tiene descomposición LU.

b) Realice pivoteo en la matriz A y demuestre que la matriz resultante tiene descomposición LU.

Problema 3.25 Considere la matriz

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix}$$

1. Usando aritmética exacta, encuentre una descomposición $A = LU$, donde L es triangular inferior con "unos" en la diagonal.

2. Usando la descomposición LU obtenida en a), encuentre la solución del sistema (usando aritmética exacta) $Ax = b$, con $b^T = (5, -3, 8)$.

3. Proponga un método de punto fijo y demuestre la convergencia (sin iterar) para resolver el sistema dado en b), y encuentre una solución aproximada con una presición de 10^{-3}.

Problema 3.26 Consideremos el problema de calcular la temperatura de una barra metálica de largo L, cuyos extremos se mantienen a temperaturas constantes T_0 y T_L conocidas. La ecuación diferencial que gobierna este fenómeno es conocida como *ecuación de conducción del calor*, la cual, en régimen estacionario está dada por

$$-kT''(x) = f(x) \qquad 0 < x < L, \tag{3.48}$$

donde $f(x)$ es una función conocida, que representa una fuente de calor externa y $k > 0$ es una constante que se denomina *coeficiente de difusión o conductividad térmica*. A esta ecuación se le agregan las condiciones de borde:

$$T(0) = T_0, \qquad T(L) = T_L. \tag{3.49}$$

Una de las técnicas más usadas para resolver el problema 3.48-3.49 es el *método de diferencias finitas*. En este método, el intervalo $[0, L]$ se divide en $N+1$ intervalos de largo $h = \frac{L}{N+1}$ y la solución es buscada en los puntos *internos* definidos por esta división, es decir, en los puntos $x_n = nh$, con $n = 1, \ldots, N$ (los valores de la temperatura en los nodos del borde $x_0 = 0$ y $x_{N+1} = L$ son conocidos). En este método, las derivadas de primer orden se aproximan por el *cuociente de diferencias finitas*

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

mientras que las derivadas de segundo orden, se aproximan por

$$f''(x) \approx \frac{f(x-h) - 2f(x) + f(x+h)}{h^2}. \tag{3.50}$$

Denotemos por T_n los valores aproximados de la evaluaciones de la temperatura exacta en los puntos de la malla, $T(x_n)$. Reemplazando $-T''(x)$ en la ecuación 3.48 por la expresión 3.50, y evaluando la ecuación en los *nodos internos* de la malla $\{x_1, \ldots, x_N\}$, se obtiene el siguiente sistema de ecuaciones lineales:

$$\begin{aligned} 2T_1 - T_2 &= h^2 f(x_1) + T_0, && (3.51)\\ -T_{j-1} + 2T_j - T_{j+1} &= h^2 f(x_j), \quad j = 2, \ldots, N-1; && (3.52)\\ 2T_{N-1} + 2T_N &= h^2 f(x_N) + T_L, && (3.53) \end{aligned}$$

que matricialmente puede ser escrito como

$$\begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_{N-1} \\ T_N \end{pmatrix} = h^2 \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_{N-1}) \\ f(x_N) \end{pmatrix} + \begin{pmatrix} T_0 \\ 0 \\ \vdots \\ 0 \\ T_L \end{pmatrix} \tag{3.54}$$

Denotemos por A la matriz y por F_h el lado derecho de este sistema.

Tomando como parámetros del problema los valores

$$\begin{aligned} &k = 1, \qquad L = 1, \qquad f(x) = 37\left(\tfrac{\pi}{2}\right)^2 \sin(\tfrac{\pi}{2}x),\\ &T_0 = 0, \qquad T_L = 37, \end{aligned} \tag{3.55}$$

se pide resolver numéricamente el sistema 3.54, usando los métodos iterativos de Jacobi, Gauss-Seidel y SOR. Para el método SOR tome distintos valores del parámetro de aceleración $\omega \in (1, 2)$. Calcule además el parámetro de aceleración óptimo ω^*.
Para ello, siga los siguientes pasos:

1. Compruebe que la matriz A es definida positiva.

2. Implemente cada uno de los 3 métodos antes mencionados y haga un estudio comparativo de estos para distintos valores de N. Por ejemplo, $N = 50, 100, 200, 1000$. Especifique el criterio de parada utilizado y el número de iteraciones para cada uno de los métodos y para los distintos valores de N. Presente sus resultados en una tabla.

3. Finalmente, determine la solución analítica del problema 3.48-3.49, con los parámetros dados por 3.55 y grafique el error absoluto en la solución, para cada uno de los métodos y para los distintos valores de N . A partir de estos gráficos, comente cual de los tres métodos aproxima mejor a la solución .

Problema 3.27 Demuestre que la matriz no singular

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

no tiene descomposición LU, pero la matriz $A - I$ si tiene descomposición LU. Dar una matriz de permutación P de modo que PA tiene descomposición LU.

Problema 3.28 Sea $a > 0$ una constante. Se tiene la siguiente matriz

$$A = \begin{pmatrix} 2\,a & 0 & \frac{2}{3}\,a^3 \\ 0 & \frac{2}{3}\,a^3 & 0 \\ \frac{2}{3}\,a^3 & 0 & \frac{2}{5}\,a^5 \end{pmatrix}$$

(a) Demuestre que A tiene descomposición de Cholesky y usela para calcular A^{-1}.

(b) calcule el número de condición de la matriz A, en términos de A y determine cuando ella está bien condicionada.

(c) Si $\mathbf{b} = (0 \ -1 \ 1)^T$. Usando la descomposición anterior resuelva el sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$.

Problema 3.29

Problema 3.30

Problema 3.31

Capítulo 4

Interpolación

Supongamos que tenemos una función $f : [a, b] \longrightarrow \mathbb{R}$ y queremos aproximarla por funciones más simples. Más general, nos son dados los datos (x_0, y_0), (x_1, y_1), $\ldots$, (x_n, y_n) donde los puntos x_i $(i = 0, \ldots, n)$ son distintos y estan sobre el eje x, los cuales suponemos ordenados por $x_0 < x_1 < \cdots < x_n$, y los puntos y_i $(i = 0, 1, \ldots, n)$ son tales que $y_i = f(x_i)$ para alguna función desconocida $f(x)$ con la cual deseamos hacer algunos cálculos. Podemos aproximar $f(x)$ por funciones simples, y hacer los cálculos con estas aproximaciones, por ejemplo, cálculo de integrales (áreas) y de derivadas.

4.1 Interpolación de Lagrange

Comenzamos aproximando los datos (x_0, y_0), (x_1, y_1), $\ldots$, (x_n, y_n), con $x_0 < x_1 < \cdots < x_n$, por funciones polinomiales a trozo o por funciones polinomiales.

4.1.1 Aproximación lineal por trozo o de grado 1

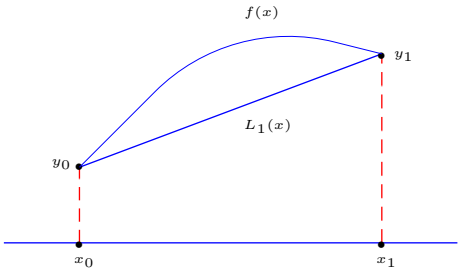
Dados los puntos (x_0, y_0) y (x_1, y_1), con $x_0 < x_1$, tales que $y_0 = f(x_0)$ y $x_1 = f(x_1)$, podemos aproximar $f(x)$ en el intervalo $[x_0, x_1]$ por la más simple de las funciones, es decir, afín . Para ello consideramos el segmento de recta que une los puntos (x_0, y_0) y (x_1, y_1). Tenemos entonces que la aproximación es dada por

$$L_1(x) = \frac{x - x_1}{x_0 - x_1}\, y_0 + \frac{x - x_0}{x_1 - x_0}\, y_1 \,. \tag{4.1}$$

Además, el error viene expresado como

$$\text{Error} = f(x) - L_1(x) = \frac{f''(\xi(x))}{2!}(x - x_0)(x - x_1)\,, \tag{4.2}$$

donde $\xi(x) \in [x_0, x_1]$.



Para obtener la expresión de $L_1(x)$, escribimos

$$L_1(x) = a_0 + a_1 x\,. \tag{4.3}$$

Usando las condiciones $L_1(x_0) = y_0$ y $L_1(x_1) = y_1$, obtenemos el siguiente sistema de ecuaciones lineales

$$\begin{aligned} L_1(x_0) &= a_0 + a_1 x_0 = y_0 \\ L_1(x_1) &= a_0 + a_1 x_1 = y_1 \end{aligned} \tag{4.4}$$

o en forma matricial

$$\begin{pmatrix} 1 & x_0 \\ 1 & x_1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$$

cuya solución es

$$\begin{aligned} a_0 &= \frac{1}{x_1 - x_0}\, y_0 + \frac{-1}{x_1 - x_0}\, y_1 \\ a_1 &= \frac{-1}{x_1 - x_0}\, y_0 + \frac{1}{x_1 - x_0}\, y_1 \end{aligned}$$

Reemplazando estos valores en la expresión para $L_1(x)$ dada en (4.3), obtenemos la fórmula (4.1).

Si queremos aproximar $f(x)$ en todo el intervalo $[a, b]$, consideramos una partición de $[a, b]$, digamos $x_0 = a < x_1 < x_2 < \cdots < x_n = b$, y en cada subintervalo $[x_i, x_{i+1}]$ tomamos una aproximación como la anterior. Para el error total sólo basta sumar los errores cometidos en cada aproximación. Esta es llamada *aproximación lineal por trozos* de $f(x)$.

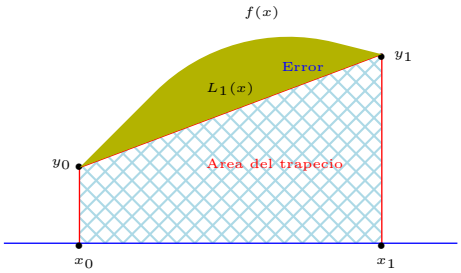
Usando este tipo de aproximación tenemos, por ejemplo,

$$\int_{x_0}^{x_1} f(x) \approx \int_{x_0}^{x_1} L_1(x)dx = \frac{x_1 - x_0}{2}(y_1 + y_0)\,, \tag{4.5}$$

es decir,

$$\int_{x_0}^{x_1} f(x)dx \approx \int_{x_0}^{x_1} L_1(x)dx = \frac{x_1 - x_0}{2}(f(x_0) + f(x_1))\,, \tag{4.6}$$

esta es llamada *regla del trapecio* para el cálculo integral.



Para el error cometido en el cálculo de la integral usando esta aproximación, tenemos que

$$\begin{aligned}\int_{x_0}^{x_1} \frac{f''(\xi(x))}{2!}(x-x_0)(x-x_1)dx &= \frac{f''(c)}{2}\int_{x_0}^{x_1}(x-x_0)(x-x_1) \\ &= \frac{f''(c)}{2}\left(\frac{x^3}{3} - \frac{x_0+x_1}{2}x^2 + x_0x_1x\right)\Big|_{x_0}^{x_1} \\ &= \frac{f''(c)}{2}\cdot\frac{-h^3}{6} \\ &= -\frac{h^3}{12}f''(c)\end{aligned}$$

para algún c en $[x_0, x_1]$, donde $h = x_1 - x_0$. Por lo tanto,

$$\int_{x_0}^{x_1} f(x)dx = \frac{h}{2}(f(x_0)+f(x_1)) - \frac{h^3}{12}f''(c)\,. \tag{4.7}$$

De (4.7) tenemos una cota para el error cometido en el cálculo de la integral, dado por

$$\left|\int_{x_0}^{x_1} f(x)dx - \frac{h}{2}(f(x_0)+f(x_1))\right| \leqslant \frac{h^3}{12}\max\{\,|f''(x)| \,:\, x\in[x_0,x_1]\}\,. \tag{4.8}$$

En general, usando la partición $a = x_0 < x_1 < \cdots < x_n = b$ del intervalo $[a,b]$, con paso constante $h = x_{i+1} - x_i$, tenemos la *fórmula de los trapecios compuesta*

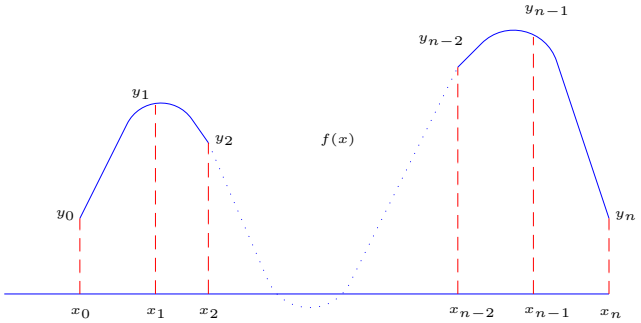
$$\int_a^b f(x) \approx \frac{h}{2}\left(f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)\right)\,. \tag{4.9}$$

El error de esta aproximación lo estudiaremos en el Capítulo 8

4.2 Polinomio de Lagrange de grado n

Dados $n+1$ puntos distintos en el intervalo $[a,b]$, digamos $x_0 = a < x_1 < \cdots < x_n = b$ y $f : [a,b] \longrightarrow \mathbb{R}$ una función $n+1$ veces derivable con $f^{(n+1)}(x)$ continua. Sean $y_i = f(x_i)$, $i = 0, 1, \ldots, n$. Entonces existe un único polinomio $L_n(x)$ de grado menor o igual que n tal que

$$f(x_k) = L_n(x_k)\,, \qquad k = 0, 1, \ldots, n\,. \tag{4.10}$$



Este polinomio es dado por

$$L_n(x) = \sum_{k=0}^{n} f(x_k) L_{n,k}(x)\,, \tag{4.11}$$

donde

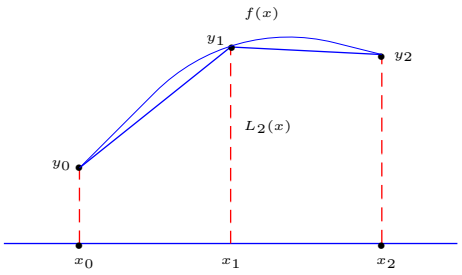
$$\begin{aligned} L_{n,k}(x) &= \frac{(x-x_0)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)} \\ &= \prod_{\substack{i\,=\,0 \\ i\,\neq\,k}}^{n} \frac{(x-x_i)}{(x_k-x_i)}\,. \end{aligned} \tag{4.12}$$

Además, se tiene que

$$\text{Error} = f(x) - L_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}(x-x_0)\cdots(x-x_n) \tag{4.13}$$

donde $\xi(x) \in [x_0, x_n]$.

La siguiente figura ilustra $L_2(x)$,



Para encontrar la fórmula (4.11), escribimos

$$L_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{4.14}$$

e imponemos las condiciones $L_n(x_j) = y_j$ para $j = 0, 1, \ldots, n$. Esto da origen al siguiente sistema de $n+1$ ecuaciones con $n+1$ incognitas

$$\left\{\begin{array}{rcl} L_n(x_0) = a_0 + a_1x_0 + a_2x_0^2 + \cdots + a_nx_0^n & = & y_0 \\ L_n(x_1) = a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n & = & y_1 \\ & \vdots & \\ L_n(x_n) = a_0 + a_1x_n + a_2x_n^2 + \cdots + a_nx_n^n & = & y_n \end{array}\right. \tag{4.15}$$

o en forma matricial

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{4.16}$$

el cual se resuelve usando la regla de Cramer, y se obtiene la fórmula (4.12) para los coeficientes del polinomio de Lagrange.

Aplicando esta aproximación para el cálculo integral, tenemos

$$\int_{x_0}^{x_n} f(x)dx \approx \int_{x_0}^{x_n} L_n(x)dx$$

y

$$\begin{aligned} \int_{x_0}^{x_n} L_n(x)dx &= \int_{x_0}^{x_n} \sum_{i=0}^{n} f(x_i)L_{n,i}(x)dx \\ &= \sum_{i=0}^{n} f(x_i) \int_{x_0}^{x_n} L_{n,i}(x)dx \end{aligned}$$

luego,

$$\begin{aligned} \left| \int_{x_0}^{x_n} f(x)dx - \int_{x_0}^{x_n} L_n(x)dx \right| &= \frac{1}{(n+1)!} \left| \int_{x_0}^{x_n} f^{(n+1)}(\xi(x)) \prod_{i=0}^{n} (x - x_i)dx \right| \\ &\leqslant \frac{1}{(n+1)!} \int_{x_0}^{x_n} \left| f^{(n+1)}(\xi(x)) \right| \left| \prod_{i=0}^{n} (x - x_i) \right| dx \\ &= \frac{|f^{(n+1)}(c)|}{(n+1)!} \int_{x_0}^{x_n} \left| \prod_{i=0}^{n} (x - x_i) \right| dx \\ &\leqslant \max \left\{ \frac{|f^{(n+1)}(x)|}{(n+1)!} \ : \ x \in [x_0, x_n] \right\} \int_{x_0}^{x_n} \left| \prod_{i=0}^{n} (x - x_i) \right| dx \\ &\leqslant \frac{nh}{(n+1)!} K_1 \cdot K_2 \,, \end{aligned}$$

donde $K_1 = \max\left\{\left|f^{(n+1)}(x)\right| : x \in [x_0, x_n]\right\}$ y $K_2 = \max\{|\prod_{i=0}^{n}(x - x_i)| : x \in [x_0, x_n]\}$, esto es,

$$\left|\int_{x_0}^{x_n} f(x)dx - \int_{x_0}^{x_n} L_n(x)dx\right| \leqslant \frac{nh}{(n+1)!}\, K_1 \cdot K_2\,. \tag{4.17}$$

4.3 Regla de Simpson

Sean $x_0 = a$, $x_2 = b$ y $x_1 = a + h$, donde $h = \dfrac{b-a}{2}$. Si usamos $L_2(x)$ para aproximar f, tenemos

$$L_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f(x_2)\,,$$

donde el resto (error) viene dado por

$$R(x) = \frac{(x-x_0)(x-x_1)(x-x_2)}{6}f^{(3)}(\xi(x))\,,$$

con $\xi(x) \in [a, b]$.

Luego

$$\begin{aligned}\int_a^b f(x)dx &= f(x_0)\int_{x_0}^{x_2}\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}dx + f(x_1)\int_{x_0}^{x_2}\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}dx\\ &\quad + f(x_2)\int_{x_0}^{x_2}\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}dx + \int_{x_0}^{x_2}\frac{(x-x_0)(x-x_1)(x-x_2)}{6}f^{(3)}(\xi(x))dx\end{aligned}$$

Tenemos $h = x_2 - x_1 = x_1 - x_0$ y $x_2 - x_0 = (x_2 - x_1) + (x_1 - x_0) = 2h$.

Ejercicio. Desarrollar el cálculo de las integrales y obtener la regla de Simpson.

Una forma alternativa para obtener la regla de Simpson es la siguiente. Desarrollamos $f(x)$ en serie de Taylor hasta orden 4 alrededor del punto x_1 en el intervalo $[x_0, x_2]$. Tenemos, $x_1 = x_0 + h$, luego

$$f(x) = f(x_1) + f'(x_1)(x-x_1) + \frac{f''(x_1)}{2}(x-x_1)^2 + \frac{f'''(x_1)}{6}(x-x_1)^3 + \frac{f^{(4)}(\xi(x))}{24}(x-x_1)^4\,,$$

con $\xi(x) \in [x_0, x_2]$.

Integrando, obtenemos

$$\begin{aligned}\int_{x_0}^{x_2} f(x)dx &= f(x_1)(x_2-x_0) + \left(\frac{f'(x_1)}{2}(x-x_1)^2\right.\\ &\quad \left. + \frac{f''(x_1)}{6}(x-x_1)^3 + \frac{f'''(x_1)}{24}(x-x_1)^4\right)\Bigg|_{x_0}^{x_2}\\ &\quad + \frac{1}{24}\int_{x_0}^{x_2} f^{(4)}(\xi(x))(x-x_1)^4 dx\,.\end{aligned}$$

Como $(x-x_1)^4 \geqslant 0$, podemos usar el teorema valor medio para integrales, y tenemos

$$\begin{aligned}\frac{1}{24}\int_{x_0}^{x_2} f^{(4)}(\xi(x))(x-x_1)^4 dx &= \frac{f^{(4)}(c)}{24}\int_{x_0}^{x_2}(x-x_1)^4 dx \\ &= \frac{1}{120}f^{(4)}(c)(x-x_1)^5\Big|_{x_0}^{x_2}, \qquad c \in]x_0, x_2[\,.\end{aligned}$$

Ahora, usando que $h = x_1 - x_0 = x_2 - x_1$, tenemos

$$\begin{aligned}(x_2-x_1)^2 - (x_0-x_1)^2 &= (x_2-x_1)^4 - (x_0-x_1)^4 = 0 \\ (x_2-x_1)^3 - (x_0-x_1)^3 &= 2h^3 \\ (x_2-x_1)^5 - (x_0-x_1)^4 &= 2h^5\,.\end{aligned}$$

Luego

$$\int_{x_0}^{x_2} f(x)dx = 2hf(x_1) + \frac{h^3}{3}f''(x_1) + \frac{f^{(4)}(c)}{60}h^5\,. \tag{4.18}$$

Usando desarrollos de Taylor, tenemos

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \frac{1}{6}f'''(x)h^3 + \frac{1}{24}f^{(4)}(\xi_1)h^4 \tag{4.19}$$

$$f(x-h) = f(x) - f'(x)h + \frac{1}{2}f''(x)h^2 - \frac{1}{6}f'''(x)h^3 + \frac{1}{24}f^{(4)}(\xi_2)h^4\,, \tag{4.20}$$

donde $x-h < \xi_2 < x < \xi_1 < x+h$. Sumando estas dos igualdades, obtenemos

$$f(x+h) + f(x-h) = 2f(x) + f''(x)h^2 + \frac{1}{24}(f^{(4)}(\xi_1) + f^{(4)}(\xi_2))h^4 \tag{4.21}$$

y resolviendo para $f''(x)$ nos queda

$$f''(x) = \frac{1}{h^2}(f(x-h) - 2f(x) + f(x+h)) - \frac{h^2}{24}(f^{(4)}(\xi_1) + f^{(4)}(\xi_1))\,. \tag{4.22}$$

Si $f^{(4)}(x)$ es continua en $[x-h, x+h]$, usando el teorema del valor medio obtenemos

$$\frac{f^{(4)}(\xi_1) + f^{(4)}(\xi_2)}{2} = f^{(4)}(\xi)\,, \tag{4.23}$$

para algún $\xi \in [x-h, x+h]$.

Luego

$$f''(x) = \frac{1}{h^2}(f(x-h) - 2f(x) + f(x+h)) - \frac{h^2}{12}f^{(4)}(\xi)\,. \tag{4.24}$$

Usando (4.24) en la fórmula (4.18) de la integral, tenemos

$$f''(x_1) = \frac{1}{h^2}(f(x_1 - h) - 2f(x_1) + f(x_1 + h)) - \frac{h^2}{12}f^{(4)}(\xi)\,. \tag{4.25}$$

Como $x_1 - h = x_0$ y $x_1 + h = x_2$, obtenemos que

$$f''(x_1) = \frac{1}{h^2}(f(x_0) - 2f(x_1) + f(x_2)) - \frac{h^2}{12}f^{(4)}(\xi(x))$$

y (4.18) se transforma en

$$\begin{aligned}
\int_{x_0}^{x_2} f(x)dx &= 2hf(x_1) + \frac{h^3}{3}\left[\frac{1}{h^2}\left(f(x_0) - 2f(x_1) + f(x_2)\right) - \frac{h^2}{12}f^{(4)}(\xi)\right] + \frac{f^{(4)}(c)}{60}h^5 \\
&= 2hf(x_1) + \frac{h}{3}(f(x_0) - 2f(x_1) + f(x_2)) - \frac{h^5}{36}f^{(4)}(\xi) + \frac{f^{(4)}(c)}{60}h^5 \\
&= \frac{h}{3}(f(x_0) + 4f(x_1) + f(x_2)) - \frac{h^5}{12}\left(\frac{1}{3}f^{(4)}(\xi) - \frac{1}{5}f^{(4)}(c)\right) \\
&= \frac{h}{3}(f(x_0) + 4f(x_1) + f(x_2)) - \frac{h^5}{12}\cdot 2\cdot\frac{f^{(4)}}{15}(\theta) \\
&= \frac{h}{3}(f(x_0) + 4f(x_1) + f(x_2)) - \frac{h^5}{90}f^{(4)}(\theta)\,.
\end{aligned}$$

En resumen,

Regla de Simpson simple . Si $x_1 = x_0 + h$ y $x_2 = x_1 + h$, entonces

$$\int_{x_0}^{x_2} f(x)dx = \frac{h}{3}(f(x_0) + 4f(x_1) + f(x_2)) - \frac{h^5}{90}f^{(4)}(\theta)\,, \tag{4.26}$$

de donde

$$\left|\int_{x_0}^{x_2} f(x)dx - \frac{h}{3}(f(x_0) + 4f(x_1) + f(x_2))\right| \leqslant \frac{h^5}{90}\,\max\{|f^{(4)}(x)| \,:\, x \in [x_0, x_2]\}\,. \tag{4.27}$$

esta fórmula es llamada *regla de Simpson simple*.

Más general, si usamos puntos $x_0 < x_1 < x_2 < x_3 < x_4 < \cdots < x_{2n-2} < x_{2n-1} < x_{2n}$ y haciendo uso de la propiedad $\int_a^c f(x) = dx = \int_a^b f(x)dx + \int_b^c f(x)dx$ para $a < b < c$, se obtenemos

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

esto es,

$$\int_{x_0}^{x_{2n}} f(x)dx = \frac{h}{3}\left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \cdots + 2f(x_{2n-2}) + 4f(x_{2n-1}) + f(x_{2n})\right) \tag{4.28}$$

esta es fórmula de integración numérica es llamada *regla de Simpson compuesta*. El error total es obtenido sumando los errores en cada parcela del cálculo. En el Capítulo 8 analizarenos este error.

La regla de los trapecios (simple y compuesta) y la regla de Simpson (simple y compuesta) forma parte de una serie de fórmulas para el cálculo aproximado de integrales, llamadas *fórmulas de Newton–Cote*, que serán estudiadas en el Capítulo 8.

4.4 Método de las diferencias divididas

Dados $n+1$ puntos distintos en el intervalo $[a,b]$, digamos $x_0 = a < x_1 < \cdots < x_n = b$ y $f : [a,b] \longrightarrow \mathbb{R}$ una función $n+1$ veces derivable con $f^{(n+1)}(x)$ continua. Sean $y_i = f(x_i)$, $i = 0, 1, \ldots, n$. Tenemos el polinomio de Lagrange $L_n(x)$ interpolando esos puntos. Ahora, queremos escribir $L_n(x)$ en una forma más sencilla, digamos como,

$$L_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \cdots + a_n(x - x_0)\cdots(x - x_{n-1}) \tag{4.29}$$

El problema que tenemos es determinar los coeficientes $a_0, \ldots, a_n$.

Evaluando, tenemos que

$$L_n(x_0) = f(x_0) = a_0\,,$$

$$L_n(x_1) = f(x_1) = f(x_0) + a_1(x_1 - x_0)\,,$$

de donde

$$a_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}\,.$$

Antes de continuar evaluando, introduzcamos la siguiente notación

$$f[x_i] = f(x_i) \tag{4.30}$$

$$f[x_i, x_{i+1}] = \frac{f[x_{i+1}] - f[x_i]}{x_{i+1} - x_i} \tag{4.31}$$

Así, si tenemos definidas $f[x_i, x_{i+1}, \ldots, x_{i+k-1}]$ y $f[x_{i+1}, \ldots, x_{i+k}]$, podemos definir la k–ésima *diferencia dividida relativa* a $x_i, x_{i+1}, \ldots, x_{i+k}$ por

$$f[x_i, x_{i+1}, \ldots, x_{i+k}] = \frac{f[x_{i+1}, \ldots, x_{i+k}] - f[x_i, \ldots, x_{i+k-1}]}{x_{i+k} - x_i}\,. \tag{4.32}$$

Ahora un cálculo directo, nos da que $a_k = f[x_0, x_1, \ldots, x_k]$, así podemos escribir

$$L_n(x) = f[x_0] + \sum_{k=1}^{n} f[x_0, \ldots, x_k](x - x_0) \cdots (x - x_{k-1}) \,. \tag{4.33}$$

La siguiente tabla muestra como se van generando las sucesivas diferencias divididas

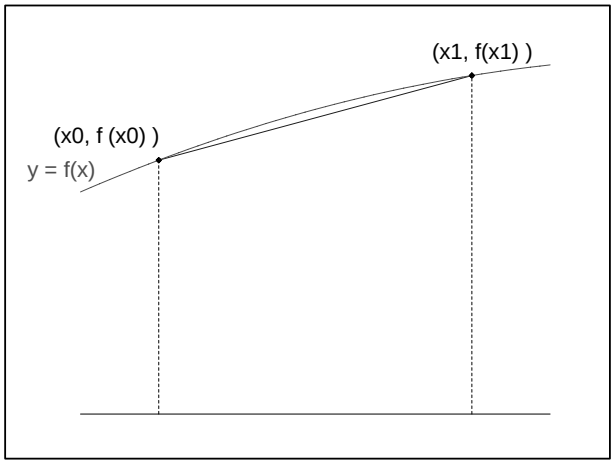
$$\begin{array}{c|ccccccc}
x_0 & f[x_0] & & & & & & \\
 & & \searrow & & & & & \\
 & & f[x_0, x_1] & = & \dfrac{f[x_1] - f[x_0]}{x_1 - x_0} & & & \\
 & & & & & \searrow & & \\
 & & \nearrow & & & & & \\
x_1 & f[x_1] & & & & f[x_0, x_1, x_2] & = & \dfrac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} \\
 & & \searrow & & & & & \\
 & & & & & \nearrow & & \\
 & & f[x_1, x_2] & = & \dfrac{f[x_2] - f[x_1]}{x_2 - x_1} & & & \\
 & & \nearrow & & & & & \\
x_2 & f[x_2] & & & & & & \\
 & & \vdots & & & & &
\end{array}$$

Veamos en forma más detallada la definición, el cálculo y la interpretación geométrica de de las diferencias divididas.

Sea $f : \mathbb{R} \longrightarrow \mathbb{R}$ una función, la cual suponemos derivable tantas veces cuanto sea necesario. Dados x_0 y x_1, como vimos, el cuociente

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

es la *diferencia dividida de primer orden* de $f(x)$. La diferencia dividida de primer orden $f[x_0,\, x_1]$ es la pendiente de la linea secante al gráfico de $y = f(x)$ uniendo los puntos $(x_0,\, f(x_0))$ y $(x_1,\, f(x_1))$.



Por ejemplo, si $f(x) = \text{sen}(x)$, $x_0 = 1.2$ y $x_1 = 1.3$ entonces

$$f\,[x_0,\, x_1] = \frac{\text{sen}(1.3) - \text{sen}(1.2)}{1.3 - 1.2} = 0.3151909944\,.$$

Las diferencias divididas de primer orden pueden ser pensadas como siendo un análogo numérico de la derivada de $f(x)$.

Si $f(x)$ es una función diferenciable y si x_0 y x_1 son próximos, entonces

$$f'\left(\frac{x_0 + x_1}{2}\right) \approx f\,[x_0,\, x_1]\,, \tag{4.34}$$

esto es, la pendiente de la linea secante uniendo los puntos $(x_0,\, f(x_0))$ y $(x_1,\, f(x_1))$ sobre el gráfico de $y = f(x)$ es aproximadamente igual a la pendiente de la linea tangente en el punto medio del intervalo desde x_0 a x_1 . Además, desde la definición de $f[x_0, x_1]$, haciendo $x_1 \to x_0$, obtenemos

$$f'(x_0) = \lim_{x_1 \to x_0} f[x_0, x_1] \stackrel{\text{def}}{=} f[x_0, x_0] \tag{4.35}$$

En el ejemplo anterior $f(x) = \text{sen}(x)$, $x_0 = 1.2$, $x_1 = 1.3$ y $f'(x) = \cos(x)$, luego $f'\left(\frac{x_0+x_1}{2}\right) = \cos\left(\frac{1.2+1.3}{2}\right) = 0.3153223624$, que es aproximadamente igual a $f\,[x_0,\, x_1] = 0.3153223624$

Como vimos las diferencias divididas de orden superior son definidas recursivamente en término de las diferencias divididas de orden menor.

Dados x_0 , x_1 y x_2 distintos (usualmente tomados en orden creciente), la *diferencia dividida de segundo orden* de $f(x)$ en esos tres puntos es

$$f\,[x_0,\, x_1,\, x_2] = \frac{f\,[x_1,\, x_2] - f\,[x_0,\, x_1]}{x_2 - x_0}\,.$$

Si x_0, x_1, x_2 y x_3 son 4 números reales distintos (usualmente tomados en orden creciente), la *diferencia dividida de tercer orden* de $f(x)$ en esos cuatro puntos es

$$f\,[x_0,\, x_1,\, x_2,\, x_3] = \frac{f\,[x_1,\, x_2,\, x_3] - f\,[x_0,\, x_1,\, x_2]}{x_3 - x_0}\,.$$

En general, si $x_0, x_1, \ldots, x_n$ son $n+1$ números distintos (usualmente tomados en orden creciente), la *diferencia dividida de orden* n de $f(x)$ en esos puntos es

$$f\,[x_0,\, x_1,\, x_2,\, \ldots,\, x_n] = \frac{f\,[x_1,\, \ldots,\, x_n] - f\,[x_0,\, \ldots,\, x_{n-1}]}{x_n - x_0}\,.$$

Observación. Existe una relación entre una diferencia dividida de orden n y la derivada de orden n de la función.

Por ejemplo, si $f(x) = \text{sen}(x)$, $x_0 = 1.2$, $x_1 = 1.21$, y $x_2 = 1.22$, entonces

$$f\,[x_0,\, x_1] = \frac{\text{sen}(1.21) - \text{sen}(1.2)}{1.21 - 1.2} = 0.315190994\,.$$

$$f\,[x_1,\,x_2] = \frac{\operatorname{sen}(1.22) - \operatorname{sen}(1.21)}{1.22 - 1.21} = 0.34833547,$$

$$f\,[x_0,\,x_1,\,x_2] = \frac{f\,[x_1,\,x_2] - f\,[x_0,\,x_1]}{1.22 - 1.2} = -0.46780450.$$

Si miramos $f\,[x_0,\,x_1]$ y $f\,[x_1,\,x_2]$ como siendo una aproximación numérica para la derivada de $f'(x)$ en los puntos medios $\frac{x_0+x_1}{2}$ y $\frac{x_1+x_2}{2}$, respectivamente, entonces

$$\frac{f\,[x_1,\,x_2] - f\,[x_0,\,x_1]}{(\frac{x_1+x_2}{2}) - (\frac{x_0+x_1}{2})} = 2\,f\,[x_0,\,x_1,\,x_2]$$

puede ser pensado como siendo una aproximación numérica para $f''(x_1)$, pues x_1 es el punto medio del intervalo desde x_0 a x_2. Luego la diferencia dividida de segundo orden $f\,[x_0,\,x_1,\,x_2]$ es una aproximación numérica para $\frac{f''(x_1)}{2}$, en otras palabras,

$$f'(x_1) \approx 2f[x_0, x_1, x_2] \tag{4.36}$$

Como $f''(x) = -\operatorname{sen}(x)$, tenemos $\frac{f''(x_1)}{2} = \frac{f''(1.21)}{2} = -0.46780800$, la cual puede ser comparada con el valor $f\,[x_0,\,x_1,\,x_2] = -0.46780450$.

4.5 Cálculo de diferencias divididas

Una manera eficiente de calcular polinomios de interpolación es hacer uso de diferencias divididas en conexión con la fórmula de interpolación de Newton.

Supongamos que son dada una función $f(x)$ y una sucesión $x_0,\,x_1,\,\ldots,\,x_n$ de valores distintos de la variable x.

La diferencias divididas de primer orden asociadas son

$$\begin{aligned} f\,[x_0,\,x_1] &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} \\ f\,[x_1,\,x_2] &= \frac{f(x_2) - f(x_1)}{x_2 - x_1} \\ &\vdots \\ f\,[x_{n-1},\,x_n] &= \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}\,. \end{aligned}$$

Las segundas diferencias divididas asociadas son

$$\begin{aligned} f\,[x_0,\,x_1,\,x_2] &= \frac{f\,[x_1,\,x_2] - f\,[x_0,\,x_1]}{x_2 - x_0} \\ f\,[x_1,\,x_2,\,x_3] &= \frac{f\,[x_2,\,x_3] - f\,[x_1,\,x_2]}{x_3 - x_1} \\ &\vdots \\ f\,[x_{n-2},\,x_{n-1},\,x_n] &= \frac{f\,[x_{n-1},\,x_n] - f\,[x_{n-2},\,x_{n-1}]}{x_n - x_{n-2}}\,. \end{aligned}$$

Note que el número de diferencias divididas decrece a cada etapa hasta que llegamos a la única diferencia dividida de orden n

$$f\left[x_0,\,\ldots,\,x_n\right]=\frac{f\left[x_1,\,\ldots,\,x_n\right]-f\left[x_0,\,\ldots,\,x_{n-1}\right]}{x_n-x_0}\,.$$

Dada una lista de valores de la variable x , digamos $[x_0,\,x_1,\,\ldots,\,x_n]$ y valores de y , digamos $[y_0,\,y_1,\,\ldots,\,y_n]$, donde $y_i=f(x_i)$ para $i=0,\,1,\,\ldots,\,n$, podemos generar todas las diferencias divididas asociadas.

Por ejemplo, dados $x_0,\,x_1,\,x_2,\,x_3,\,x_4$ y $y_0,\,y_1,\,y_2,\,y_3,\,y_4$. Tenemos 4 diferencias divididas de primer orden

$$\frac{y_1-y_0}{x_1-x_0}\,,\quad \frac{y_2-y_1}{x_2-x_1}\,,\quad \frac{y_3-y_2}{x_3-x_2}\,,\quad \frac{y_4-y_3}{x_4-x_3}$$

tres diferencias divididas de segundo orden

$$\frac{\dfrac{y_2-y_1}{x_2-x_1}-\dfrac{y_1-y_0}{x_1-x_0}}{x_2-x_0}\,,\qquad \frac{\dfrac{y_3-y_2}{x_3-x_2}-\dfrac{y_2-y_1}{x_2-x_1}}{x_3-x_1}\,,\qquad \frac{\dfrac{y_4-y_3}{x_4-x_3}-\dfrac{y_3-y_2}{x_3-x_2}}{x_4-x_2}$$

dos diferencias divididas de tercer orden

$$\frac{\dfrac{\dfrac{y_3-y_2}{x_3-x_2}-\dfrac{y_2-y_1}{x_2-x_1}}{x_3-x_1}-\dfrac{\dfrac{y_2-y_1}{x_2-x_1}-\dfrac{y_1-y_0}{x_1-x_0}}{x_2-x_0}}{x_3-x_0}\,,$$

$$\frac{\dfrac{\dfrac{y_4-y_3}{x_4-x_3}-\dfrac{y_3-y_2}{x_3-x_2}}{x_4-x_2}-\dfrac{\dfrac{y_3-y_2}{x_3-x_2}-\dfrac{y_2-y_1}{x_2-x_1}}{x_3-x_1}}{x_4-x_1}$$

una diferencia dividida de cuarto orden

$$\frac{\dfrac{\dfrac{\dfrac{y_4-y_3}{x_4-x_3}-\dfrac{y_3-y_2}{x_3-x_2}}{x_4-x_2}-\dfrac{\dfrac{y_3-y_2}{x_3-x_2}-\dfrac{y_2-y_1}{x_2-x_1}}{x_3-x_1}}{x_4-x_1}-\dfrac{\dfrac{\dfrac{y_3-y_2}{x_3-x_2}-\dfrac{y_2-y_1}{x_2-x_1}}{x_3-x_1}-\dfrac{\dfrac{y_2-y_1}{x_2-x_1}-\dfrac{y_1-y_0}{x_1-x_0}}{x_2-x_0}}{x_3-x_0}}{x_4-x_0}$$

Por ejemplo, tomando un conjunto de puntos (nodos) equiespaciados calculemos un polinomio de grado 6 que aproxima a $\operatorname{sen}(x)$ sobre el intervalo $[0,\,\frac{\pi}{2}]$. Para ello tomamos $h=\pi/12$ y generamos los valores de x_i y los correspondientes valores $y_i=\operatorname{sen}(x_i)$, para $i=0,\ldots,6$. Tenemos entonces $x_0=0$, $x_1=0.2617993878$, $x_2=0.5235987756$, $x_3=0.7853981634$, $x_4=1.047197551$, $x_5=1.308996939$, $x_6=1.570796327$ y $y_0=0$, $y_1=0.2588190451$, $y_2=0.5000000000$, $y_3=0.7071067812$, $y_4=0.8660254037$, $y_5=0.9659258263$, $y_6=1$, esto nos genera las siguiente lista de diferencias divididas

$$\text{diferencias divididas} = [0,\ 0.9886159295,\ -0.1286720769,\ -0.1526656066,\ 0.02059656984,\ 0.006517494278,\ -0.0009653997733]\,.$$

Ejemplo 66 Si $n=1$, tenemos dos puntos $(x_0, f(x_0))$ y $(x_1, f(x_1))$, luego

$$p(x) = f(x_0) + \frac{(f(x_1)-f(x_0))\,(x-x_0)}{x_1-x_0}\,,$$

esta es la ecuación de la linea recta pasando a través de los puntos $(x_0, f(x_0))$ y $(x_1, f(x_1))$.

4.6 Interpolación de Hermite

Dados $n+1$ puntos distintos en el intervalo $[a,b]$, digamos $x_0 = a < x_1 < \cdots < x_n = b$ y $f : [a,b] \longrightarrow \mathbb{R}$ una función $2n+2$ veces derivable con $f^{(2n+2)}(x)$ continua, sean $y_i = f(x_i)$, $i = 0, 1, \ldots, n$.

Problema. Encontrar un polinomio $P(x)$ tal que

$$\begin{cases} P(x_i) &= f(x_i) \\ P'(x_i) &= f'(x_i) \end{cases} \tag{4.37}$$

para $i = 0, 1, \ldots, n$.

Teorema 4.1 *Dados $n+1$ puntos distintos en el intervalo $[a,b]$, digamos $a \leqslant x_0 < x_1 < \cdots < x_n \leqslant b$ y $f : [a,b] \longrightarrow \mathbb{R}$ una función $2n+2$ veces derivable con $f^{(2n+2)}(x)$ continua, sean $y_i = f(x_i)$, $i = 0, 1, \ldots, n$. Entonces existe un único polinomio de grado a lo más $2n+1$, $H_{2n+1}(x)$, tal que*

$$\begin{cases} H_{2n+1}(x_i) &= f(x_i) \\ H'_{2n+1}(x_i) &= f'(x_i) \end{cases} \tag{4.38}$$

para $i = 0, 1, \ldots, n$ donde $H_{2n+1}(x)$ es dado por

$$H_{2n+1}(x) = \sum_{j=0}^{n} f(x_j) H_{n,j}(x) + \sum_{j=0}^{n} f'(x_j) \hat{H}_{n,j}(x) \tag{4.39}$$

con

$$H_{n,j}(x) = \left[1 - 2(x-x_j)L'_{n,j}(x_j)\right] (L_{n,j}(x))^2 \tag{4.40}$$

y

$$\hat{H}_{n,j}(x) = (x-x_j)(L_{n,j}(x))^2\,, \tag{4.41}$$

donde

$$L_{n,j}(x) = \frac{(x-x_0)\cdots(x-x_{j-1})(x-x_{j+1})\cdots(x-x_n)}{(x_j-x_0)\cdots(x_j-x_{j-1})(x_j-x_{j+1})\cdots(x_j-x_n)}$$

son los coeficientes en la interpolación de Lagrange de $f(x)$ *. Además,*

$$f(x) - H_{2n+1}(x) = \frac{(x-x_0)^2\cdots(x-x_1)^2\cdots(x-x_n)^2}{(2n+2)!} f^{(2n+2)}(\xi(x))\,, \tag{4.42}$$

con $x_0 < \xi(x) < x_n$ *.*

Integrando, tenemos

$$\int_{x_0}^{x_n} f(x)dx = \int_{x_0}^{x_n} H_{2n+1}(x) + \int_{x_0}^{x_n} \frac{(x-x_0)^2\cdots(x-x_n)^2}{(2n+2)!} f^{(2n+2)}(\xi(x))dx \tag{4.43}$$

y

$$\int_{x_0}^{x_n} f(x) \approx \int_{x_0}^{x_n} H_{2n+1}(x)\,. \tag{4.44}$$

4.7 Ejemplos resueltos

Problema 4.1 Encontrar el polinomio de interpolación para los puntos $(1,5)\,,\ (2,3)\,,\ (4,2)\,,$ $(5,4)\,,\ (6,\,3)\,.$

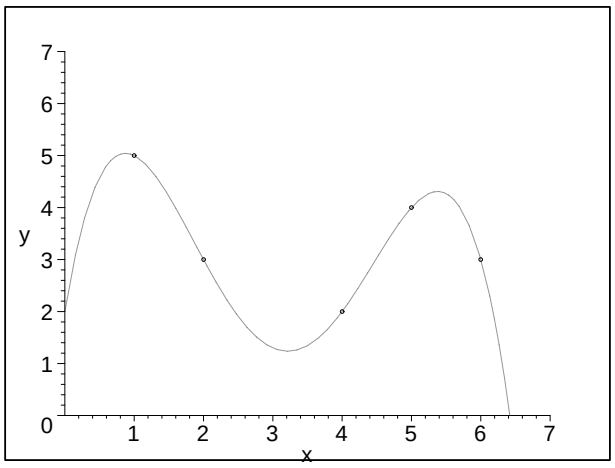
Solución. Tenenos

$$x - \textit{valores} = [1,\,2,\,4,\,5,\,6]\,,\ y - \textit{valores} = [5,\,3,\,2,\,4,\,3]$$

luego, calculando las diferencias divididas, obtenemos

$$p(x) := 7 - 2\,x + \frac{(x-1)\,(x-2)}{2} + \frac{(x-1)\,(x-2)\,(x-4)}{12} - \frac{2\,(x-1)\,(x-2)\,(x-4)\,(x-5)}{15}$$

La siguiente figura muestra el resultado obtenido



Expandiendo los términos del polinomio de interpolación de Newton y ordenando, obtenemos

$$p(x) = -\frac{2}{15}\,x^4 + \frac{101}{60}\,x^3 - \frac{397}{60}\,x^2 + \frac{121}{15}\,x + 2$$

Problema 4.2 Encontrar el polinomio de interpolación de grado 6 para $\text{sen}(x)$ en el intervalo $[0,\, \frac{\pi}{2}]$ y que coincide de $\text{sen}(x)$ en 7 puntos equiespaciados entre 0 y $\frac{\pi}{2}$ inclusive.

Solución. Tenemos entonces que $h = \pi/12 \approx 0.261799387799149$ y generamos los valores de x , obteniendo

$$\begin{aligned} x - \textit{valores} &= [0.,\, 0.261799387799149,\, 0.523598775598298,\, 0.785398163397447,\\ &\quad 1.04719755119660,\, 1.30899693899574,\, 1.57079632679489] \end{aligned}$$

y los valores de y ,

$$\begin{aligned} y - \textit{valores} &= [0.,\, 0.258819045102520,\, 0.499999999999999,\, 0.707106781186547,\\ &\quad 0.866025403784440,\, 0.965925826289066,\, 1.00000000000000] \end{aligned}$$

de donde, relizando el cálculo de las diferencias divididas, otenemos el polinomio

$$\begin{aligned} p(x) &= 0.988615929465368\,x - 0.128672076727087\,x\,(x - 0.261799387799149)\\ &\quad - 0.152665606983500\,x\,(x - 0.261799387799149)\,(x - 0.523598775598298) +\\ &\quad 0.0205965705401489\,x\,(x - 0.261799387799149)\,(x - 0.523598775598298)\\ &\quad (x - 0.785398163397447) + 0.00651749335845159\,x\,(x - 0.261799387799149)\\ &\quad (x - 0.523598775598298)\,(x - 0.785398163397447)\,(x - 1.04719755119660) -\\ &\quad 0.000965398750105050\,x\,(x - 0.261799387799149)\,(x - 0.523598775598298)\\ &\quad (x - 0.785398163397447)\,(x - 1.04719755119660)\,(x - 1.30899693899574) \end{aligned}$$

En los 7 puntos de los nodos el valor dado por la fórmula de interpolación, esencialmente coinciden con los valores de y dados por la función, excepto por un pequeño error. Por ejemplo,

$p(1.047197551) = 0.8660254034$ y $\text{sen}(1.047197551) = 0.8660254037$, $p(1) = 0.8414709003$ y $\text{sen}(1) = 0.8414709848$. Tenemos los siguientes datos

$$\begin{aligned} p(x_i) - f(x_i) &= [[0,\ 0],\ [0.2617993878,\ 0.2588190451],\ [0.5235987756,\ 0.5000000000], \\ &\quad [0.7853981634,\ 0.7071067812],\ [1.047197551,\ 0.8660254038], \\ &\quad [1.308996939,\ 0.9659258263],\ [1.570796327,\ 1.000000000]] \end{aligned}$$

Tenemos

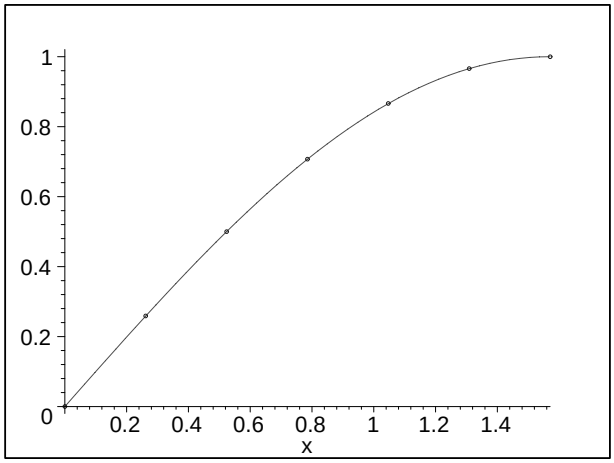


Gráfico del polinomio de interpolación

Desarrollando y ordenando, nos queda

$$\begin{aligned} p(x) &= -0.000965398750105050\,x^6 + 0.01030860538\,x^5 - 0.00209041508\,x^4 - 0.1654864753\,x^3 \\ &\quad - 0.0003303719\,x^2 + 1.000034956\,x \end{aligned}$$

Podemos graficar el error absoluto para $p(x)$ como una aproximación para $\text{sen}(x)$ en el intervalo $[0,\ \frac{\pi}{2}]$.

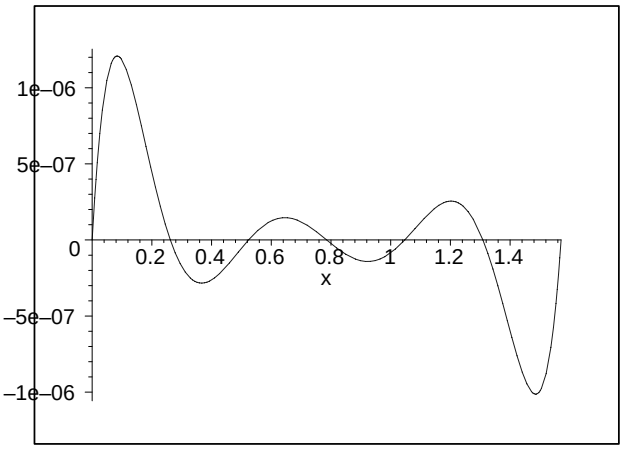


Gráfico del error del polinomio de interpolación

Problema 4.3 De una función $f(x)$ se conocen sus valores y_i en los puntos x_i, $i = 0, 1, 2, 3, 4$. Se desea concer un valor aproximado de $f(2.5)$

x_i	y_i
$x_0 = 0.0$	$y_0 = 1.90$
$x_1 = 0.5$	$y_1 = 2.39$
$x_2 = 1.0$	$y_2 = 2.71$
$x_3 = 1.5$	$y_3 = 2.98$
$x_4 = 2.0$	$y_4 = 3.20$
$x_5 = 3.0$	$y_5 = 3.20$
$x_6 = 3.5$	$y_6 = 2.98$
$x_7 = 4.0$	$y_7 = 2.74$

La forma más directa de calcular $L_7(2.5)$ es escribir

$$L_{7,0}(x) = \frac{(x-x_1)(x-x_2)(x-x_3)(x-x_4)(x-x_5)(x-x_6)(x-x_7)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)(x_0-x_4)(x_0-x_5)(x_0-x_6)(x_0-x_7)}$$

evaluando $L_{7,0}(2.5)$, tenemos $A_0 = L_{7,0}(2.5) = 0.0178571428572$

$$L_{7,1}(x) = \frac{(x-x0)(x-x_2)(x-x_3)(x-x_4)(x-x_5)(x-x_6)(x-x_7)}{(x_1-x0)(x_1-x_2)(x_1-x_3)(x_1-x_4)(x_1-x_5)(x_1-x_6)(x_1-x_7)}$$

evaluando $A_1 = L_{7,1}(2.5) = -0.142857142857$

$$L_{7,2}(x) = \frac{(x-x_0)(x-x_1)(x-x_3)(x-x_4)(x-x_5)(x-x_6)(x-x_7)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)(x_2-x_4)(x_2-x_5)(x_2-x_6)(x_2-x_7)}$$

evaluando, tenemos $A_2 = L_{7,2}(2.5) = 0.500000000000$.

$$L_{7,3}(x) = \frac{(x-x_0)(x-x_1)(x-x_2)(x-x_4)(x-x_5)(x-x_6)(x-x_7)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)(x_3-x_4)(x_3-x_5)(x_3-x_6)(x_3-x_7)}$$

evaluando, tenemos $A_3 = L_{7,3}(2.5) = -1.00000000000$

$$L_{7,4}(x) = \frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)(x-x_5)(x-x_6)(x-x_7)}{(x_4-x_0)(x_4-x_1)(x_4-x_2)(x_4-x_3)(x_4-x_5)(x_4-x_6)(x_4-x_7)}$$

evaluando, tenemos $A_4 = L_{7,4}(2.5) = 1.25000000000$

$$L_{7,5}(x) = \frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)(x-x_4)(x-x_6)(x-x_7)}{(x_5-x_0)(x_5-x_1)(x_5-x_2)(x_5-x_3)(x_5-x_4)(x_5-x_6)(x_5-x_7)}$$

evaluando, tenemos $A_5 = L_{7,5}(2.5) = 0.500000000000$,

$$L_{7,6}(x) = \frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)(x-x_4)(x-x_5)(x-x_7)}{(x_6-x_0)(x_6-x_1)(x_6-x_2)(x_6-x_3)(x_6-x_4)(x_6-x_5)(x_6-x_7)}$$

evaluando, tenemos $A_6 = L_{7,6}(2.5) = -.142857142857$

$$L_{7,7}(x)\frac{(x-x_0)\,(x-x_1)\,(x-x_2)\,(x-x_3)\,(x-x_4)\,(x-x_5)\,(x-x_6)}{(x_7-x_0)\,(x_7-x_1)\,(x_7-x_2)\,(x_7-x_3)\,(x_7-x_4)\,(x_7-x_5)\,(x_7-x_6)}$$

y tenemos $A_7 = L_{7,7}(2.5) = 0.0178571428572$.

Por lo tanto,

$$\begin{aligned} L_7(2.5) &= A_0 \times 1.9 + A_1 \times 2.39 + A_2 \times 2.71 + A_3 \times 2.98 + A_4 \times 3.2 \\ &\quad + A_5 \times 3.2 + A_6 \times 2.98 + A_7 \times 2.74 \\ &= 3.29071428572 \end{aligned}$$

Otra forma, indirecta, es escribir cada uno de los polinomios coeficientes del polinomio de Lagrange, calcular el polinomio de Lagrange en cuestión y después evaluar.

Tenemos

$$\begin{aligned} L_{7,0}(x) &= -0.0158730158730\,x^7 + 0.246031746032\,x^6 + 0.999999999999 \\ &\quad - 1.55158730159\,x^5 - 5.03571428571\,x + 5.12103174603\,x^4 \\ &\quad + 9.70436507936\,x^2 - 9.46825396824\,x^3 \end{aligned}$$

$$\begin{aligned} L_{7,1}(x) &= 0.101587301588\,x^7 - 1.52380952382\,x^6 + 12.8000000001\,x \\ &\quad + 9.16825396832\,x^5 - 38.8571428574\,x^2 - 28.1904761907\,x^4 \\ &\quad + 46.5015873019\,x^3 \end{aligned}$$

$$\begin{aligned} L_{7,2}(x) &= -0.266666666667\,x^7 + 3.86666666667\,x^6 - 16.8000000000\,x \\ &\quad - 22.2000000000\,x^5 + 67.8000000001\,x^2 + 63.8333333334\,x^4 \\ &\quad - 95.2333333335\,x^3 \end{aligned}$$

$$\begin{aligned} L_{7,3}(x) &= 0.355555555554\,x^7 - 4.97777777776\,x^6 + 14.9333333333\,x \\ &\quad + 27.2888888888\,x^5 - 65.2444444442\,x^2 - 73.7777777775\,x^4 \\ &\quad + 101.422222222\,x^3 \end{aligned}$$

$$\begin{aligned} L_{7,4}(x) &= -0.222222222222\,x^7 + 3.00000000000\,x^6 - 6.99999999999\,x \\ &\quad - 15.7222222222\,x^5 + 31.7500000000\,x^2 + 40.2500000000\,x^4 \\ &\quad - 52.0555555555\,x^3 \end{aligned}$$

$$\begin{aligned} L_{7,5}(x) &= 0.0888888888886\,x^7 - 1.11111111111\,x^6 + 1.86666666666\,x \\ &\quad + 5.35555555554\,x^5 - 8.77777777775\,x^2 - 12.6111111111\,x^4 \\ &\quad + 15.1888888888\,x^3 \end{aligned}$$

$$\begin{aligned} L_{7,6}(x) &= -0.0507936507936\,x^7 + 0.609523809523\,x^6 - 0.914285714285\,x \\ &\quad - 2.83174603174\,x^5 + 4.34285714285\,x^2 + 6.47619047618\,x^4 \\ &\quad - 7.63174603174\,x^3 \end{aligned}$$

$$\begin{aligned} L_{7,7}(x) &= 0.00952380952380\,x^7 - 0.109523809524\,x^6 + 0.150000000000\,x \\ &\quad + 0.492857142857\,x^5 - 0.717857142856\,x^2 - 1.10119047619\,x^4 \\ &\quad + 1.27619047619\,x^3 \end{aligned}$$

Luego,

$$\begin{aligned} L_7(x) &= \mathrm{L}_{7,0}(x)\,y_0 + \mathrm{L}_{7,1}(x)\,y_1 + \mathrm{L}_{7,2}(x)\,y_2 + \mathrm{L}_{7,3}(x)\,y_3 \\ &\quad + \mathrm{L}_{7,4}(x)\,y_4 + \mathrm{L}_{7,5}(x)\,y_5 + \mathrm{L}_{7,6}(x)\,y_6 + \mathrm{L}_{7,7}(x)\,y_7 \end{aligned}$$

por lo tanto,

$$\begin{aligned} L_7(x) &= -0.0301587301587\,(x-0.5)\,(x-1.0)\,(x-1.5)\,(x-2.0)\,(x-3.0)\,(x-3.5)\,(x-4.0) \\ &\quad + 0.242793650795\,x\,(x-1.0)\,(x-1.5)\,(x-2.0)\,(x-3.0)\,(x-3.5)\,(x-4.0) \\ &\quad - 0.722666666668\,x\,(x-0.5)\,(x-1.5)\,(x-2.0)\,(x-3.0)\,(x-3.5)\,(x-4.0) \\ &\quad + 1.05955555555\,x\,(x-0.5)\,(x-1.0)\,(x-2.0)\,(x-3.0)\,(x-3.5)\,(x-4.0) \\ &\quad - 0.711111111110\,x\,(x-0.5)\,(x-1.0)\,(x-1.5)\,(x-3.0)\,(x-3.5)\,(x-4.0) \\ &\quad + 0.284444444444\,x\,(x-0.5)\,(x-1.0)\,(x-1.5)\,(x-2.0)\,(x-3.5)\,(x-4.0) \\ &\quad - 0.151365079365\,x\,(x-0.5)\,(x-1.0)\,(x-1.5)\,(x-2.0)\,(x-3.0)\,(x-4.0) \\ &\quad + 0.0260952380952\,x\,(x-0.5)\,(x-1.0)\,(x-1.5)\,(x-2.0)\,(x-3.0)\,(x-3.5) \end{aligned}$$

desarrollando, nos queda

$$\begin{aligned} L_7(x) &= 1.90000000000 - 0.0024126984178\,x^7 + 0.0311746033\,x^6 - 0.1385079373\,x^5 \\ &\quad + 0.211507937\,x^4 + 0.085825394\,x^3 - 0.634825396\,x^2 + 1.2572380950\,x\,. \end{aligned}$$

Luego, el valor pedido es

$$L_7(2.5) = 3.29071428572\,.$$

Ahora podemos calcular una aproximacion de la integral de $f(x)$ usando el polinomio de Lagrange esto es

$$\int_0^4 f(x) \approx \int_0^4 L_7(x) = 11.5714225246\,.$$

Usando la regla de los trapecios con paso constante $h = 0.5$, el valor de la integral es como sigue. Como ya conocemos el valor que tiene $L_7(x)$ en $x = 2.5$, el cual representa un valor aproximado para $f(x)$ en $x = 2.5$. Definimos los nuevos valores x_i e $y_i = f(x_i)$, $i = 0, \ldots, 8$

x_i	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
$f(x_i)$	1.90	2.39	2.71	2.98	3.20	3.29071428572	3.20	2.98	2.74

Tenemos, que

$$Trap_f = \frac{h}{2} \times (y_0 + 2 \times y_1 + 2 \times y_2 + 2 \times y_3 + 2 \times y_4 + 2 \times y_5 + 2 \times y_6 + 2 \times y_7 + y_8)$$

evaluando, obtenemos

$$Trap_f := 11.5353571428\,.$$

El error que hemos cometido en el cálculo comparado con el valor dado por la interpolación de Lagrange

$$\text{Error}(Trap_f) = |Trap_f - \int_0^4 L(x)|$$

nos da

$$\text{Error}(Trap_f) = 0.0360653818\,.$$

Usando la regla de Simpson con paso constante $h = 0.5$, el valor de la integral es como sigue. Como ya conocemos el valor que tiene $L_7(x)$ en $x = 2.5$, el cual representa un valor aproximado para $f(x)$ en $x = 2.5$ y hemos redefinido los nuevos valores x_i e $y_i = f(x_i)$, $i = 0, \ldots, 8$. Tenemos que

$$\text{Simpson}_f = \frac{h}{3} \times (y_0 + 4 \times y_1 + 2 \times y_2 + 4 \times y_3 + 2 \times y_4 + 4 \times y_5 + 2 \times y_6 + 4 \times y_7 + y_8)$$

evaluando, obtenemos

$$\text{Simpson}_f = 11.5704761905\,.$$

El error que hemos cometido en el cálculo de la integral usando el polinomio de Lagrange y la aproximación de Simpson es

$$\text{Error (Simpson}_f) = |\text{Simpson}_f - \int_0^4 L(x)| = 0.0009463341$$

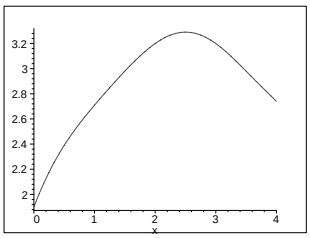


Gráfico de $L_7(x)$ en el intervalo $[0,4]$

Podemos continuar nuestros cálculos, y tenemos *simpson8* ≈ 11.5704761904, con error *Error* ≈ 0.0009463342, *simpson16* ≈ 11.5713597470, con error *Error* ≈ 0.0000627776 y *simpson32* ≈ 11.5714185442, con error *Error* $\approx 0.39804\,10^{-5}$, y para la regla de los trapecios, obtenemos *Trap* ≈ 11.5704929653.

Problema 4.4 Consideremos la funcion $f(x) = x^2 \exp(-x^2)$. Deseamos calcular un valor apromixado para la integral de $f(x)$ en el intervalo $[-2,2]$ usando el polinomio de Lagrange que interpola $f(x)$ en los puntos $x_0 = -2$, $x_1 = -1$, $x_2 = 0$, $x_3 = 1$ y $x_4 = 2$.

Solución. Tenemos que $f(x) = x^2\, e^{-x^2}$

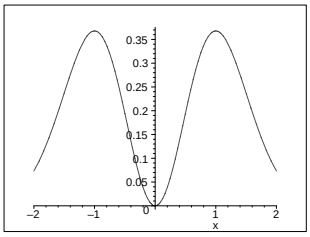


gráfico de f en $[-2,2]$

Como la función es simétrica respecto del origen, se tiene que $f(-1) = f(1) = 0.3678794412$, $f(2) = f(-2) = 0.07326255556$ y $f(0) = 0$.

$$x_0 = -2.0$$
$$x_1 = -1.0$$
$$x_2 = 0.$$
$$x_3 = 1.0$$
$$x_4 = 2.0$$

$$L_4(x) = L_{4,0}(x)\, f(x_0) + L_{4,1}(x)\, f(x_1) + L_{4,2}(x)\, f(x_2) + L_{4,3}(x)\, f(x_3) + L_{4,4}(x)\, f(x_4)$$

Luego,

$$\begin{aligned} L_4(x) &= 0.0732625555548\, L_{4,0}(x) + 0.367879441171\, L_{4,1}(x) + \\ &\quad 0.367879441171\, L_{4,3}(x) + 0.0732625555548\, L_{4,4}(x) \end{aligned}$$

$$L_{4,0}(x) = \frac{(x-x_1)\,(x-x_2)\,(x-x_3)\,(x-x_4)}{(x_0-x_1)\,(x_0-x_2)\,(x_0-x_3)\,(x_0-x_4)}$$

leugo, reemplazando nos queda

$$L_{4,0}(x) = 0.0416666666668\,x^4 - 0.0833333333336\,x^3 - 0.0416666666668\,x^2 + 0.0833333333336\,x$$

$$L_{4,1}(x) = \frac{(x-x_0)\,(x-x_2)\,(x-x_3)\,(x-x_4)}{(x_1-x_0)\,(x_1-x_2)\,(x_1-x_3)\,(x_1-x_4)}$$

$$L_{4,1}(x) = -0.166666666667\,x^4 + 0.166666666667\,x^3 - 0.666666666668\,x + 0.666666666668\,x^2\,.$$

Como $f(x_2) = 0$, no es necesario calcular $L_{4,2}(x)$, pero de todas maneras lo hacemos;

$$L_{4,2}(x) = \frac{(x-x_0)\,(x-x_1)\,(x-x_3)\,(x-x_4)}{(x_2-x_0)\,(x_2-x_1)\,(x_2-x_3)\,(x_2-x_4)}\,.$$

Expandiendo $L_{4,2}(x)$ nos queda

$$L_{4,2}(x) = 0.250000000000\,x^4 + 1.00000000000 - 1.25000000000\,x^2$$

$$L_{4,3}(x) = \frac{(x-x_0)\,(x-x_1)\,(x-x_2)\,(x-x_4)}{(x_3-x_0)\,(x_3-x_1)\,(x_3-x_2)\,(x_3-x_4)}\,,$$

de donde

$$L_{4,3}(x) = -0.166666666666\,x^4 - 0.166666666666\,x^3 + 0.666666666664\,x + 0.666666666664\,x^2$$

$$L_{4,4}(x) = \frac{(x-x_0)\,(x-x_1)\,(x-x_2)\,(x-x_3)}{(x_4-x_0)\,(x_4-x_1)\,(x_4-x_2)\,(x_4-x_3)}$$

$$L_{4,4}(x) = 0.0416666666666\,x^4 + 0.0833333333332\,x^3 - 0.0833333333332\,x - 0.0416666666666\,x^2$$

Luego,

$$L_4(x) = -0.116521267427\,x^4 + 0.4\,10^{-12}\,x^3 + 0.484400708598\,x^2 - 0.1\,10^{-11}\,x$$

Cálculo aproximado de la integral, usando un software, nos da

$$\int_{-2}^{2} f(x) = 0.845450112985\,.$$

Usando el polinomio de Lagrange $L_4(x)$, obtenenos la siguiente aproximación para el valor de la integral

$$\int_{-2}^{2} f(x) \approx \int_{.-2}^{2} L_4(x) = 1.09199822279\,.$$

El error cometido al calcular la integral usando la interpolacion de Lagrange comparado con el resultado de la integrla de $f(x)$ es

$$\text{Error Lagrange} = 0.246548109805\,,$$

el cual es muy grande.

Usando la regla de los trapecios para aproximar la integral de $f(x)$, tenemos

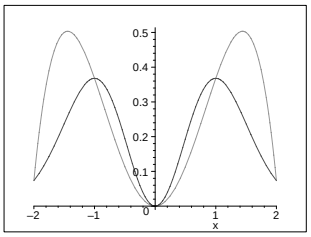
$$h = 1$$
$$y0 = 0.0732625555548$$
$$y1 = 0.367879441171$$
$$y2 = 0$$
$$y3 = 0.367879441171$$
$$y4 = 0.0732625555548$$

luego,

$$Trap_f = 0.809021437895$$

El error cometido es entonces

$$\text{Error Trapecio} = 0.036428675090$$



$$simpson8 = 1.08811418054$$
$$Err = 0.00388404225$$
$$simpson32 = 1.09198305075$$

$$Err = 0.00001517204$$

$$trap = 1.04463387609$$

Problema 4.5 Ejemplo del fenómeno de Runge

Para un ejemplo numérico del fenómeno de Runge consideramos un polinomio de grado 10 interpolando la función $f(x) = \frac{1}{1+25\,x^2}$ en el intervalo $[-1, 1]$, tomamos el conjunto de datos equiespaciado. Este es un ejemplo clásico para ilustrar el problema asociado con el uso de polinomios interpolantes para aproximar una función.

$$f(x) = \frac{1}{1 + 25\,x^2}$$

Tomando $h = \dfrac{1}{5}$, obtenemos

$$p(x) = -\frac{390625}{1768}\,x^{10} + \frac{109375}{221}\,x^8 - \frac{51875}{136}\,x^6 + \frac{54525}{442}\,x^4 - \frac{3725}{221}\,x^2 + 1$$

Notemos que el polinomio interpolante oscila violentamente y no aproxima "muy bien" a la funcion original.

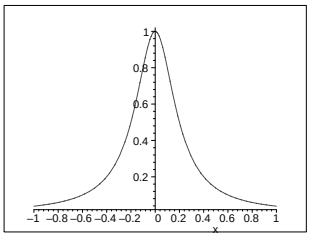


gráfico de $f(x)$

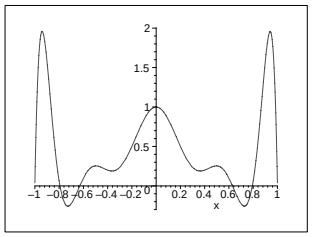
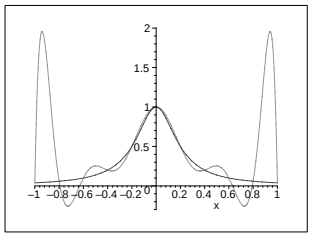


gráfico de $p(x)$



graficos de $f(x)$ y de $p(x)$

4.8 Ejercicios

Problema 4.1 Considere la función $f(x) = e^{-x^2}$ en el intervalo $[-1, 1]$. Construya los polinomios de interpolación de Newton, con puntos equiespaciados y $h = 1/10$ y $h = 1/20$.

Problema 4.2 Encuentre el polinomio de interpolación de Newton para los puntos dados a seguir

$$\left(-2,\, \frac{1}{4}\right),\quad \left(-1,\, \frac{1}{2}\right),\quad (0,\, 1),\quad \left(1,\, \frac{3}{2}\right),\quad \left(2,\, \frac{5}{4}\right).$$

Muestre la gráfica del polinomio de interpolación.

Problema 4.3 Para la función

$$f(x) = \frac{1}{1 + 25\,x^2}$$

muestre explicitamente el polinomio de interpolación de Newton

$$\begin{aligned} p(x) \;=\; & f(x_0) + (x - x_0)\, f[x_0,\, x_1] + (x - x_0)\,(x - x_1)\, f[x_0,\, x_1,\, x_2] \\ & + (x - x_0)\,(x - x_1)\,(x - x_2)\, f[x_0,\, x_1,\, x_2,\, x_3] + \\ & \cdots + (x - x_0)\,(x - x_1)\cdots(x - x_{n-1})\, f[x_0,\, \ldots,\, x_n] \end{aligned}$$

considerando las siguientes listas de valores de x :

$[0, \frac{1}{4}]$; $[0, \frac{1}{4}, \frac{1}{2}]$; $[0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}]$; $[0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1]$; $[0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}]$; $[0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}, \frac{3}{2}]$; $[0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}, \frac{3}{2}, \frac{7}{4}]$; $[0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}, \frac{3}{2}, \frac{7}{4}, 2]$.

Grafique en cada caso el polinomio resultante y compare con el gráfico de la función original Cuál es su conclusión?

Problema 4.4 Para la función $f(x) = \ln(1 + x)$, encuentre explicitamente el polinomio de interpolación de Newton en su forma

$p(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2] + (x - x_0)(x - x_1)(x - x_2) f[x_0, x_1, x_2, x_3] + \cdots + (x - x_0)(x - x_1) \cdots (x - x_{n-1}) f[x_0, \ldots, x_n]$

para las siguientes listas de valores de x

$[0, \frac{1}{5}]$; $[0, \frac{1}{5}, \frac{2}{5}]$; $[0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}]$; $[0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}]$; $[0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1]$.

Grafique en cada caso el polinomio obtenido y la función original, y compare Cuál es su conlcusión?

Problema 4.5 Calcule el polinomio de interpolación de Lagrange para $f(x)$, si se conocen los los siguientes datos

x	3	8	5	1	7
$f(x)$	-459	-18864	-3105	-13	-11263

Usando la aproximación polinomial de f obtenida en (a), calcule un valor aproximado para $\int_1^8 f(x)dx$.

Problema 4.6 Considere la función $f(x) = e^{x^2}$.

a) Usando los puntos $x_0 = 0$, $x_1 = 0.25$, $x_2 = 0.5$, $x_3 = 0.75$, $x_4 = 1.0$. Calcule el polinomio de interpolación de Lagrange para $f(x)$.

b) Exprese una cota para el error en la aproximación de Lagrange obtenida en la parte a). (Simplifique al máximo la expresión obtenida para el error).

c) Usando el polinomio de interpolación de Lagrange obtenido en la parte a), calcular una aproximación para la integral $\int_0^1 f(x)\,dx$, y obtener una buena cota para el error.

Problema 4.7 Considere la función $f(x) = \mathrm{e}^x$.

1. Usando los puntos $x_0 = 0$, $x_1 = 0.25$, $x_2 = 0.5$, $x_3 = 0.75$, $x_4 = 1.0$. Calcule el polinomio de interpolación de Hermite para $f(x)$.

2. Exprese una cota para el error en la aproximación de Hermite obtenida en la parte 1). (Simplifique al máximo la expresión obtenida para el error)

3. Usando el polinomio de interpolación de Hermite obtenido en la parte 1), calcular una aproximación para la integral $\int_0^1 f(x)\,dx$, y obtener una buena cota para el error (compare con el valor exacto de la integral).

Problema 4.8 Sea $f : [a, b] \to \mathbb{R}$ una función suficientemente derivable un número suficiente de veces.

a) Si f es simétrica respecto al origen, es decir, $f(-x) = f(x)$, y se consideran los puntos $x_0, \ldots, x_{2n}$ distribuidos simétricamente, con $x_n = 0$ ¿Son los polinomios

$$L_{2n,k}(x) = \frac{(x - x_0) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_{2n})}{(x - x_0) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_{2n})}$$

simétricos respecto al origen? ¿Es el polinomio de Lagrange $L_{2n}(x)$ simétrico respecto al origen? Justifique su respuesta. Ilustre usando la función $f(x) = e^{-x^2}$, con $x \in [-, 1]$.

b) ¿Qué puede decir de $L_{2n,k}(x)$ y de $L_{2n}(x)$ si f es antisimétrica, respecto al origen es decir, $f(-x) = -f(x)$? Justifique su respuesta. Ilustre usando la función $f(x) = \text{sen}(x)$, con $x \in [-1, 1]$.

Problema 4.9 Sea $f : [0, 4] \longrightarrow \mathbb{R}$ una función suficientemente derivable, la cual satisface $f(0) = 0$, $f(1) = 1$, $f(2) = 0$, $f(3) = -1$, $f(4) = 0$, y $|f^{(5)}(x)| \leqslant \frac{\pi^5}{32}$.

1. Determine el polinomio de Lagrange $L_4(x)$ asociado a f.
2. Estime el error $|f(x) - L_4(x)|$.
3. Usando $L_4(x)$, determine una aproximación y el error cometido para el cálculo de $\int_0^4 f(x)dx$.

Problema 4.10 Considere la función $f(x) = \text{sen}(x)$, con $-2\pi \leqslant x \leqslant 2\pi$.

1. Calcule los desarrollos de Taylor de $f(x)$ en torno a $x = 0$, hasta los ordenes 3, 5, 9 estimando en cada caso el error. Grafique los polinomios de Taylor y compare con la gráfica de $f(x)$.
2. Considere la partición $-2\pi < -\pi < 0 < \pi < 2\pi$, y calcule el polinomio de Lagrange para $f(x)$, estimando el error. Aproxime el valor de $\int_{-2\pi}^{2\pi} f(x)$ usando el polinomio de Lagrange, estimando el error.
3. Considerando la partición anterior, encuentre el polinomio de Hermite para f. Aproxime el valor de $\int_{-2\pi}^{2\pi} f(x)$, estimando el error.
4. ¿Cuál es su conclusión para los itemes 1), 2), y 3)?
5. Refinando la partición anterior, obtenemos una nueva partición, $-2\pi < -3\pi/2 < -\pi < -\pi/2 < 0 < \pi/2 < \pi < 3\pi/2 < 2\pi$. Usando esta nueva partición repita los ejercicios 2) y 3) anteriores.
6. Calcule aprimadamente la derivada de dos maneras distintas para $f(x)$ en $x = \pi$ usando $h = \pi/2$. Calcule también la segunda derivada aproximadamente de $f(x)$ en $x = \pi$. Como se conoce explicitamente las derivadas primeras y segundas de $f(x)$, estime el error que cometió en sus aproximaciones.

Problema 4.11 Considere la función $f(x) = \text{sen}(2\pi x)$.

a) Usando los puntos $x_0 = 0$, $x_1 = 0.25$, $x_2 = 0.5$, $x_3 = 0.75$, $x_4 = 1.0$. Calcule el polinomio de interpolación de Hermite para $f(x)$. Exprese una cota para el error en la aproximación obtenida en la parte. (Simplifique al máximo la expresión obtenida para el error)

b) Usando el polinomio de interpolación de Hermite obtenido en la parte a), calcular una aproximación para la integral $\int_0^1 f(x)\,dx$, y obtener una buena cota para el error.

Problema 4.12 Sea $f : [0,1] \longrightarrow \mathbb{R}$ una función al menos tres veces derivable con continuidad, que satisface $f(0) = 0$, $f'(0) = 1$, $f(0.5) = 1$, $f'(0.5) = 0$, $f(1) = 0$ y $f'(1) = 1$.

1. Encuentre un polinomio P, de menor grado posible, que aproxime a f y que satisface las mismas condiciones que f en los puntos fijados $x_0 = 0$, $x_1 = 0.5$ y $x_2 = 1$.
2. Determine una expresión para error y una buena cota para este (simplifique al máximo).

Problema 4.13 Considere una función $f : [0,4] \longrightarrow \mathbb{R}$, al menos cinco veces derivable, de la cual se sabe que $f(0) = 0$, $f(1) = 1.7183$, $f(2) = 6.3891$, $f(3) = 19.0855$, $f(4) = 53.5982$, y tal que $|f^{(k)}(x)| \leqslant \mathrm{e}^x$ para todo $x \in [0,4]$

1. Encuentre el polinomio de Lagrange $L_4(x)$ correspondiente a f.
2. Calcule $\int_0^4 f(x)dx$ aproximadamente y acote el valor absoluto del error cometido en la aproximación.

Problema 4.14 Sea $f : \mathbb{R} \longrightarrow \mathbb{R}$ una función, al menos 5 veces derivable con continuidad, para la cual conocemos lo siguiente: $f(0) = 0$, $f(0.25) = 1$, $f(0.5) = 0$, $f(0.75) = -1$, $f(1) = 0$, y $|f^{(n)}(x)| \leqslant (2\pi)^n$ para todo $x \in \mathbb{R}$ y todo $n \geqslant 1$.

(a) Usando la información dada, encuentre el polinomio de interpolación Lagrange para $f(x)$.

(b) Calcule una aproximación para $\int_0^1 f(x)dx$ usando el polinomio obtenido en a).

(c) Encuentre una cota para el error que tenemos al aproximar $f(x)$ por el polinomio de Lagrange obtenido en a) en el intervalo $[0,1]$.

(d) Encuentre el error para la aproximación de $\int_0^{0.5} f(x)dx$ por la integral del polinomio de interpolación de Lagrange obtenido en a)

Problema 4.15 a) Sea $f(x) = a_0 + a_1x + \cdots + a_nx^n$ un polinomio de grado n. Dadas las condiciones de interpolación, $x_0 < x_1 < \cdots < x_n$ y $f(x_j) = y_j$, $j = 0, 1, \ldots, n$. Pruebe que el polinomio de interpolación correspondiente, $P_n(x)$, coincide con $f(x)$.

b) Dado $f(x) = 1 + x - x^2$. Verifique a) para este caso particular.

Problema 4.16 Sea $f : [0,4] \longrightarrow \mathbb{R}$ una función suficientemente derivable, la cual satisface $f(0) = 0$, $f(1) = 1$, $f(2) = 0$, $f(3) = -1$, $f(4) = 0$, y $|f^{(5)}(x)| \leqslant \frac{\pi^5}{32}$ para todo $x \in [0,4]$.

(a) Determine el polinomio de Lagrange $L_4(x)$ asociado a f.

(b) Estime el error $|f(x) - L_4(x)|$.

(c) Usando $L_4(x)$, determine una aproximación y el error cometido para el cálculo de $\int_0^4 f(x)dx$.

Problema 4.17 Sea $f : \mathbb{R} \longrightarrow \mathbb{R}$ una función, al menos 5 veces derivable con continuidad, para la cual conocemos lo siguiente: $f(0) = 1$, $f(0.25) = 0$, $f(0.5) = -1$, $f(0.75) = 0$, $f(1) = 1$, y $|f^{(n)}(x)| \leqslant (2\pi)^n$ para todo $x \in \mathbb{R}$ y todo $n \geqslant 1$.

1. Usando la información dada, encuentre el polinomio de interpolación Lagrange para $f(x)$.
2. Calcule una aproximación para $\displaystyle\int_0^1 f(x)dx$ usando el polinomio obtenido en 1).
3. Encuentre una cota para el error que tenemos al aproximar $f(x)$ por el polinomio de Lagrange obtenido en 1) en el intervalo $[0, 1]$.
4. Encuentre el error para la aproximación de $\displaystyle\int_0^{0.5} f(x)dx$ por la integral del polinomio de interpolación de Lagrange obtenido en 1)

Problema 4.18 Considere la función $f : [-1, 1] \longrightarrow \mathbb{R}$ dada por $f(x) = \dfrac{1}{1 + 25x^2}$. Usando divisiones uniformes, calcule y grafique $L_n(n)$, para $n = 2, 3, 4, \ldots, 10$ para esas subdivisiones. Grafique $f(x)$ y $L_n(x)$. Compare Cuál es su conclusión?

Problema 4.19 Una función $f(x)$ satisface $f(0) = 1$, $f(1) = 1$, $f(2) = 2$. Calcule la interpolación de Lagrange para f, y use esta para estimar $f\left(\dfrac{3}{2}\right)$. Grafique.

Problema 4.20 Una función $f(x)$ satisface $f(-1) = 1, f(0) = 2$, $f(2) = 1$. Encuentre la aproximación de f mediante interpolación de Lagrange. Estime $f(1)$. Si es sabido que $|f'''(x)|$, es acotada por 1.12 para $x \in [-1, 2]$. Encuentre el error que se produce al aproximar $f(1)$ por la interpolación de Lagrange. Grafique.

Problema 4.21 Dados los siguientes datos

x	0	1	2	3	4	5
y	2	2	3	5	6	7

Calcule el polinomio de interpolación de Lagrange para esos datos. Grafique el resultado.

Sean $x_0 = 0$, $x_1 = 1.5$, $x_2 = 2$. Suponga que $f(x_0) = 1$, $f(x_1) = 0$ y $f(x_2) = -1$. Calcule las diferencias divididas (todas) y escriba el poinomio de Lagrange. Grafique.

Problema 4.22 Sea $f(x) = x^3$. Sean x_0, x_1, x_2 tres puntos distintos, con $x_0 < x_1 < x_2$. Calcule $f[x_0, x_1, x_2]$.

Problema 4.23 Para una función $f(x)$ se conocen $f[-1] = 2, f[-1, 1] = 1, f[-1, 1, 2] = -2, f[-1, 1, 2, 4] = 2$. Escriba el polinomio de interpolación de grado a lo más 3 con nodos en -1, 1, 2, 4. Usando ese polinomio, estime $f(0)$.

Problema 4.24 Complete la siguiente tabla de diferencias divididas

k	0	1	2	3
x_k	-1	0	1	3
y_k	1	0	-1	2
$f[x_{k-1}, x_k]$	?	-1	-1	?
$f[x_{k-2}, x_{k-1}, x_k]$	?	?	?	?
$f[x_{k-3}, x_{k-2}, x_{k-1}, x_2]$	?	?	?	?

y calcule el polinomio de interpolación de grado a lo más 3 para $(x_k, y_k)\,,\ k = 0, 1, 2, 3\,.$

Problema 4.25 Para aproximar $I(f) = \int_{-1}^{1} f(x)\,dx$ se propone la siguiente fórmula:

$$I_a(f) = Af(0) + Bf'(-1) + Cf'(1).$$

(a) Calcular los valores de las constantes $A\,,\ B$ y C de modo que la fórmula I_a sea exacta en $\mathcal{P}_2(\mathbb{R})$ (el espacio vectorial de los polinomios de grado menor o igual que 2).

(b) Usando la fórmula obtenida en la parte (a), aproxime la integral

$$\int_0^{\pi} \exp(\sin(x))\,dx.$$

Problema 4.26 Se desea construir una función $f : \mathbb{R} \to \mathbb{R}$ tal que $f(x) = 0$ para $x \leqslant 0$ y $f(x) = 1$ para $x \geqslant 1$. Para $x \in [0, 1]$, necesitaremos definir una función s, polinomial por pedazos, que conecte ambas ramas de f, esto es, $s(0) = 0$, $s(1) = 1$, y para tener continuidad de las tangentes, $s'(0) = 0$ y $s'(1) = 0$. Se busca entonces $s : [0, 1] \to \mathbb{R}$ definida por dos polinomios $p, q \in \mathcal{P}_3(\mathbb{R})$ de modo que

$$s(x) = \begin{cases} p(x) & \text{si} \quad x \in [0, \frac{1}{2}], \\ q(x) & \text{si} \quad x \in (\frac{1}{2}, 1], \end{cases}$$

que sea de clase $C^2([0, 1])$ (es decir, que tenga segunda derivada continua en $[0, 1]$), y que interpole a los datos $\{0, \frac{1}{2}, 1\}$ en la malla $\{0, \frac{1}{2}, 1\}$ (ver figura).

(a) Sea m la derivada de s en $\frac{1}{2}$ (o sea $m = s'(\frac{1}{2})$). Encuentre las expresiones de $p(x)$ y $q(x)$ en función de m.

(b) Determine el valor de la constante m de modo que la función $s(x)$ resultante sea de clase $C^2([0, 1])$. En este caso, escriba explícitamente $s(x)$ en el intervalo $[0, 1]$.

Problema 4.27

Problema 4.28

Problema 4.29

Problema 4.30

Problema 4.31

Problema 4.32

Capítulo 5

Derivadas Numéricas

La primera fórmula para calcular derivadas numéricamente proviene del cálculo diferencial, y es dada por

$$f'(x) = \frac{f(x+h) - f(x)}{h} + R(h^2)\,, \tag{5.1}$$

donde $\lim_{h\to 0} R(h^2) = 0\,.$

Para tener mejores aproximaciones podemos usar expansión de Taylor alrededor de x para $f(x+h)$ y $f(x-h)$

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) + \cdots & (5.2)\\ f(x) &= f(x) & (5.3)\\ f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(x) + \cdots & (5.4) \end{aligned}$$

y haciendo alguna combinaciones de esa expansiones.

Por ejemplo, considerando la expansión hasta orden 2, tenemos

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(\xi_1(x)) & (5.5)\\ f(x) &= f(x) & (5.6)\\ f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(\xi_2(x)) & (5.7) \end{aligned}$$

con $\xi_1(x)$ entre x y $x+h$ y $\xi_2(x)$ entre $x-h$ y x. Restando (5.6) de (5.5), nos queda

$$f(x+h) - f(x) = hf'(x) + \frac{h^2}{2}f''(\xi_1(x))\,,$$

de donde, despejando $f'(x)$ obtenemos la fórmula

$$f'(x) = \underbrace{\frac{f(x+h) - f(x)}{h}}_{\text{aproximación}} - \underbrace{\frac{h}{2}f''(\xi_1(x))}_{\text{error}}\,. \tag{5.8}$$

Considerando ahora el desarrollo hasta orden 3, tenemos

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(\theta_1(x)) \tag{5.9}$$

$$f(x) = f(x) \tag{5.10}$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(\theta_2(x)) \tag{5.11}$$

con $\theta_1(x)$ entre x y $x+h$ y $\theta_2(x)$ entre $x-h$ y x.

Restando (5.11) de (5.9), nos queda

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

Si $f'''(x)$ es continua, entonces

$$\frac{f'''(\theta_1(x)) + f'''(\theta_2(x))}{2} = f'''(\theta(x))$$

para algún $\theta(x)$ entre $x-h$ y $x+h$. Luego,

$$f'(x) = \underbrace{\frac{f(x+h) - f(x-h)}{2h}}_{\text{arpoximación}} - \underbrace{\frac{h^2}{6}f'''(\theta(x))}_{\text{error}}\,. \tag{5.12}$$

esta es llamada *fórmula de las diferencias centradas para la primera derivada.*

Para obtener fórmulas para derivadas de orden 2, consideramos desarrollos hasta orden 4.

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) + \frac{h^4}{24}f^{(4)}(\eta_1(x)) \tag{5.13}$$

$$f(x) = f(x) \tag{5.14}$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(x) + \frac{h^4}{24}f^{(4)}(\eta_2(x)) \tag{5.15}$$

con $\eta_1(x)$ entre x y $x+h$ y $\eta_2(x)$ entre $x-h$ y x. Sumando (5.13) y (5.15) y restando (5.14) multiplicada por 2, nos queda

$$f(x+h) - 2f(x) + f(x-h) = h^2 f''(x) + \frac{h^4}{12}\,\frac{f^{(4)}(\eta_1(x)) + f^{(4)}(\eta_2(x))}{2}\,.$$

Si $f^{(4)}(x)$ es continua, entonces

$$\frac{f^{(4)}(\eta_1(x)) + f^{(4)}(\eta_2(x))}{2} = f^{(4)}(\eta(x))$$

con $\eta(x)$ entre $x-h$ y $x+h$. Luego,

$$f(x+h) - 2f(x) + f(x-h) = h^2 f''(x) + \frac{h^4}{12}f^{(4)}(\eta(x))$$

con $\eta(x)$ entre $x-h$ y $x+h$. De donde

$$f''(x) = \underbrace{\frac{f(x+h)-2f(x)+f(x-h)}{h^2}}_{\text{aproximación}} - \underbrace{\frac{h^2}{12}f^{(4)}(\eta(x))}_{\text{error}} \tag{5.16}$$

esta es llamada *fórmula de las diferencias centradas para la segunda derivada.*

Veamos ahora como obtener otras fórmulas para derivadas numérica usando interpolacion de Lagrange.

Sea $f:[a,b]\longrightarrow\mathbb{R}$ tal que $f''(x)$ es continua. Sea $x_0\in\,]a,b[$ y $h\neq 0$ pequeño. Definimos $x_1=x_0+h$, $h\neq 0$ suficientemente pequeño para que $x_1\in[a,b]$. Tenemos la aproximación de Lagrange de grado 1, $L_1(x)$ para $f(x)$ determinada por x_0 y x_1, con su error. Esta es dada por

$$\begin{aligned} f(x) &= L_1(x)+\frac{(x-x_0)(x-x_1)}{2!}f''(\xi(x)) \\ &= f(x_0)\frac{x-x_1}{x_0-x_1}+f(x_1)\frac{x-x_0}{x_1-x_0}+\frac{(x-x_0)(x-x_1)}{2}f''(\xi(x)) \\ &= f(x_0)\frac{(x-x_0-h)}{-h}+f(x_0+h)\frac{(x-x_0)}{h}+\frac{(x-x_0)(x-x_0-h)}{2}f''(\xi(x)) \end{aligned}$$

donde $\xi(x)\in[a,b]$.

Derivando la igualdad anterior, obtenemos

$$\begin{aligned} f'(x) &= \frac{f(x_0+h)-f(x_0)}{h}+\frac{d}{dx}\left(\frac{(x-x_0)(x-x_0-h)}{2}f''(\xi(x))\right) \\ &= \frac{f(x_0+h)-f(x_0)}{h}+\frac{2(x-x_0)-h}{2}f''(\xi(x)) \\ &\quad +\frac{(x-x_0)(x-x_0-h)}{2}\frac{d}{dx}(f''(\xi(x)))\,. \end{aligned}$$

Luego

$$f'(x)\approx\frac{f(x_0+h)-f(x_0)}{h}\,.$$

Ahora $\dfrac{d}{dx}(f''(\xi(x)))=f'''(\xi(x))\xi'(x)$, y no tenemos información sobre esto, luego el error de truncación no lo podemos estimar. Cuando $x=x_0$, el coeficiente que acompaña a $\dfrac{d}{dx}f''(\xi(x))$ se anula, y entonces

$$f'(x_0)=\frac{f(x_0+h)-f(x_0)}{h}-\frac{h}{2}f''(\xi) \tag{5.17}$$

Ahora, si $x_0,\ldots,x_n$ son $n+1$ puntos distintos en el intervalo $I=[a,b]$ y $f(x)$ es $n+1$ veces derivable, con $f^{(n+1)}(x)$ continua. Tenemos que

$$f(x) = \sum_{k=0}^{n} f(x_k) L_{n,k}(x) + \frac{(x-x_0)\cdots(x-x_n)}{(x+1)!} f^{(n+1)}(\xi(x))$$

con $\xi(x) \in [a,b]$. Derivando, obtenemos

$$\begin{aligned} f'(x) &= \sum_{k=0}^{n} f(x_k) L'_{n,k}(x) + \frac{d}{dx}\left(\frac{(x-x_0)\cdots(x-x_n)}{(n+1)!} f^{(n+1)}(\xi(x))\right) \\ &= \sum_{k=0}^{n} f(x_k) L'_{n,k}(x) + \frac{d}{dx}\left(\frac{(x-x_0)\cdots(x-x_n)}{(n+1)!}\right) \cdot f^{(n+1)}(\xi(x)) \\ &\quad + \frac{(x-x_0)\cdots(x-x_n))}{(n+1)!} \frac{d}{dx} f^{(n+1)}(\xi(x)) \end{aligned}$$

el factor del último sumando se anula en cuando $x = x_j$, para $j = 0, \ldots, n$.

Luego

$$f'(x_j) = \sum_{k=0}^{n} f(x_k) L'_{n,k}(x) + \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \prod_{\substack{i=0 \\ i \neq j}}^{n} (x_j - x_i) \tag{5.18}$$

Ejemplo. Fórmula con tres modos.

En este caso, son dados x_0, x_1 y x_2, y $f(x)$ una función 3 veces derivable, con $f'''(x)$ continua. Entonces

$$\begin{aligned} L_{2,0}(x) &= \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}, & L'_{2,0}(x) &= \frac{2x-x_1-x_2}{(x_0-x_1)(x_0-x_2)} \\ L_{2,1}(x) &= \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}, & L'_{2,1}(x) &= \frac{2x-x_0-x_2}{(x_1-x_0)(x_1-x_2)} \\ L_{2,2}(x) &= \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}, & L'_{2,2}(x) &= \frac{2x-x_0-x_1}{(x_2-x_0)(x_2-x_1)} \end{aligned}$$

Reemplazando en la fórmula (5.18), nos queda

$$\begin{aligned} f'(x_j) &= f(x_0)\frac{2x_j-x_1-x_2}{(x_0-x_1)(x_0-x_2)} + f(x_1)\frac{2x_j-x_0-x_2}{(x_1-x_0)(x_1-x_2)} \\ &\quad + f(x_2)\frac{2x_j-x_0-x_1}{(x_2-x_0)(x_2-x_1)} + \frac{1}{6} f^{(3)}(\xi_j) \prod_{\substack{i=0 \\ i \neq j}}^{2} (x_j - x_i) \end{aligned}$$

donde $j = 0, 1, 2$ y ξ_j depende de x_j.

Ahora, tomamos $a_1 = x_0 + h$, $x_2 = x_1 + h = x_0 + 2h$ ($h \neq 0$). Para $x_j = x_0$, tenemos que $x_1 = x_0 + h$ y $x_2 = x_0 + 2h$. Luego

$$f'(x_0) = \frac{1}{h}\left(-\frac{3}{2}f(x_0) + 2f(x_1) - \frac{1}{2}f(x_2)\right) + \frac{h^2}{3}f^{(3)}(\xi_0)\,,$$

para $x_j = x_1$, tenemos

$$f'(x_1) = \frac{1}{h}\left(-\frac{1}{2}f(x_0) + \frac{1}{2}f(x_2)\right) - \frac{h^2}{6}f^{(3)}(\xi_1)$$

y para $x_j = x_2$,

$$f'(x_2) = \frac{1}{h}\left(\frac{1}{2}f(x_0) - 2f(x_1) + \frac{3}{2}f(x_2)\right) + \frac{h^2}{3}f^{(3)}(\xi_2)\,.$$

Como $x_1 = x_0 + h$ y $x_2 = x_0 + 2h$, tenemos

$$\begin{aligned}
f'(x_0) &= \frac{1}{h}\left(-\frac{3}{2}f(x_0) + 2f(x_0+h) - \frac{1}{2}f(x_0+2h)\right) + \frac{h^2}{3}f^{(3)}(\xi_0)\\
f'(x_0+h) &= \frac{1}{h}\left(-\frac{1}{2}f(x_0) + \frac{1}{2}f(x_0+2h)\right) - \frac{h^2}{6}f^{(3)}(\xi_1)\\
f'(x_0+2h) &= \frac{1}{h}\left(\frac{1}{2}f(x_0) - 2f(x_0+h) + \frac{3}{2}f(x_0++2h)\right) + \frac{h^2}{6}f^{(3)}(\xi_2)\,,
\end{aligned}$$

es decir, reemplazamdo x_0 para x_0+h en la segunda ecuación y x_0 para x_0+2h en la tercera ecuación obtenemos

a) $f'(x_0) = \dfrac{1}{2h}\left(-3f(x_0) + 4f(x_0+h) - f(x_0+2h) + \dfrac{h^2}{3}f^{(3)}(\xi_0)\right.$

b) $f'(x_0) = \dfrac{1}{2h}\left(-f(x_0-h) + f(x_0+h)\right) - \dfrac{h^2}{6}f^{(3)}(\xi_1)$

c) $f'(x_0) = \dfrac{1}{2h}(f(x_0-2h) - 4f(x_0-h) + 3f(x_0)) + \dfrac{h^2}{3}f^{(3)}(\xi_2)$

De modo análogo, puede deducirse las fórmulas

d) $f'(x_0) = \dfrac{1}{12h}\left(f(x_0-2h) - 8f(x_0-h) + 8f(x_0+h) - f(x_0+2h)\right) + \dfrac{h^4}{30}f^{(5)}(\xi)$

e) $f'(x_0) = \dfrac{1}{12h}\left(-25f(x_0) + 48f(x_0+h) - 36f(x_0+2h) + 16f(x_0+3h) - 3f(x_0+4h)\right))+$
$\dfrac{h^4}{5}f^{(5)}(\xi)$

5.1 Ejercicios

Problema 5.1 Usando las fórmulas anteriores (todas) calcule $f'(2.0)$, donde $f(x) = xe^x$, primero con $h = 0.1$ y despues con $h = -0.1$. Analice los resultados en comparación con la derivada $f'(x)$ calculada en forma exacta y evaluada en $x = 2.0$.

Problema 5.2 Para calcular numéricamente la derivada de una función $f(x)$, podemos usar las siguiente aproximaciones

(i) $f'(x_0) \approx \dfrac{f(x_0+h) - f(x_0)}{h}$,

(ii) $f'(x_0) \approx \dfrac{f(x_0+h) - f(x_0-h)}{2h}$.

(a) Considerando $f(x) = \operatorname{sen}(x)$, $x_0 = \pi/4$, y $h_n = \dfrac{1}{10^n}$, con $n = 1, \ldots, 5$, estime $f'(\pi/4)$, usando (i) y (ii) anteriores.

(b) En ambas fórmulas (a) y (b) anteriores existe pérdidad de dígitos significativos (es decir, son numéricamente inestables) pues existe resta de números parecidos ¿Cuál de ellas es más inestable? Justifique su respuesta en forma teórica y numérica.

Problema 5.3 Obtenga los valores de a_1 y a_2 de modo que la fórmula de derivación numérica $f'(1/2) \approx a_1 f(0) + a_2 f(1/2)$ sea exacta para las funciones 1, x.

Problema 5.4

Problema 5.5

Problema 5.6

Problema 5.7

Problema 5.8

Poner lista de fórmulas para derivadas numéricas

Capítulo 6

Spline Cúbicos

Sea $f : [a, b] \longrightarrow \mathbb{R}$ y sean $a = x_0 < x_1 < \cdots < x_n = b$. Un spline cúbico $S(x)$ para $f(x)$ es una función que satisface

1. $S(x)$ es un polinomio cúbico en cada intervalo $[x_j, x_{j+1}]$. Notación $S\big|_{[x_j,x_{j+1}]} = S_j$, $j = 0, 1, \ldots, n-1$,
2. $S(x_j) = f(x_j)$, $j = 0, 1, \ldots, n$,
3. $S_j(x_{j+1}) = S_{j+1}(x_{j+1})$, $j = 0, 1, \ldots, n-2$,
4. $S'_j(x_{j+1}) = S'_{j+1}(x_{j+1})$, $j = 0, 1, \ldots, n-2$,
5. $S''_j(x_{j+1}) = S''_{j+1}(x_{j+1})$, $j = 0, 1, \ldots, n-2$,
6. una de las siguientes condiciones de frontera se cumple
   (a) $S''(x_0) = S''(x_n) = 0$, *frontera libre o natural.*
   (b) $S'(x_0) = f'(x_0)$ y $S'(x_n) = f'(x_n)$, *frontera sujeta.*

6.1 Construcción de Spline cúbicos

Consideremos los polinomios cúbicos

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3, \tag{6.1}$$

$j = 0, 1, \ldots, n-1$.

Ahora, aplicando la condición (3), tenemos

$$\begin{aligned} a_{j+1} &= S_{j+1}(x_{j+1}) = S_j(x_{j+1}) \\ &= a_j + b_j \underbrace{(x_{j+1} - x_j)}_{h_j} + c_j \underbrace{(x_{j+1} - x_j)^2}_{h_j^2} + d_j \underbrace{(x_{j+1} - x_j)^3}_{h_j^3} \end{aligned} \tag{6.2}$$

$j = 0, 1, \ldots, n-2$.

Usaremos la notación $h_j = x_{j+1} - x_j$, para $j = 0, 1, \ldots, n-1$, y

$$a_n = f(x_n)\,. \tag{6.3}$$

La ecuación (6.2) se escribe entonces como

$$a_{j+1} = a_j + b_j h_j + c_j h_j^2 + d_j h_j^3\,, \quad j = 0, 1, \ldots, n-1. \tag{6.4}$$

Ahora definamos

$$b_n = S'(x_n)\,. \tag{6.5}$$

Tenemos entonces

$$S'_j(x) = b_j + 2c_j(x - x_j) + 3d_j(x - x_j)^2\,, \tag{6.6}$$

Luego, $S'_j(x_j) = b_j$, $j = 0, 1, \ldots, n-1$.

Aplicando la condición (4) obtenemos

$$b_{j+1} = S'_{j+1}(x_{j+1}) = S'_j(x_{j+1}) = b_j + 2c_j h_j + 3d_j h_j^2 \tag{6.7}$$

$j = 0, 1, \ldots, n-1$.

Definamos

$$c_n = \frac{S''(x_n)}{2} \tag{6.8}$$

aplicando la condición (5), obtenemos

$$c_{j+1} = c_j + 3d_j h_j\,, \quad j = 0, 1, \ldots, n-1 \tag{6.9}$$

despejando d_j de esta ecuación obtenemos

$$d_j = \frac{c_{j+1} - c_j}{3h_j} \tag{6.10}$$

Reemplazando en la ecuación para a_{j+1} (6.4) y para b_{j+1} (6.7), nos queda

$$a_{j+1} = a_j + b_j h_j + \frac{h_j^2}{3}(2c_j + c_{j+1}) \tag{6.11}$$

$$b_{j+1} = b_j + h_j(c_j + c_{j+1}) \tag{6.12}$$

De la ecuación (6.11) despejamos b_j , y nos queda

$$b_j = \frac{1}{h_j}(a_{j+1} - a_j) - \frac{h_j}{3}(2c_j + c_{j+1}) \tag{6.13}$$

Para $j-1$ obtenemos

$$b_{j-1} = \frac{1}{h_{j-1}}(a_j - a_{j-1}) - \frac{h_{j-1}}{3}(2c_{j-1} + c_j)\,. \tag{6.14}$$

Reemplazando en la ecuación para b_{j+1}, con el índice reducido en 1, obtenemos el siguiente sistema de ecuaciones lineales

$$h_{j-1}c_{j-1} + 2(h_{j-1} + h_j)c_j + h_j c_{j+1} = \frac{3}{h_j}(a_{j+1} - a_j) - \frac{3}{h_{j-1}}(a_j - a_{j-1}) \tag{6.15}$$

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

Notemos que h_j y a_j son conocidos. Los otros coeficientes b_j y d_j, $j = 0, 1, \ldots, n-1$ se obtienen de

$$b_j = \frac{1}{h_j}(a_{j+1} - a_j) - \frac{h_j}{3}(2c_j + c_{j+1}) \tag{6.16}$$

$$d_j = \frac{c_{j+1} - c_j}{3h_j}\,. \tag{6.17}$$

Teorema 6.1 *Sea $f : [a, b] \longrightarrow \mathbb{R}$ y $a = x_0 < x_1 < \cdots < x_n = b$. Entonces f tiene un único spline cúbico natural en los nodos $x_0, x_1, \ldots, x_n$.*

Demostración. En este caso las condiciones de frontera significan que $c_n = \frac{S''(x_n)}{2} = 0$ y que $0 = S''(x_0) = 2c_0 + 6d_0(x_0 - x_0)$, luego $c_0 = 0$.

De las dos ecuaciones $c_0 = 0$ y $c_n = 0$ junto con las ecuaciones lineales (6.15) se tiene el sistema de ecuaciones lineales $Ax = b$, donde A es la matriz $(n+1) \times (n+1)$ dada por

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ h_0 & 2(h_0 + h_1) & h_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & h_1 & 2(h_1 + h_2) & h_2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & h_{n-2} & 2(h_{n-2} + h_{n-1}) & h_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

y los vectores b y x son dados por

$$b = \begin{pmatrix} 0 \\ \frac{3}{h_1}(a_2 - a_1) - \frac{3}{h_0}(a_1 - a_0) \\ \vdots \\ \vdots \\ \frac{3}{h_{n-1}}(a_n - a_{n-1}) - \frac{3}{h_{n-2}}(a_{n-1} - a_{n-2}) \\ 0 \end{pmatrix} \quad \text{y} \quad x = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ \vdots \\ c_n \end{pmatrix}$$

La matriz A es diagonal dominante, luego existe solución única $c_0, c_1, \ldots, c_n$ del sistema.

Teorema 6.2 *Si* $f : [a,b] \longrightarrow \mathbb{R}$ *es derivable en* a *y* b, *y* $a = x_0 < x_1 < \cdots < x_n = b$. *Entonces existe un único spline cúbico con las condiciones de frontera* $S'(a) = f'(a)$ *y* $S'(b) = f'(b)$.

Demostración. Como $f'(a) = S'(a) = S'(x_0) = b_0$. Tenemos,

$$b_j = \frac{1}{h_j}(a_{j+1} - a_j) - \frac{h_j}{3}(2c_j + c_{j+1}) \tag{6.18}$$

con $j = 0$ nos queda la ecuación

$$b_0 = f'(a) = \frac{1}{h_0}(a_1 - a_0) - \frac{h_0}{3}(2c_0 + c_1)$$

y de aquí obtenemos que

$$2h_0c_0 + h_0c_1 = \frac{3}{h_0}(a_1 - a_0) - 3f'(a).$$

De modo análogo

$$f'(b) = b_n = b_{n-1} + h_{n-1}(c_{n-1} + c_n)$$

y con $j - 1$ la ecuación (6.18) nos da que

$$\begin{aligned} f'(b) &= \frac{a_n - a_{n-1}}{h_{n-1}} - \frac{h_{n-1}}{3}(2c_{n-1} + c_n) + h_{n-1}(c_{n-1} + c_n) \\ &= \frac{a_n - a_{n-1}}{h_{n-1}} + \frac{h_{n-1}}{3}(c_{n-1} + 2c_n) \end{aligned}$$

y que

$$h_{n-1}c_{n-1} + 2h_{n-1}c_n = 3f'(b) - \frac{3}{h_{n-1}}(a_n - a_{n-1}).$$

El sistema de ecuaciones lineales (6.15) junto con las ecuaciones

$$2h_0c_0 + h_0c_1 = \frac{3}{h_0}(a_1 - a_0) - 3f'(a)$$

y

$$h_{n-1}c_{n-1} + 2h_{n-1}c_n = 3f'(b) - \frac{3}{h_{n-1}}(a_n - a_{n-1})$$

determinan un sistema de ecuaciones lineales $Ax = b$, donde

$$A = \begin{pmatrix} 2h_0 & h_0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ h_0 & 2(h_0+h_1) & h_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & h_1 & 2(h_1+h_2) & h_2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & h_{n-2} & 2(h_{n-2}+h_{n-1}) & h_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & h_{n-1} & 2h_{n-1} \end{pmatrix}$$

$$b = \begin{pmatrix} \frac{3}{h_0}(a_1-a_0) - 3f'(a) \\ \frac{3}{h_1}(a_2-a_1) - \frac{3}{h_0}(a_1-a_0) \\ \vdots \\ \vdots \\ \frac{3}{h_{n-1}}(a_n-a_{n-1}) - \frac{3}{h_{n-2}}(a_{n-1}-a_{n-2}) \\ 3f'(b) - \frac{3}{h_{n-1}}(a_n-a_{n-1}) \end{pmatrix} \quad \text{y} \quad x = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ \vdots \\ c_n \end{pmatrix}$$

La matriz A es diagonal dominante, luego existe solución única $c_0, c_1, \ldots, c_n$ del sistema

6.2 Otra forma de construir spline cúbicos naturales

Sean $z_i = S''(x_i)$, para $0 \leqslant i \leqslant n$. Sobre $[x_i, x_{i+1}]$, debemos tener que $S_i''(x)$ es una interpolación lineal y $S_i''(x_i) = z_i$, $S_i''(x_{i+1}) = z_{i+1}$. Luego podemos escribir

$$S_i''(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}}\, z_i + \frac{x - x_i}{x_{i+1} - x_i}\, z_{i+1}\,.$$

Integrando $S''(x)$ dos veces, obtenemos

$$S_i(x) = (x_{i+1} - x)^3\, \frac{z_i}{6h_i} + (x - x_i)^3\, \frac{z_{i+1}}{6h_i} + cx + d \tag{6.19}$$

donde $h_i = x_{i+1} - x_i$ y c, d son las constantes de integración. Ahora, imponiendo las condiciones

$$\begin{aligned} S_i(x_i) &= f(x_i) = y_i \\ S_i(x_{i+1}) &= y_{i+1}\,. \end{aligned}$$

Obtenemos las ecuaciones

$$\begin{cases} h_i^3 \dfrac{z_i}{6h_i} + cx_i + d = y_i \\ \\ h_i^3 \dfrac{z_i}{6h_i} + cx_{i+1} + d = y_{i+1} \end{cases}$$

de donde $c = \dfrac{y_{i+1} - y_i}{h_i} - \dfrac{(z_{i+1} - z_i)}{6} h_i$ y $d = \dfrac{y_i x_{i+1} - y_{i+1} x_i}{h_i} + h_i \dfrac{x_i z_{i+1} - x_{i+1} z_i}{6}$.

Reemplazando en la ecuación (6.19) nos queda

$$S_i(x) = (x_{i+1}-x)^3\frac{z_i}{6h_i} + \frac{(x-x_i)^3 z_{i+1}}{6h_i} + \left(\frac{(y_{i+1}-y_i)}{h_i} - \frac{(z_{i+1}-z_i)}{6}h_i\right)x$$
$$+\frac{y_i x_{i+1} - y_{i+1}x_i}{h_i} + h_i\,\frac{x_i z_{i+1} - x_{i+1}z_i}{6} \tag{6.20}$$

Para encontrar z_i y z_{i+1} imponemos la condición $S'_{i-1}(x_i) = S'_i(x_i)$.

Ahora, derivando (6.20) y reemplazando obtenemos

$$S'_i(x_i) = -\frac{1}{3}h_i z_i - \frac{1}{6}h_i z_{i+1} + b_i$$

y

$$S'_{i-1}(x_i) = \frac{1}{6}h_{i-1}z_{i-1} + \frac{1}{3}h_{i-1}z_i + b_i$$

imponiendo la condición $S'_{i-1}(x_i) = S'_i(x_i)$ nos queda

$$h_{i-1}z_{i-1} + 2(h_{i-1}+h_i)z_i + h_i z_{i+1} = 6(b_i - b_{i-1}), \qquad i = 1,\ldots,n-1$$

Para encontrar z_i, $(1 \leqslant i \leqslant n-1)$, considerando que $z_0 = z_n = 0$, se tiene un sistema de ecuaciones, simétrico, tridiagonal, diagonal dominante, de la forma

$$\begin{pmatrix} u_1 & h_1 & 0 & 0 & 0 & \cdots & 0 \\ h_1 & u_2 & h_2 & 0 & 0 & \cdots & 0 \\ 0 & h_2 & u_3 & h_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & h_{n-3} & u_{n-2} & h_{n-2} \\ 0 & 0 & \cdots & 0 & 0 & h_{n-2} & u_{n-1} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-2} \\ z_{n-1} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-2} \\ v_{n-1} \end{pmatrix}$$

donde

$$\begin{cases} h_i = t_{i+1} - t_i \\ u_i = 2(h_i + h_{i-1}) \\ b_i = \dfrac{6}{h_i}(x_{i+1} - x_i), \quad S_j(t_j) = x_j \\ v_i = b_i - b_{i-1} \end{cases}$$

6.3 Ejercicios

Problema 6.1 Sea $f : [a,b] \longrightarrow \mathbb{R}$ un polinomio cúbico. Denote por S_1 y S_2, los spline cúbicos libre y sujeto determinados por f ¿S_1 y S_2 coinciden con f? Justique su respuesta.

Problema 6.2 Determine los valores de a, b y c de modo que la función

$$f(x) = \begin{cases} x^3, & x \in [0,1] \\ \frac{1}{2}(x-1)^3 + a(x-1)^2 + b(x-1) + c, & x \in [1,3] \end{cases}$$

es un spline cúbico.

Problema 6.3 Sea $s(x)$ el spline cúbico natural con nodos $(0,2)$, $(1,0)$, $(2,3)$ y $(3,1)$. Suponga que

$$s(x) = \begin{cases} 2 - \frac{11}{3}x + \frac{5}{3}x^3 & \text{si} \quad 0 \leqslant x < 1 \\ 7 - \frac{56}{3}x + 15x^2 - \frac{10}{3}x^3 & \text{si} \quad 1 \leqslant x < 2 \\ -33 + \frac{124}{3}x + Ax^2 + Bx^3 & \text{si} \quad 2 \leqslant x \leqslant 3 \end{cases}$$

Encuentre A y B.

Problema 6.4 Si $s(x) = 0$ para $x < 2$ y $s(x) = (x-2)^3$ para $x \geqslant 2$ ¿es $s(x)$ un spline cúbico? Justifique.

Problema 6.5 Suponga que

$$s(x) = \begin{cases} x^3 + ax^2 - 4x + c & 0 \leqslant x \leqslant 2 \\ -x^3 + ax^2 + bx + 34, & 2 \leqslant x \leqslant 4 \end{cases}$$

Encuentre las constantes a, b y c de modo que $s(x)$ es dos veces continuamente derivable en $[0,4]$ ¿Es $s(x)$ para esas constantes un spline cúbico?

Problema 6.6 Sea

$$s(x) = \begin{cases} 1 - x + ax^2 + x^3 & \text{si} \quad 0 \leqslant x \leqslant 1 \\ 3 + bx + cx^2 - x^3 & \text{si} \quad 1 < x \leqslant 2 \end{cases}$$

Determine a, b, y c de modo que $s(x)$ sea un spline cúbico natural sobre el intervalo $[0,2]$

Problema 6.7 Considere la función $f : [-1,1] \longrightarrow \mathbb{R}$ dada por $f(x) = \dfrac{1}{1+25x^2}$. Usando divisiones uniformes, calcule y grafique $L_n(x)$, para $n = 2,3,4,\ldots,10$. Calcule y grafique las interpolaciones por spline cúbicos naturales y libres (imponga las condiciones adecuadas en cada caso) para esas subdivisiones. Grafique $f(x)$, $L_n(x)$ y $s_n(x)$. Compare ¿Cuál es su conclusión?

Problema 6.8

Problema 6.9

Problema 6.10

Problema 6.11

Problema 6.12

Problema 6.13

Problema 6.14

Capítulo 7

Ajuste de Curvas

7.1 Ajuste de curvas

Sean $\{x_1, x_2, \ldots, x_n\}$ un conjunto de puntos, con $x_1 < x_2 < \cdots < x_n$, y sean $\{y_1, y_2, \ldots, y_n\}$ un conjunto de datos, es decir, tenemos los puntos $(x_1, y_1)\,, (x_2, y_2), \ldots, (x_n, y_n)\,.$

Problema 1 Determinar una función que relacione las variables.

En general esto no es posible hacer en forma exacta. Buscamos entonces una función $f(x)$ tal que

$$f(x_k) = y_k + e_k\,, \tag{7.1}$$

donde e_k es el error de medición.

Problema 2 Cómo encontramos la mejor aproximación, es decir, que pase lo más cerca posible y no necesariamente sobre esos puntos? Para responder a esta pregunta, hay que considerar los errores (los cuales son también llamados *desviaciones* o *residuos*).

Para ello debemos encontrar $f(x)$ tal que

$$e_k = f(x_k) - y_k\,, \qquad 1 \leqslant k \leqslant n\,, \tag{7.2}$$

donde e_k son llamados errores o desviaciones o residuos, sean en algún sentido "pequeños".

Existen varias formas que pueden usarse en (7.2) para medir la distancia entre la curva $y = f(x)$ y los datos.

a) **Error máximo**

$$E_\infty(f) = \max\{|f(x_k) - y_k| : 1 \leqslant k \leqslant n\}\,.$$

b) **Error medio**

$$E_1(f) = \frac{1}{n}\sum_{k=1}^{n} |f(x_k) - y_k|\,.$$

c) **Error medio cuadrático**

$$E_2(f) = \left(\frac{1}{n} \sum_{k=1}^{n} |f(x_k) - y_k|^2 \right)^{1/2} .$$

Observación. Minimizar $E_2(f)$ equivale a minimizar

$$E_2(f)^2 = \frac{1}{n} \sum_{k=1}^{n} |f(x_k) - y_k|^2 ,$$

y esto equivale a minimizar

$$F(f) = \sum_{k=1}^{n} |f(x_k) - y_k|^2 .$$

Ejemplo 67 consideremos la función $y = f(x) = 8.6 - 1.6x$ y el conjunto de a ajustar datos $(-1, 10), (0, 9), (1, 7), (2, 5), (3, 4), (4, 3), (5, 0), (6, -1)$. Tenemos $e_k = f(x_k) - y_k$. Luego

$$\begin{aligned} E_\infty(f) &= \max\{0.2, 0.4, 0, 0.4, 0.2, 0.8, 0.6, 0\} = 0.8 \\ E_1(f) &= \frac{1}{8} \times 2.6 \approx 0.325 \\ E_2(f) &= \left(\frac{14}{8} \right)^{1/2} \approx 0.4183300133 . \end{aligned}$$

7.2 Ajuste por rectas: recta de regresión

En este caso, tenemos los datos

$$\{(x_k, y_k)\}_{k=1}^{n}, \quad \text{con} \quad x_1 < x_2 < \cdots < x_n .$$

La *recta de regresión* o *recta óptima*, en el sentido de los mínimos cuadrados es la recta de ecuación $y = f(x) = Ax + B$ que minimiza el error cuadrático medio $E_2(f)$.

Problema 3 Determinar A y B.

Tenemos $y = Ax + B$ y sea $d_k =$ distancia entre (x_k, y_k) y $(x_k, Ax_k + B)$

DIBUJO

Esta distancia es

$$d_k = |Ax_k + B - y_k| .$$

Sea $E(A, B) = \sum_{k=1}^{n} (Ax_k + B - y_k)^2 = \sum_{k=1}^{n} d_k^2$.

Problema 4 Determinar el valor mínimo de $E(A,B)$.

Para esto, debemos resolver las ecuaciones

$$\begin{cases} \dfrac{\partial E}{\partial A} &= 0 \\ \dfrac{\partial E}{\partial B} &= 0\,, \end{cases}$$

llamadas *ecuaciones normales de Gauss*. Tenemos

$$\frac{\partial E}{\partial A} = \sum_{k=1}^{n} 2(Ax_k + B - y_k)x_k = \sum_{k=1}^{n} 2(Ax_k^2 + Bx_k - x_k y_k)\,,$$

luego $\dfrac{\partial E}{\partial A} = 0$ si y sólo si

$$\left(\sum_{k=1}^{n} x_k^2\right) A + \left(\sum_{k=1}^{n} x_k\right) B = \sum_{k=1}^{n} x_k y_k\,.$$

Ahora

$$\frac{\partial E}{\partial B} = \sum_{k=1}^{n} 2(Ax_k + B - y_k) = 0$$

si y sólo si

$$\left(\sum_{k=1}^{n} x_k\right) A + nB = \sum_{k=1}^{n} y_k$$

Ejemplo 68 Hacer el que esta antes.

7.3 Ajuste potencial $y = Ax^M$

En este caso, buscamos una curva de ajuste de la forma $y = Ax^M$, donde M es una constante conocida.

Como antes, sea

$$E(A) = \sum_{k=1}^{n} (Ax_k^M - y_k)^2\,.$$

Debemos resolver $\dfrac{\partial E}{\partial A} = 0$. Tenemos

$$\begin{aligned}\frac{\partial E}{\partial A} &= \sum_{k=1}^{n} 2(Ax_k^M - y_k)x_k^M \\ &= 2\sum_{k=1}^{n} Ax_k^{2M} - x_2^M y_k \,.\end{aligned}$$

Luego, $\dfrac{\partial E}{\partial A} = 0$ si y sólo si

$$A\sum_{k=1}^{n} x_k^{2M} = \sum_{k=1}^{n} x_k^M y_k \,,$$

es decir,

$$A = \sum_{k=1}^{n} x_k^M y_k \Big/ \sum_{k=1}^{n} x_k^{2M} \,,$$

Ejemplo 69 Consideremos la siguiente tabla de datos

x_k	y_k
-1	10
0	9
1	7
2	5
3	4
4	3
5	0
6	-1

y el exponente $M = 2$. Es dejado al lector realizar los cálculos numéricos.

7.4 Ajuste con curvas del tipo $y = Ce^{Ax}$, con $C > 0$

En este caso, tenemos que $y = Ce^{Ax}$. Tomando logarítmo natural nos queda $\ln(y) = Ax + \ln(C)$, lo cual puede ser escrito en la forma $Y = AX + B$, donde

$$\begin{cases} X &= x \\ B &= \ln(C) \\ Y &= \ln(y) \end{cases}$$

y el problema se transforma en

$$Y = Ax + B \,.$$

Este es llamado *métodos de linealización de datos.*

Dados (x_k, y_k), tenemos

$$(X_k, Y_k) = (x_k, \ln(y_k))$$

y las ecuaciones normales de Gauss son

$$\begin{aligned} \left(\sum_{k=1}^{n} X_k^2\right) A + \left(\sum_{k=1}^{n} X_k\right) B &= \sum_{k=1}^{n} X_k Y_k \\ \left(\sum_{k=1}^{n} X_k\right) A + nB &= \sum_{k=1}^{n} Y_k \,. \end{aligned}$$

Una vez calculados A y B, se obtiene $C = e^B$.

Ejemplo 70 Encontrar la curva de ajuste de la forma $y = Ce^{Ax}$, para los datos $(0, 1.5)$, $(1, 1.25)$, $(2, 3.5)$, (3.5), $(4, 7.5)$.

7.5 Método no lineal de los mínimos cuadrados para $y = Ce^{Ax}$

Este consiste en encontrar el mínimo de la función

$$E(A, C) = \sum_{k=1}^{n} (Ce^{Ax_k} - y_k)^2 \,.$$

Tenemos

$$\begin{aligned} \frac{\partial E}{\partial A} &= \sum_{k=1}^{n} 2(Ce^{Ax_k} - y_k) C x_k e^{Ax_k} = 0 \\ \frac{\partial E}{\partial C} &= \sum_{k=1}^{n} 2(Ce^{Ax_k} - y_k) e^{Ax_k} = 0 \end{aligned}$$

de donde,

$$\begin{cases} C \sum_{k=1}^{n} x_k e^{2Ax_k} - \sum_{k=1}^{n} x_k y_k e^{Ax_k} &= 0 \\ C \sum_{k=1}^{n} e^{2Ax_k} - \sum_{k=1}^{n} y_k e^{Ax_k} &= 0 \,. \end{cases}$$

Estas son ecuaciones no lineales en A y C, y se pueden resolver, por ejemplo, usando Newton en varias variables.

7.6 Combinaciones lineales en mínimos cuadrados

En este caso, nos son dados los datos $\{f_j(x)\}_{j=1}^m$.

Problema. Encontrar los coeficientes $\{c_j\}_{j=1}^m$ tal que la función $f(x)$ definida por

$$f(x) = \sum_{j=1}^{m} c_j f_j(x)$$

minimice la suma de los cuadrados de los errores.

Tenemos

$$E(c_1, c_2, \ldots, c_m) \quad = \quad \sum_{k=1}^{n} (f(x_k) - y_k)^2 = \sum_{k=1}^{n} \left(\left(\sum_{j=1}^{m} c_j f_j(x_k) \right) - y_k \right)^2 .$$

Para minimizar, debemos resolver las ecuaciones

$$\frac{\partial E}{\partial c_i} = 0 \,, \quad i = 1, \ldots, m \,.$$

Tenemos

$$\frac{\partial E}{\partial c_i} = \sum_{k=1}^{n} 2 \left(\left(\sum_{j=1}^{m} c_j f_j(x_k) \right) - y_k \right) f_i(x_k) = 0 \,, \qquad i = 1, \ldots, m$$

lo cual nos da el siguiente sistema de ecuaciones lineales

$$\sum_{j=1}^{m} \left(\sum_{k=1}^{n} f_i(x_k) f_j(x_k) \right) c_j = \sum_{k=1}^{n} f_i(x_k) y_k \,, \qquad i = 1, 2, \ldots, m \,.$$

Esto lo podemos escribir de otra forma, introduciendo notación matricial, para ello definimos

$$F = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_m(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_m(x_n) \end{bmatrix}_{n \times m}$$

$$F^T = \begin{bmatrix} f_1(x_1) & f_1(x_2) & \cdots & f_1(x_n) \\ f_2(x_1) & f_2(x_2) & \cdots & f_2(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ f_m(x_1) & f_m(x_2) & \cdots & f_m(x_n) \end{bmatrix}_{m \times n}$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} .$$

El elemento de la i–ésima fila del producto $F^T Y$ coincide con el i–ésimo elemento de la matriz columna que contiene los términos independientes en el sistema, esto es,

$$\sum_{k=1}^{n} f_i(x_k) y_k = \text{ fila}_i \left(F^T \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right) .$$

Ahora, el producto $F^T F$ es una mariz $m \times m$. El elemento (i, j) de $F^T F$ coincide con el coeficiente c_j en la i–ésima ecuación del sistema, esto es,

$$\sum_{k=1}^{n} f_i(x_k) f_j(x_k) = f_i(x_1) f_j(x_1) + f_i(x_2) f_j(x_2) + \cdots + f_i(x_n) f_j(x_n) .$$

Luego el problema se reduce a resolver el sistema de ecuaciones lineales

$$F^T F C = FTY$$

cuya incógnita es C.

7.7 Ajuste polonomial

Este es un caso particular del anterior, pues si

$$f(x) = c_1 + c_2 x + \cdots + c_{n+1} x^n$$

entonces f se obtiene como combinación lineal de las funciones linealmente independiente $\{f_j(x)\}_{j=1}^{n+1} = \{x^{j-1}\}_{j=1}^{n+1}$.

Ejemplo 71 Parábola óptima para mínimos cuadrados.

En este caso, nos son dados los datos $\{(x_k, y_k)\}_{k=1}^{n}$ y buscamos una curva de asjuste de la forma $y = f(x) = Ax^2 + B + C$.

Problema. Determinar los coeficientes A, B y C.

Sea

$$E(A, B, C) = \sum_{k=1}^{n} (Ax_k^2 + Bx_k + C - y_k)^2 .$$

Para minimizar, debemos resolver las ecuaciones

$$\left.\begin{array}{rcccl} 0 & = & \dfrac{\partial E}{\partial A} & = & 2\displaystyle\sum_{k=1}^{n}(Ax_k^2+Bx_k+C-y_k)x_k^2 \\ 0 & = & \dfrac{\partial E}{\partial B} & = & 2\displaystyle\sum_{k=1}^{n}(Ax_k^2+Bx_k+C-y_k)x_k \\ 0 & = & \dfrac{\partial E}{\partial C} & = & 2\displaystyle\sum_{k=1}^{n}(Ax_k^2+Bx_k+C-y_k) \end{array}\right\}$$

esto es,

$$\begin{aligned} \left(\sum_{k=1}^{n}x_k^4\right)A+\left(\sum_{k=1}^{n}x_k^3\right)B+\left(\sum_{k=1}^{n}x_k^2\right)C &= \sum_{k=1}^{n}y_kx_k^2 \\ \left(\sum_{k=1}^{n}x_k^3\right)A+\left(\sum_{k=1}^{n}x_k^2\right)B+\left(\sum_{k=1}^{n}x_k\right)C &= \sum_{k=1}^{n}x^ky_k \\ \left(\sum_{k=1}^{n}x_k^2\right)A+\left(\sum_{k=1}^{n}x_k\right)B+nC &= \sum_{k=1}^{n}y_k\,. \end{aligned}$$

Capítulo 8

Integración Numérica

En el Capítulo sobre Interpolación hemos deducidos las fórmulas de la regla de los Trapecios y la de Simpson.

Sean $x_0 < x_1 < \cdots < x_n$ puntos dados y sea $f : [x_0, x_n] \longrightarrow \mathbb{R}$ una función suficientemente diferenciable. Una fórmula del tipo

$$\int_{x_0}^{x_n} f(x) = Q[f] + E[f]\,, \tag{8.1}$$

donde

$$Q[f] = \sum_{k=0}^{n} \omega_k f(x_k) \tag{8.2}$$

es llamada una fórmula de *cuadratura* o de *integración numérica*, $Q[f]$ es llamada *cuadratura* y $E[f]$ es el *error de truncamiento* de la fórmula. Los puntos x_i $(i = 0, \ldots, n)$ son llamados los *nodo* de la cuadratura y los números ω_i son llamados los *pesos* de la cuadratura. Por ejemplo, las fórmulas de los trapecios, de Simpsons, de Simpson $\left(\frac{3}{8}\right)$, las de Newton–Cote, y muchas otras fórmulas de cuadratura pueden ser encontradas en textos de métodos numéricos.

El *grado de presición* de una fórmula de cuadratura como (8.1) es el menor número natural n, tal que $E(P_i) = 0$ para todo polinomio de grado menor o igual que n y existe un polinomio P_{n+1} de grado $n+1$ tal que $E[P_{n+1}] \neq 0$.

Ejemplos.

1. Fórmula de los trapecios simple. Esta es dada por

   $$\int_{x_0}^{x_1} f(x)dx = \frac{h}{2}[f(x_0) + f(x_1)] - \frac{h^3}{12} f''(c)\,,$$

   donde $h = x_1 - x_0$ y $c \in]x_0, x_1[$. En este caso, tenemos los pesos $\omega_0 = \omega_1 = \frac{h}{2}$, los nodos son x_0 y x_1, luego $Q[f] = \frac{h}{2}f(x_0) + \frac{h}{2}f(x_1)$ y $E[f] = -\frac{h^3}{12} f''(c)$.

   Esta fórmula de cuadratura tiene grado de precisión igual a 1 si f'' es continua.

2. Regla de Simpson simple. Esta fórmula de cuadratura es dada por

   $$\int_{x_0}^{x_2} f(x) = \frac{h}{3}\,[f(x_0) + 4f(x_1) + f(x_2)] - \frac{h^5}{90}\, f^{(4)}(c)$$

donde $c \in]x_0, x_2[$ y $h = x_2 - x_1 = x_1 - x_0$ (paso constante). En este caso, los nodos son $x_0 < x_1 < x_2$, y los pesos son $\omega_0 = \omega_2 = \frac{h}{3}$ y $\omega_1 = \frac{4h}{3}$, luego $Q[f] = \frac{h}{3}f(x_0) + \frac{4h}{3}f(x_1) + \frac{h}{3}f(x_2)$ y $E[f] = -\frac{h^5}{90}f^{(4)}(c)$. Esta fórmula de cuadratura tiene grado de presición 3 si $f^{(4)}$ es continua.

3. Regla de Simpson $\left(\frac{3}{8}\right)$ simple. En esta fórmula de cuadratura tomamos los nodos $x_0 < x_1 < x_2 < x_3$, considerando paso constante $h = x_{i+1} - x_i$ y los pesos $\omega_0 = \omega_3 = \frac{3h}{8}$, $\omega_1 = \omega_2 = \frac{9h}{8}$, se tiene la fórmula de cuadratura

$$\int_{x_0}^{x_3} f(x)dx = \underbrace{\frac{3}{8}\,h\,[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)]}_{Q[f]} - \underbrace{\frac{3}{80}\,h^5\,f^{(4)}(c)}_{E[f]}$$

donde $c \in]x_0, x_3[$, la cual tiene grado de presición 3 si $f^{(4)}$ es continua.

4. Regla de Boole simple. Esta fórmula de cuadratura es dada como sigue. Tomamos los nodos $x_0 < x_1 < x_2 < x_3 < x_4$, con paso constante $h = x_{i+1} - x_i$, y los pesos $\omega_0 = \omega_4 = \frac{14}{45}\,h$, $\omega_1 = \omega_3 = \frac{64}{45}\,h$ y $\omega_2 = \frac{24}{45}\,h$. La regla de Boole es dada entonces por

$$\int_{x_0}^{x_4} f(x)dx = \underbrace{\frac{2}{45}\,h\,[7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)]}_{Q[f]} - \underbrace{\frac{8}{945}\,h^7\,f^{(6)}(c)}_{E[f]},$$

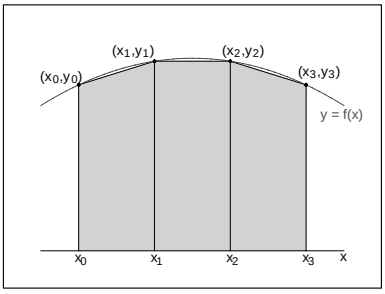
con $c \in]x_0, x_4[$. Esta fórmula de cuadratura tiene grado de presición5 si $f^{(6)}$ es continua.

8.1 Regla de los trapecios

Ya vimos que la regla de los Trapecios simple viene dada por

$$\int_{x_j}^{x_{j+1}} f(x) = \frac{h_j}{2}(f(x_j) + f(x_{j+1})) - \frac{h_j^3}{12}f''(\xi_j)\,, \tag{8.3}$$

donde $h_j = x_{j+1} - x_j$ y $\xi_j \in]x_j, x_{j+1}[$.



Ahora, si tenemos los nodos $x_0 < x_1 < \cdots < x_n$, con paso $h_j = x_{j+1} - x_j$ y queremos aproximar el valor de

$$\int_{x_0}^{x_n} f(x)dx$$

usando el hecho que $\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$, integramos sobre los subintervalos $[x_j, x_{j+1}]$, para $j = 0, 1, \ldots, n-1$ y aplicando la fórmula de los trapecios simples, obtenemos

$$\int_{x_0}^{x_n} f(x) = \sum_{j=0}^{n-1} \int_{x_j}^{x_{j+1}} f(x) = \underbrace{\sum_{j=0}^{n-1} \frac{h_j}{2} \left(f(x_j) + f(x_{j+1})\right)}_{Q[f]} - \underbrace{\sum_{j=0}^{n-1} \frac{h_j^3}{12} f''(\xi_j)}_{E_{total} \text{ error total}}, \tag{8.4}$$

donde $\xi_j \in]x_j, x_{j+1}[$.

Analicemos el error total $E_{\text{total}} = -\sum_{j=0}^{n-1} \frac{h_j^3}{12} f''(\xi_j)$. Si consideramos paso constante $h = h_j$ para $j = 0, 1, \ldots, n-1$, y $f''(x)$ continua, entonces obtenemos

$$\begin{aligned} E_{\text{total}} &= -\frac{h^3}{12} \sum_{j=0}^{n-1} f''(\xi_j), \qquad \xi_j \in]x_j, x_{j+1}[\\ &= -\frac{h^3}{12} n f''(\xi), \qquad \xi \in]x_0, x_n[\\ &= -\frac{h^2}{12} nh f''(\xi), \quad \text{como} \quad h = \frac{x_n - x_0}{n} \\ &= -\frac{h^2}{12}(x_n - x_0) f''(\xi). \end{aligned}$$

Luego, en el caso de paso constante h, tenemos la fórmula

$$\int_{x_0}^{x_n} f(x) = \underbrace{\frac{h}{2} \left[f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)\right]}_{Q[f]} - \underbrace{\frac{h^2}{12}(x_n - x_0) f''(\xi)}_{E[f]}, \tag{8.5}$$

la cual es llamada *Regla de los trapecios compuesta*.

8.2 Regla de Simpson

Falta dibujos

Para obtener la regla de Simpson simple, suponemos por simplicidad, que tenemos los nodos $x_0 < x_1 < x_2$ y el paso $h = x_2 - x_1 = x_1 - x_0$ es constante. Integrando el polinomio de Lagrange de grado 2, que interpola a f en esos tres nodos, obtenemos

$$\int_{x_0}^{x_2} f(x) = \frac{h}{3} \left[f(x_0) + 4f(x_1) + f(x_2)\right] - \frac{h^5}{90} f^{(4)}(c) \tag{8.6}$$

donde $c \in]x_0, x_4[$.

Ahora si tenemos los nodos $x_0 < x_1 < \cdots < x_n$, como antes suponemos que tenemos paso constante $h = x_{k+1} - x_k$, integrando en cada subintervalo de la forma $[x_j, x_{j+2}]$, tenemos

$$\int_{x_j}^{x_{j+2}} f(x) = \frac{h}{3}(f(x_j) + 4f(x_{j+1}) + f(x_{j+2})) - \frac{h^5}{90} f^{(4)}(\theta_j), \tag{8.7}$$

donde $\theta_j \in]x_j, x_{j+2}[$. Ahora, como

$$\int_{x_0}^{x_n} f(x)dx = \int_{x_0}^{x_2} f(x)dx + \int_{x_2}^{x_4} f(x)dx + \cdots + \int_{x_{n-2}}^{x_n} f(x)dx\,,$$

note que n debe ser par y $n \geqslant 2$. De esto, tenemos

$$\begin{aligned}\int_{x_0}^{x_n} f(x)dx &= \frac{h}{3}[(f(x_0)+4f(x_1)+f(x_2))+(f(x_2)+4f(x_3)+f(x_4))+\cdots \\ &\quad +(f(x_{n-2})+4f(x_{n-1})+f(x_n))] - \frac{h^5}{90}\,\frac{n}{2}\,f^{(4)}(\theta)\,,\end{aligned}$$

con $\theta \in]x_0, x_n[$. Reordenando, nos queda

$$\int_{x_0}^{x_n} f(x)dx = \underbrace{\frac{h}{3}\left(f(x_0)+f(x_n)+4\sum_{j=1}^{\frac{n-2}{2}} f(x_{2j-1})+2\sum_{j=1}^{\frac{n-2}{2}} f(x_{2j})\right)}_{E[f]} \underbrace{- h^4\,\frac{(x_n-x_0)}{180}\,f^{(4)}(\theta)}_{E[f]}\,. \tag{8.8}$$

Esta es la *regla de Simpson compuesta*.

8.3 Regla de Simpson $(\frac{3}{8})$

Consideramos paso constante $h = x_{j+1} - x_j$. Sea $P_j(x)$ el polinomio de interpolación de Lagrange con nodos x_j, x_{j+1}, x_{j+2} y x_{j+3}, con paso h constante. Tenemos

$$\int_{x_j}^{x_{j+3}} f(x)dx \approx \int_{x_j}^{x_{j+3}} P_j(x)dx \;\approx\; \frac{3}{8}h\,[f(x_j)+3f(x_{j+1})+3f(x_{j+2})+f(x_{j+3})]\,, \tag{8.9}$$

esta es la *regla de Simpson* $\left(\frac{3}{8}\right)$ *simple*.

Ahora, como

$$\int_{x_0}^{x_n} f(x)dx = \int_{x_0}^{x_3} f(x)dx + \int_{x_3}^{x_6} f(x)dx + \cdots + \int_{x_{n-3}}^{x_n} f(x)dx\,,$$

notemos que $n = 3m$ y $m \geqslant 1$. Reordenando, obtenemos

$$\int_{x_0}^{x_n} f(x)dx \approx \frac{3}{8}h\left(f(x_0)+f(x_n)+3\sum_{j=0}^{\frac{n-3}{3}} f(x_{3j+1})+3\sum_{j=0}^{\frac{n-3}{3}} f(x_{3j+2})+2\sum_{j=0}^{\frac{n-3}{3}} f(x_{3j+3})\right)\,, \tag{8.10}$$

esta es la regla de Simpson $(\frac{3}{8})$, cuyo error es dado por

$$E[f] = -\frac{(x_n-x_0)^5}{6480}\,f^{(4)}(\theta)\,, \tag{8.11}$$

donde $\theta \in]x_0, x_n[$. Las reglas de los trapecios y la de Simpson pertenecen a una clase de métodos para aproximar integrales, llamadas *fórmulas de Newton–Cotes*.

8.4 Fórmulas de Newton–Cotes cerradas de $(n+1)$ puntos

Notemos que para obtener la fórmula de la regla de los trapecios integramos el polinomio de interpolación de Lagrange de grado 1 entre los puntos x_j y x_{j+1}, para obtener la fórmula de la regla ed Simpson integramos el polinomio de interpolación de Lagrange de grado 2 entre los puntos x_j, x_{j+1} y x_{j+2}, para la regla de Simpson $\frac{3}{8}$ integramos el polinomio de interpolación de Lagrange de grado 3 entre los puntos x_j, x_{j+1}, x_{j+2} y x_{j+3}. Para otras fórmulas podemos continuar integrando el polinomio de interpolación de Lagrange, digamos de grado k, entre los puntos $x_j, \ldots, x_{j+k}$.

Las fórmulas de Newton–Cotes cerradas de $(n+1)$ puntos utilizan nodos $x_j = x_0 + jh$, $j = 0, 1, \ldots, n$, $h = \frac{b-a}{n}$, $x_0 = a$ y $x_n = b$. Son llamadas *cerradas* pues incluyen los extremos x_0 y x_n. Ellas tienen la forma

$$\int_{x_0}^{x_n} f(x)dx \approx \sum_{j=0}^{n} A_{n,j} f(x_j)\,,$$

donde

$$A_{n,j} = \int_{x_0}^{x_n} L_{n,j}(x)dx = \int_{x_0}^{x_n} \prod_{\substack{k\,=\,0\\ k\,\neq\,j}}^{n} \frac{(x-x_k)}{(x_j-x_k)}\,,$$

son obtenidas integrando el polinomio de Lagrange de grado n, que recordemos es dado por

$$L_n(x) = \sum_{j=0}^{n} f(x_j) L_{n,j}(x)\,,$$

donde $L_{n,j} = \prod_{\substack{i\,=\,0\\ i\,\neq\,j}}^{n} \frac{x-x_i}{x_j-x_i}$, que interpola a f en los nodos $x_0 < x_1 < \cdots < x_n$

Teorema 8.1 *Denotemos por $\sum_{j=0}^{n} A_j f(x_j)$ la fórmula de Newton–Cotes cerrada de $(n+1)$ puntos, con $x_0 = a$, $x_n = b$ y $h = \frac{b-a}{n}$. Entonces existe $\theta \in]a,b[$, tal que*

(α) $\displaystyle\int_a^b f(x)dx = \sum_{j=0}^{n} A_j f(x_j) + \frac{h^{n+3}}{(n+2)!} f^{(n+2)}(\theta) \int_0^n t^2(t-1)\cdots(t-n)dt\,,$

si n es par y f es $n+2$ veces derivable en $[a,b]$ y $f^{(n+2)}(x)$ es continua.

(β) $\displaystyle\int_a^b f(x)dx = \sum_{j=0}^{n} A_j f(x_j) + \frac{h^{n+2}}{(n+1)!} f^{(n+1)}(\theta) \int_0^n t(t-1)\cdots(t-n)dt\,,$

si n es impar y f es $n+1$ veces derivable en $[a,b]$ y $f^{(n+1)}(x)$ es continua.

8.5 Fórmulas abiertas de Newton–Cotes

Estas usan nodos $x_j = x_0 + jh$, $j = 0, 1, \ldots, n$, donde $h = \frac{b-a}{n+2}$ y $x_0 = a + h$. Luego $x_n = b - h$. Marcamos los extremos $x_{-1} = a$ y $x_{n+1} = b$ y tenemos

$$\int_a^b f(x)dx = \int_{x_{-1}}^{x_{n+1}} f(x)dx \approx \sum_{j=0}^{n} A_j f(x_j)\,,$$

donde $A_j = \int_a^b L_{n,j}(x)dx$.

Con las notaciones anteriores, tenemos que existe $\theta \in]a,b[$ tal que

(α') $\int_a^b f(x)dx = \sum_{j=0}^{n} A_j f(x_j) + \frac{h^{n+3}}{(n+2)!} f^{(n+2)}(\theta) \int_{-1}^{n} t^2(t-1)\cdots(t-n)dt$,

si n es par y f es $n+2$ veces derivable en $[a,b]$ y $f^{(n+2)}(x)$ es continua.

(β') $\int_a^b f(x)dx = \sum_{j=0}^{n} A_j f(x_j) + \frac{h^{n+2}}{(n+1)!} f^{(n+1)}(\theta) \int_{-1}^{n} t(t-1)\cdots(t-n)dt$,

si n es impar y f es $n+1$ veces derivable en $[a,b]$ y $f^{(n+1)}(x)$ es continua.

Por ejemplo,

$n = 0$, $\int_{x_{-1}}^{x_1} f(x) = 2hf(x_0) + \frac{h^3}{3} f''(\theta)$, $x_{-1} < \theta < x_1$,

$n = 1$, $\int_{x_{-1}}^{x_2} f(x) = \frac{3h}{2}(f(x_0) + f(x_1)) + \frac{3}{4}h^3 f''(\theta)$, $x_{-1} < \theta < x_2$,

8.6 Integración de Romberg

Comenzamos por obtener aproximaciones para $\int_a^b f(x)$ usando la regla de los trapecios con m_j nodos, $m_1 = 1$, $m_2 = 2, \ldots, m_j = 2^{j-1}$. Los valores del paso h_j correspondiente a m_j son dados por $h_j = \frac{b-a}{m_j} = \frac{b-a}{2^{j-1}}$. Luego, tenemos

$$\int_a^b f(x) = \frac{h_j}{2}\left(f(a) + f(b) + 2\sum_{i=1}^{2^{j-1}-1} f(a+ih_j)\right) - \frac{(b-a)}{12}h_j^2 f''(\theta_j)\,, \tag{8.12}$$

donde $\theta_j \in]a,b[$.

Usando la notación

$$\begin{aligned} R_{1,1} &= \frac{h_1}{2}(f(a)+f(b)) = \frac{b-a}{2}(f(a)+f(b)) \\ R_{2,1} &= \frac{h_2}{2}(f(a)+f(b)+2f(a+h_2)) \\ &= \frac{b-a}{4}\left(f(a)+f(b)+2f\left(a+\frac{b-a}{2}\right)\right) \\ &= \frac{1}{2}(R_{1,1}+h_1 f(a+h_2)) \end{aligned}$$

En general,

$$R_{j,1} = \frac{1}{2}\left(R_{j-1,1} + h_{j-1}\sum_{i=0}^{2^{j-2}} f(a+(2i-1)h_j)\right)$$

$j = 2, 3, \ldots, n$. Tenemos

$$\int_a^b f(x) - R_{j,1} = K_1 h_j^2 + \sum_{i=2}^{\infty} K_i h_j^{2i} \,,$$

donde K_i para cada i es independiente de h_j .

También tenemos las relaciones

$$R_{k,j} = R_{k,j-1} + \frac{R_{k,j-1} - R_{k-1,j-1}}{4^{j-1} - 1}$$

$k = 2, \ldots, n$ y $j = 2, \ldots, k$.

Estos resultados pueden tabularse como

$$\begin{array}{lllll} R_{1,1} & & & & \\ R_{2,1} & R_{2,2} & & & \\ R_{3,1} & R_{3,2} & R_{3,3} & & \\ \vdots & \vdots & \vdots & \ddots & \\ R_{n,1} & R_{n,2} & R_{n,3} & \cdots & R_{n,n} \\ \downarrow & & & & \\ \int_a^b f(x) & & & & \end{array}$$

8.7 Cuadratura gaussiana

Queremos aproximar el valor de

$$\int_a^b f(x) \,.$$

Haciendo el cambio de variable $x = \frac{1}{2}((b-a)t + (a+b))$ tenemos que

$$\int_a^b f(x)dx = \int_{-1}^1 f\left(\frac{(b-a)t + (b+a)}{2}\right) \frac{b-a}{2}\, dt$$

por lo tanto buscamos un método para obtener aproximaciones a una integral del tipo

$$\int_{-1}^1 f(x)dx \,.$$

Aquí se trata de determinar constantes c_1 y c_2 y valores x_1 y x_2 de modo que

$$\int_{-1}^1 f(x) \approx c_1 f(x_1) + c_2 f(x_2) \,, \tag{8.13}$$

y el resultado sea exacto cuando $f(x)$ es un polinomio de grado menor o igual que 3.

Escribamos $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$. Tenemos entonces

$$\int f(x) = a_0 \int 1 + a_1 \int x + a_2 \int x^2 + a_3 \int x^3 \,,$$

luego necesitamos $c_1, c_2\,, x_1$ y x_2 tales que

$$\begin{aligned} c_1 \cdot 1 + c_2 \cdot 1 &= \int_{-1}^{1} 1 = 2 \\ c_1 x_1 + c_2 x_2 &= \int_{-1}^{1} x = 0 \\ c_1 x_1^2 + c_2 x_2^2 &= \int_{-1}^{1} x^2 = \frac{2}{3} \\ c_1 x_1^3 + c_2 x_2^3 &= \int_{-1}^{1} x^3 = 0 \end{aligned}$$

resolviendo esas ecuaciones obtenemos $c_1 = 1\,,\ c_2 = 1\,,\ x_1 = -\frac{\sqrt{3}}{3}$ y $x_2 = \frac{\sqrt{3}}{3}$. Luego

$$\int_{-1}^{1} f(x) \approx f\left(-\frac{\sqrt{3}}{3}\right) + f\left(\frac{\sqrt{3}}{3}\right) .$$

En general, el problema es resuelto considerando los polinomios de Legendre $\{P_0(x), P_1(x), \ldots\}$. Estos son polinomios que satisfacen

1. $P_n(x)$ es un polinomio de grado $n\,,\ n = 0, 1, \ldots$.
2. $\int_{-1}^{1} P(x) P_n(x) dx = 0$ cuando $P(x)$ es un polinomio de grado menor que n.

Los primeros polinomios de Legendre son

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= x^2 - \frac{1}{3} \\ P_3(x) &= x^3 - \frac{3}{5}x \\ &\vdots \end{aligned}$$

Tenemos el siguiente teorema

Teorema 8.2 *Supongamos que $x_1, x_2, \ldots, x_n$ son las raíces del polinomio de Legendre $P_n(x)$ y que los números c_i son dados por*

$$c_i = \int_{-1}^{1} \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j}\, dx\,, \qquad i = 1, 2, \ldots, n$$

Si $P(x)$ es un polinomio de grado menor que $2n$. Entonces

$$\int_{-1}^{1} P(x) = \sum_{i=1}^{n} c_i P(x_i)\,.$$

8.8 Ejemplos resueltos

Ejemplo 72 Considere la siguiente tabla de valores

x	0	1	2	3	4	5	6
$f(x)$	5	4	1	-1	-1	0	3

a) Encuentre la interpolación de Lagrange $L_6(x)$ de esos datos. Usando el método de Newton, encuentre una aproximación para la raíz de $L_6(x)$ en el intervalo $[2,3]$.

b) Usando la integración numérica de Simpson, y también integrando $L_6(x)$ directamente, obtenga aproximaciones para $\int_0^6 f(x)dx$.

Solución.

a) Tenemos que $L_6(x) = L_{6,0}(x)y_0+L_{6,1}(x)y_1+L_{6,2}(x)y_2+L_{6,3}(x)y_3+L_{6,4}(x)y_4+L_{6,5}(x)y_5+L_{6,6}(x)y_6$, donde

$$\begin{aligned}
L_{6,0}(x) &= \frac{(x-1)(x-2)(x-3)(x-4)(x-5)(x-6)}{720} \\
&= \frac{1}{720}\,x^6 - \frac{7}{240}\,x^5 + \frac{35}{144}\,x^4 - \frac{49}{48}\,x^3 + \frac{203}{90}\,x^2 - \frac{49}{20}\,x + 1 \\
L_{6,1}(x) &= -\frac{x(x-2)(x-3)(x-4)(x-5)(x-6)}{120} \\
&= -\frac{1}{120}\,x^6 + \frac{1}{6}\,x^5 - \frac{31}{24}\,x^4 + \frac{29}{6}\,x^3 - \frac{87}{10}\,x^2 + 6\,x \\
L_{6,2}(x) &= \frac{x(x-1)(x-3)(x-4)(x-5)(x-6)}{48} \\
&= \frac{1}{48}\,x^6 - \frac{19}{48}\,x^5 + \frac{137}{48}\,x^4 - \frac{461}{48}\,x^3 + \frac{117}{8}\,x^2 - \frac{15}{2}\,x
\end{aligned}$$

$$\begin{aligned}
L_{6,3}(x) &= -\frac{x(x-1)(x-2)(x-4)(x-5)(x-6)}{36} \\
&= -\frac{1}{36}x^6+\frac{1}{2}x^5-\frac{121}{36}x^4+\frac{31}{3}x^3-\frac{127}{9}x^2+\frac{20}{3}x \\
L_{6,4}(x) &= \frac{x(x-1)(x-2)(x-3)(x-5)(x-6)}{48} \\
&= \frac{1}{48}x^6-\frac{17}{48}x^5+\frac{107}{48}x^4-\frac{307}{48}x^3+\frac{33}{4}x^2-\frac{15}{4}x \\
L_{6,5}(x) &= -\frac{x(x-1)(x-2)(x-3)(x-4)(x-6)}{120} \\
&= -\frac{1}{120}x^6+\frac{2}{15}x^5-\frac{19}{24}x^4+\frac{13}{6}x^3-\frac{27}{10}x^2+\frac{6}{5}x \\
L_{6,6}(x) &= \frac{x(x-1)(x-2)(x-3)(x-4)(x-5)}{720} \\
&= \frac{1}{720}x^6-\frac{1}{48}x^5+\frac{17}{144}x^4-\frac{5}{16}x^3+\frac{137}{360}x^2-\frac{1}{6}x
\end{aligned}$$

Ahora, es fácil que ver el polinomio de Lagrange es dado por

$$L_6(x) = 5+\frac{5}{6}x-\frac{1}{4}x^3+\frac{7}{18}x^4-\frac{1}{12}x^5+\frac{1}{180}x^6-\frac{341}{180}x^2$$

y de aquí tenemos que la transformación de Newton de $L_6(x)$ es dada por

$$\mathrm{N}(x) = x-\frac{5+\frac{5}{6}x-\frac{1}{4}x^3+\frac{7}{18}x^4-\frac{1}{12}x^5+\frac{1}{180}x^6-\frac{341}{180}x^2}{\frac{5}{6}-\frac{3}{4}x^2+\frac{14}{9}x^3-\frac{5}{12}x^4+\frac{1}{30}x^5-\frac{341}{90}x}$$

Comenzando la busqueda de la raíz en $x_0 = 2$ la convergencia es rápida y la raíz buscada es aproximadamente igual a $\bar{x} = 2,37$, y $L_6(\bar{x}) = 0,0096448$. (Obviamente podemos obtener una mejor aproximación, pero eso sólo implica más iteraciones).

b) La fórmula de integración de Simpson es $\int_a^b f(x)dx \approx \frac{h}{3}(f_0+4f_1+2f_2+4f_3+2f_4+\cdots+2f_{n-2}+4f_{n-1}+f_n)$, donde $f_i = f(x_i) = y_i$, y $h = \frac{b-a}{n}$. En nuestro caso, $h = 1$ y tenemos $\int_0^6 f(x)dx \approx (5+4\cdot 4+2\cdot 1+4\cdot -1+2\cdot -1+4\cdot 0+3)/3 = 20/3 = 6,66666....$

Por otra parte, tenemos que

$$L_6(x) = 5+\frac{5}{6}x-\frac{1}{4}x^3+\frac{7}{18}x^4-\frac{1}{12}x^5+\frac{1}{180}x^6-\frac{341}{180}x^2$$

integrando, obtenemos

$$\int_0^6 f(x)dx \approx \int_0^6 L_6(x)dx = 6,571428571$$

Ejemplo 73 Considere la siguiente tabla de valores

k	x_k	$f(x_k)$	$f'(x_k)$
0	1	5	-2
1	2	-1	-1
2	3	3	4

a) Encuentre la aproximación de Hermite $H_5(x)$ de esos datos. Usando el método de Newton, encuentre una raíz de $H_5(x)$.

b) Usando la regla de los trapecios, y también integrando directamente $\int_1^3 H_5(x)dx$, calcular aproximaciones de $\int_1^3 f(x)dx$.

Solución

a) Tenemos que

$$\begin{aligned}
L_{2,0}(x) &= \frac{(x-2)(x-3)}{(1-2)(1-3)} = \frac{x^2}{2} - \frac{5x}{2} + 3 \\
L'_{2,0}(x) &= x - \frac{5}{2} \\
L_{2,1}(x) &= \frac{(x-1)(x-3)}{(2-1)(2-3)} = -x^2 + 4x - 3 \\
L'_{2,1}(x) &= -2x + 4 \\
L_{2,2}(x) &= \frac{(x-1)(x-2)}{(3-1)(3-2)} = \frac{x^2}{2} - \frac{3x}{2} + 1 \\
L'_{2,2}(x) &= x - \frac{3}{2}
\end{aligned}$$

Ahora, como $H_{n,j}(x) = (1 - 2(x - x_j)L'_{n,j}(x_j))(L_{n,j}(x))^2$ tenemos

$$\begin{aligned}
H_{2,0}(\,x\,) &= (\,3\,x - 2\,)\left(\frac{1}{2}\,x^2 - \frac{5}{2}\,x + 3\right)^2 \\
&= \frac{3}{4}\,x^5 - 8\,x^4 + \frac{131}{4}\,x^3 - \frac{127}{2}\,x^2 + 57\,x - 18 \\
\hat{H}_{2,0}(x) &= (\,x - 1\,)\left(\frac{1}{2}\,x^2 - \frac{5}{2}\,x + 3\right)^2 \\
&= \frac{1}{4}\,x^5 - \frac{11}{4}\,x^4 + \frac{47}{4}\,x^3 - \frac{97}{4}\,x^2 + 24\,x - 9
\end{aligned}$$

$$\begin{aligned}
H_{2,1}(x) &= (-x^2+4\,x-3\,)^2 \\
&= x^4-8x^3+22x^2-24x+9 \\
\hat{H}_{2,1}(\,x\,) &= (\,x-2\,)\,(-x^2+4\,x-3\,)^2 \\
&= x^5-10\,x^4+38\,x^3-68\,x^2+57\,x-18 \\
H_{2,2}(\,x\,) &= (\,x-8\,)\left(\frac{1}{2}\,x^2-\frac{3}{2}\,x+1\right)^2 \\
&= \frac{1}{4}\,x^5-\frac{7}{2}\,x^4+\frac{61}{4}\,x^3-29\,x^2+25\,x-8 \\
\hat{H}_{2,2}(\,x\,) &= (\,10-3x\,)\left(\frac{1}{2}\,x^2-\frac{3}{2}\,x+1\right)^2 \\
&= \frac{1}{4}\,x^5-\frac{9}{4}\,x^4+\frac{31}{4}\,x^3-\frac{51}{4}\,x^2+10\,x-3
\end{aligned}$$

Luego el polinomio de Hermite aproximando los datos es

$$H_5(x)=x^5-\frac{27x^4}{2}+67x^3-\frac{299x}{2}+145x-45$$

La transformada de Newton de $H_5(x)$ es

$$\mathrm{N}(\,x\,)=x-\frac{-45+145\,x+x^5-\dfrac{299}{2}\,x^2+67\,x^3-\dfrac{27}{2}\,x^4}{145+5\,x^4-299\,x+201\,x^2-54\,x^3}$$

Comenzando la busqueda de la raíz con $x_0=1.6$ la convergencia es rápida y el valor aproximado de la raíz es $\bar{x}=1.71$, el correspondiente valor $H_5(1.71)=-0.0026091$. Otra alternativa es comenzar con $x_0=2.4$ obtenemos la raíz $\tilde{x}=2.44$, el correspondiente valor $H_5(2.44)=-0.002691$.

b) Calculando

$$\int_1^3 f(x)dx\approx\int_1^3 H_2(x)dx=-45\,x+\frac{145}{2}\,x^2+\frac{1}{6}\,x^6-\frac{299}{6}\,x^3+\frac{67}{4}\,x^4-\frac{27}{10}\,x^5|_1^3=2,266666667$$

Por la regla de los trapecios nos queda

$$\int_1^3 f(x)dx\approx\frac{5-1}{2}+\frac{-1+3}{2}=3$$

8.9 Ejercicios

Problema 8.1 Sea $f(x) = 1+e^{-x}\text{sen}(4x)$. Aproxime el valor de la integral $\int_0^1 f(x)dx$, usando

1. la regla de los trapecios simple,
2. Regla de Simpson simple,
3. Regla de Simpson $(\frac{3}{8})$ simple
4. Regla de Boole simple.

Compare los resultados obtenidos en cada aproximación con el resultado exacto de la integral

$$\int_0^1 f(x)dx = \frac{21e - 4\cos(4) - \text{sen}(4)}{17e} \approx 1.3082506046426\ldots$$

Problema 8.2 Deduzca la fórmula del punto medio simple para aproximar integrales, la cual es dada por

$$\int_a^b f(x)dx \approx (b-a)\, f\left(\frac{a+b}{2}\right) \tag{8.14}$$

Usando esta fórmula, deduzca la fórmula del punto medio compuesta para aproximar el valor de la integral $\int_{x_0}^{x_n} f(x)dx$, usando los nodos $x_0 < x_1 < \cdots < x_n$, con paso constante $h = x_{j+1} - x_j$, la cual es dada por

$$\int_{x_0}^{x_n} f(x)dx = h\sum_{k=1}^{n} f(\bar{x}_k) - \frac{x_n - x_0}{24}\, h^2\, f''(\alpha) \tag{8.15}$$

donde $\bar{x}_k = \frac{x_{k-1}+x_k}{2}$ y $\alpha \in]x_0, x_n[$.

Problema 8.3 Se desea calcular una aproximación al valor de $\pi = 3.14159265358979323846264338 32\ldots$. Sabemos que

$$\pi = 4\int_0^1 \frac{1}{1+x^2}\, dx\,.$$

Use la regla de Simpson con $n = 10$ para obtener una aproximación al valor de π dado arriba.

Problema 8.4 Para calcular las integrales

$$E_n = \int_0^1 x^n e^{x-1}\, dx\,, \quad n = 1, 2, \ldots,$$

podemos usar integración por partes

$$\int_0^1 x^n e^{x-1} dx = x^n e^{x-1}\big|_0^1 - \int_0^1 n x^{n-1} e^{x-1}\, dx$$

con lo que obtenemos el siguiente algoritmo numérico

$$E_n = 1 - nE_{n-1}\,, \quad n = 2, 3, \ldots,$$

donde $E_1 = 1/e$.

1. Es este algoritmo convergente?

2. Es numéricamente estable?

3. Usando cuadratura gaussiana de la forma

$$\int_a^b f(x)dx = Af(0) + Bf(1/2) + Cf(1)\,.$$

Calcule las constantes A, B y C y aplique la fórmula para estimar el valor de la integral E_n .

4. Cuál valor de la integral es más exacto (cuadratura o iterativamente)?

Problema 8.5 Si $ab > 0$, podemos hacer el cambio de variable y transformar la integral

$$\int_a^b f(x)dx = \int_{1/b}^{1/a} \frac{1}{t^2}\, f(1/t)dt\,.$$

Considere la evaluación de la integral impropia

$$N(x) = \int_{-\infty}^{x} \frac{1}{2\pi}\, e^{-x^2/2}\, dx$$

1. Transforme $N(x)$ usando la expresión anterior para que pueda utilizar la fórmula de Simpson.

2. Aplique la fórmula de integración de Simpson compuesta con M puntos a cada una de las integrales obtenidas en la parte (a).

Problema 8.6 Se sabe que $\displaystyle\int_0^3 e^x dx = 19.08553692....$ Estime el valor de la integral usando trapezoide, Simpson y Simpson $\left(\frac{3}{8}\right)$.

Problema 8.7 Sea f una función continua. Use cuadratura de Gauss sobre $[-1, 1]$ con 3 nodos para obtener la fórmula

$$\int_{-1}^{1} f(x) \approx \frac{5}{9} f\left(\frac{-\sqrt{15}}{5}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\frac{\sqrt{15}}{5}\right)\,.$$

Problema 8.8 Deduzca la fórmula de cuadratura gaussiana con 3 modos reescalada al intervalo $[0, 1]$, la cual es dada por

$$\int_0^1 f(x)dx \approx \frac{5}{18} f\left(\frac{5-\sqrt{15}}{10}\right) + \frac{4}{9} f\left(\frac{1}{2}\right) + \frac{5}{18} f\left(\frac{5+\sqrt{15}}{10}\right)$$

y use esta para calcular un valor aproximado para $\displaystyle\int_0^1 f(x)x^3 dx$

Problema 8.9 Determine a, b y c tal que la fórmula de cuadratura

$$Q(f) = af(0) + bf\left(\frac{1}{2}\right) + cf(1)$$

es exacta para la integral

$$\int_0^1 f(x)x^3 dx$$

si $f(x)$ es un polinomio de grado menor o igual que 2. Encuentre el error en $Q(x^4)$, como una aproximación de $\displaystyle\int_0^1 x^7 dx$.

Problema 8.10 Determine c_1, c_2, c_3 y c_4 tal que la fórmula de cuadratura

$$Q(f) = c_1 f(0) + c_2 f(1) + c_3 f(3) + c_4 f(4)$$

es exacta para la integral

$$\int_0^\infty f(x)e^{-x} dx$$

si f es un polinomio de grado menor o igual que 3.

Problema 8.11 Obtenga los valores de las constantes c_1, c_2, x_1 y x_2 tales que la siguiente fórmula de integración numérica

$$\int_{-1}^{1} F(x)dx = c_1 F(x_1) + c_2 F(x_2)$$

sea exacta para todos los polinomios de grado menor o igual a 3. Use dicha fórmula para aproximar la integral de la función $f(x) = (x-3)^5$ en el intervalo $[2, 4]$.

Problema 8.12 Queremos aproximar el valor de la integral $I(f) = \int_0^c f(x)dx$, donde $f(x) = \cos^2(x)e^{-x}$ y c es una constante positiva. Esta integral puedes ser aproximada con error absoluto menor que $\varepsilon > 0$ dado, para ellos usamos una fórmula de integración numérica. Fijando $\varepsilon = 10^{-3}$, estime la cantidad de intervalos que son necesarios, en terminos de c, si usamos la regla de los trapecios compuesta para aproximar $I(f)$.

Problema 8.13 Sea $f : [x_0, x_n] \longrightarrow \mathbb{R}$ una función a la cual deseamos aproximar el valor de $\int_{x_0}^{x_n} f(x)dx$. Denotemos por I_1 el valor obtenido usando al regla de los trapecios con un paso constante h_1 e I_2 el valor obtenido usando la regla de los trapecios con paso constante h_2, con la condición $h_1 \neq h_2$. Defina

$$I_k = I_1 + \frac{I_1 - I_2}{\frac{h_2^2}{h_1^2} - 1}$$

Muestre con varios ejemplos que el valor I_k, llamado *extrapolación de Richardson*, es una mejor aproximación que I_1 y que I_2 para la integral.

Problema 8.14

Problema 8.15

Problema 8.16

Capítulo 9

Solución numérica de ecuaciones diferenciales ordinarias

Problema 1. Encontrar una solución al problema de valor inicial (PVI)

$$\left\{ \begin{array}{rcl} x' & = & f(t,x)x(t_0) \\ x(t_0) & = & x_0\,. \end{array} \right. \tag{9.1}$$

Por ejemplo,

$$\left\{ \begin{array}{rcl} x' & = & x\ \mathrm{tang}(t+3) \\ x(-3) & = & 1\,. \end{array} \right.$$

El problema 1 tiene dos partes

(i) Existencia de soluciones,

(ii) Unidad de soluciones.

Para esto se tienen los siguientes teoremas.

Teorema 9.1 *Sea* $R \subset \mathbb{R}^2$ *un rectángulo centrado en* (t_0, x_0), *digamos*

$$R = \{(t,x) \ : \ |t-t_0| \leqslant \alpha, |x-x_0| \leqslant \beta\}\,.$$

Si $f : R \longrightarrow \mathbb{R}$ *es continua en* R, *entonces el problema (9.8) tiene una solución* $x(t)$ *para* $|t-t_0| < \min\{\alpha, \beta/M\}$ *donde* $M = \max\{|f(t,x)| \ : \ (t,x) \in R\}$

Teorema 9.2 *Si* f *y* $\dfrac{\partial f}{\partial x}$ *son continuas en*

$$R = \{(t,x) \ : \ |t-t_0| \leqslant \alpha, |x-x_0| \leqslant \beta\}\,,$$

entonces el problema (9.8) tiene una única solución $x(t)$ *para* $|t-t_0| \leqslant \min\{\alpha, \beta/M\}$.

También tenemos el teorema siguiente

Teorema 9.3 *Si f es continua en $a \leqslant t \leqslant b$ y $-\infty < x < \infty$ y satisface*

$$|f(t,x) - f(t,x_2)| \leqslant L|x_1 - x_2|$$

entonces el problema (9.8) tiene una única solución en $[a,b]$

Observación. Si $|g(x_1) - g(x_2)| \leqslant L|x_1 - x_2|$, decimos que g es una función *Lipschitz*, y L es llamada una *constante de Lipschitz* de g.

Por ejemplo, si g es derivable y g' es acotada en $[a,b]$ entonces, por el teorema del valor medio, g es Lipschitz.

En lo que sigue suponemos que el problema (9.8) tiene solución única. Tenemos entonces el siguiente problema.

Problema 2. Encontrar una solución exacta o una aproximación a una solución exacta del problema (9.8).

9.1 Método de Euler

DIBUJO

Tenemos la ecuación diferencial

$$\frac{dx}{dt} = f(t,x)$$

definimos los incrementos

$$x_{i+1} = x_i + f(t_i, x_i)h \qquad (\textit{método de Euler})$$

Ejemplo. Consideremos el problema de valores iniciales

$$\begin{cases} \dfrac{dx}{dt} &= -2t^3 + 12t^2 - 20t + 8.5\,, \qquad 0 \leqslant t \leqslant 4 \\ x(0) &= 1\,. \end{cases}$$

Integrando directamente nos queda

$$x(t) = -0.5t^4 + 4t^3 - 10t^2 + 8 + 8.5t + 1\,.$$

Tenemos

$$\frac{dx}{dt} = f(t,x) = -2t^3 + 12t^2 - 20t + 8.5$$

Luego, $x(0.5) = x(0) + f(0,1) \times 0.5$, donde $x(0) = 1$ y la pendiente en $t = 0$ es $f(0,1) = 8.5$. Luego $x_A = x(0.5) = 1 + 8.5 \times 0.5 = 5.25$ y la solución exacta en $t = 0.5$ es $x_T = x(0.5) = 3,21875$. Se deja al lector completar los cálculos.

Recordemos que el error es

$$E(x_A) = x_T - x_A$$

en nuestro caso $E(x_A) = 3.21875 - 5.25 = -2.03125$ y el error relativo (porcentaje de error)

$$E_R(x_A) = \frac{x_T - x_A}{x_T} = -\frac{2.03125}{3.21875} = -63.1\%\,.$$

9.1.1 Análisis del error para el módulo de Euler

Sea $x'(t) = \dfrac{dx}{dt} = f(t,x)$ como en el problema 9.1.

Usando desarrollo de Taylor alrededor de (t_i, x_i) obtenemos

$$x_{i+1} = x_i + x_i' h + \frac{x_i''}{2}h^2 + \cdots + \frac{x_i^{(n)}}{n!}h^n + R_n$$

donde $h = t_{i+1} - t_i$ y $R_n = \dfrac{x^{(n+1)}(\theta)}{(n+1)!}h^{n+1}$, con θ entre t_i y t_{i+1}. Escrito de otra forma

$$x_{i+1} \quad = \quad x_i + f(t_i,x_i)h + \frac{f'(t_i,x_i)}{2!}h^2 + \cdots + \frac{f^{(n-1)}(t_i,x_i)}{n!}h^n + O(h^{n+1})$$

Luego, $\text{Error} = \dfrac{f'(t_i,x_i)}{2}h^2 + \cdots$ y podemos considerar el error aproximado

$$\mathrm{E}_a = \frac{f'(t_i,x_i)}{2}h^2 \quad \text{(error de trancamiento local)}$$

Observación. Vimos que

$$\underbrace{x_{i+1} = x_i + f(t_i,x_i)h}_{\text{método de Euler}} + \underbrace{\frac{f'(t_i,x_i)}{2}h^2 + \frac{f''(t_i,x_i)}{3!}h^3 + \cdots}_{\text{error}}$$

Ahora, para estimar el error aproximado hasta el orden 3, tenemos que calcular $f''(t_i,x_i)$. Sabemos del cálculo diferencial que

$$f'(t_i, x_i) = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\frac{dx}{dt}$$

luego,

$$f''(t_i, x_i) = \frac{\partial}{\partial t}\left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\cdot\frac{dx}{dt}\right) + \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\frac{dx}{dt}\right)$$

desarrollando esto, se obtiene una expresión larga y no muy cómoda para $f''(t, x)$.

9.2 Método de Heun

DIBUJO

Tenemos

$$x_{i+1} = f(t_i, x_i)\,,$$

una extrapolación lineal de x_{i+1} es dada por

$$x^0_{i+1} = x_i + f(t_i, x_i)h \qquad (\text{ ec. predictor})$$

y una mejora de x_{i+1} que permite el cálculo de una estimación de la pendiente es

$$x_{i+1} = f(t_{i+1}, x^0_{i+1})\,.$$

Luego, podemos considerar la pendiente promedio

$$\overline{x}'_i = \frac{x'_i + x'_{i+1}}{2} = \frac{f(t_i, x_i) + f(t_{i+1}, x^0_{i+1})}{2}$$

y usaremos esta pendiente para extropolar linealmente desde x_i a x_{i+1} usando el método de Euler, es decir,

$$x_{i+1} = x_i + \frac{f(t_i, x_i) + f(t_{i+1}, x^0_{i+1})}{2}\,h \qquad (\text{ (ec. corrector))}$$

En resumen, el método de Heun nos queda como

$$\begin{cases} x_{i+1}^0 &= x_i + f(t_i, x_i)h & \text{(ec. Predictor)} \\ x_{i+1} &= x_i + \dfrac{f(t_i, x_i) + f(t_{i+1}, x_{i+1}^0)}{2}\, h & \text{(ec. Corrector)} \end{cases}$$

Escrito de otra forma, nos queda

$$x_{i+1} = x_i + \frac{f(t_i, x_i) + f(t_{i+1}, x_i + hf(t_i, x_i))}{2}\, h$$

9.3 Método del punto medio o método de Euler mejorado

En este caso consideramos

$$\begin{cases} x_{i+1/2} &= x_i + f(t_i, x_i)\dfrac{h}{2} \\ t_{i+1/2} &= t_i + \dfrac{h}{2} = \dfrac{t_i + t_{i+1}}{2}\,. \end{cases}$$

DIBUJO

Usaremos ese valor promedio para calcular la pendiente en el punto medio $(t_{i+1/2}, x_{i+1/2})$ y después extrapolamos a x_{i+1}, usando este valor intermedio, es decir, tenemos

$$x_{i+1} = x_i + f(t_{i+1/2}, x_{i+1/2})\frac{h}{2} \qquad \text{(Euler mejorado)}$$

esto es,

$$x_{i+1} = x_i + f(t_i + \frac{h}{2}, x_i + \frac{h}{2}\, f(t_i, x_i))\, \frac{h}{2}\,.$$

9.4 Métodos de Runge–Kutta

Consideramos una forma generalizada para lo anterior, escribiendo

$$x_{i+1} = x_i + \phi(t_i, x_i, h)h\,,$$

es decir, ϕ depende del incremento h. La función ϕ es llamada *función incremento*. En general esta se escribe como

$$\phi = a_1 k_1 + a_2 k_2 + \cdots + a_n k_n \,,$$

donde $a_i\,,\ i = 1, 2, \ldots, n\,,$ son constantes y

$$\left\{ \begin{array}{lcl} k_1 & = & f(t_i, x_i) \\ k_2 & = & f(t_i + p_1 h, x_i + q_{1,1} k_1 h) \\ k_3 & = & f(t_i + p_2 h, x_i + (q_{2,1} k_1 + q_{2,2} k_2) h) \\ & \vdots & \\ k_n & = & f(t_i + p_{n-1} h, x_i + (q_{n-1,1} k_1 + q_{n-1,2} k_2 + \cdots + q_{n-1,n-1} k_{n-1}) h) \,. \end{array} \right.$$

Note que los k_j son relaciones de recurrencias.

Problema. Hacer una elección razonable para tener relaciones manejables y una buena aproximación a la solución. En esto consiste una de las técnicas más conocidas, nos referimos a las técnicas de Runge–Kutta.

9.4.1 Runge–Kutta de orden 2

En este caso, tenemos

$$x_{i+1} = x_i + (a_1 k_1 + a_2 k_2) h \qquad \text{(R-K2,1)}$$

donde

$$\left\{ \begin{array}{lcl} k_1 & = & f(t_i, x_i) \\ k_2 & = & f(t_i + p_1 h, x_i + q_{1,1} k_1 h) \,. \end{array} \right.$$

Problema. Encontrar $a_1\,,\ a_2\,,\ p_1$ y $q_{1,1}\,.$

Estos se calculan igualando el término de segundo orden de (R-K2,1) con la expansión de Taylor de f alrededor del punto $(t_i, x_i)\,.$ De esto se obtienen las ecuaciones

$$\left\{ \begin{array}{lcl} a_1 + a_2 & = & 1 \\ a_2 q_{1,1} & = & \dfrac{1}{2} = a_2 p_1 \end{array} \right.$$

de donde

$$\left\{ \begin{array}{lcl} a_1 & = & 1 - a_2 \\ p_1 & = & q_{1,1} = \dfrac{1}{2a_2} \,, \end{array} \right.$$

y a_2 lo podemos elegir libremente.

Por ejemplo, considerando $a_2 = \dfrac{1}{2}$ obtenemos $a_1 = \dfrac{1}{2}$ y $p_1 = q_{1,1} = 1\,,$ y nos queda

$$x_{i+1} = x_i + (k_1 + k_2)\,\frac{h}{2} \qquad \text{(método de Heun con un sólo corrector)}$$

donde

$$\begin{cases} k_1 &= f(t_i, x_i) \\ k_2 &= f(t_i + h, x_i + k_1 h)\,, \end{cases}$$

es decir,

$$x_{i+1} \;=\; x_i + (f(t_i, x_i) + f(t_i + h, x_i + hf(t_i, x_i))\,\frac{h}{2}\,.$$

Considerando $a_2 = 1$ nos queda $a_1 = 0$ y $p_1 = q_{1,1} = 1/2$, y obtenemos

$$\begin{aligned} x_{i+1} &= x_i + k_2 h && \text{(método de punto medio)} \\ &= x_i + f(t_i + \frac{h}{2}, x_i + \frac{h}{2}\, f(t_i, x_i)) \end{aligned}$$

donde

$$\begin{cases} k_1 &= f(t_i, x_i) \\ \\ k_2 &= f\left(t_i + \dfrac{h}{2}, x_i + \dfrac{1}{2} k_1 h\right)\,. \end{cases}$$

9.4.2 Método de Ralston

Tomando $a_2 = \dfrac{2}{3}$, obtenemos $a_1 = \dfrac{1}{3}$, $p_1 = q_{1,1} = \dfrac{3}{4}$, y

$$x_{i+1} = x_i + \left(\frac{1}{3} k_1 + \frac{2}{3} k_2\right) h$$

donde

$$\begin{cases} k_1 &= f(t_i, x_i) \\ \\ k_2 &= f\left(t_i + \dfrac{3}{4} h, x_i + \dfrac{3}{4} k_1 h\right)\,. \end{cases}$$

9.4.3 Método de Runge–Kutta de orden 3

Esta vez, consideramos desarrollos de Taylor hasta orden 3. Obtenemos

$$x_{i+1} = x_i + \frac{1}{6}(k_1 + 4k_2 + k_3)h \qquad \text{(R-K3)}$$

donde

$$\begin{cases} k_1 &= f(t_i, x_i) \\ k_2 &= f\left(t_i + \frac{1}{2} h, x_i + \frac{1}{2} k_1 h\right) \\ k_3 &= f\left(t_i + h, x_i - k_1 h + 2k_2 h\right) \end{cases}$$

9.4.4 Método de Runge–Kuta de orden 4

Esta vez, consideramos desarrollos de Taylor hasta orden 4. Obtenemos

$$x_{i+1} = x_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)h \qquad \text{(R-K4)}$$

donde

$$\begin{cases} k_1 &= f(t_i, x_i) \\ k_2 &= f\left(t_i + \frac{h}{2}, x_i + \frac{1}{2}k_1 h\right) \\ k_3 &= f\left(t_i + \frac{h}{2}, x_i + \frac{1}{2}k_2 h\right) \\ k_4 &= f(t_i + h, x_i + k_3 h) \end{cases}$$

9.4.5 Método de Runge–Kuta de orden superior

Considerando dearrollos de Taylor hasta order 6, obtenemos

$$x_{i+1} = x_i + \frac{1}{90}(7k_1 + 32k_3 + 12k_4 + 32k_5 + 7k_6)h \qquad \text{(R-K6) (método de Butcher)}$$

donde

$$\begin{cases} k_1 &= f(t_i, x_i) \\ k_2 &= f(t_i + \frac{1}{4}h, x_i + \frac{1}{4}k_1 h) \\ k_3 &= f(t_i + \frac{1}{4}h, x_i + \frac{1}{8}k_1 h + \frac{1}{8}k_2 h) \\ k_4 &= f(t_i + \frac{h}{2}, x_i - \frac{1}{2}k_2 h + k_3 h) \\ k_5 &= f(t_i + \frac{3}{4}h, x_i + \frac{3}{16}k_1 h + \frac{9}{16}k_4 h) \\ k_6 &= f(t_i + h, x_i - \frac{3}{7}k_1 h + \frac{2}{7}k_2 h + \frac{12}{7}k_3 h - \frac{12}{7}k_4 h + \frac{8}{7}k_5 h)\,. \end{cases}$$

9.5 Métodos multipaso

Consideremos el problema

$$\begin{cases} x' &= f(t, x)\,, \quad a \leqslant t \leqslant b \\ x(a) &= \alpha\,. \end{cases}$$

Un método multipaso de paso m para resolver este problema de valor inicial es uno cuya ecuación de diferencia para obtener x_{i+1} en el punto t_{i+1} puede representarse por la siguiente

ecuación, donde m es un entero mayor o igual que 1.

$$\begin{aligned} x_{i+1} &= a_{m-1}x_i + a_{m-2}x_{i-1} + \cdots + a_0 x_{i-(m-1)} \\ &\quad + h[b_m f(t_{i+1}, x_{i+1}) + b_{m-1} f(t_i, x_i) \\ &\quad + \cdots + b_0 f(t_{i-(m-1)}, x_{i-(m-1)})] \,. \end{aligned} \tag{9.2}$$

para $i = m-1, m, \ldots, N-1$, donde $h = \dfrac{b-a}{N}$, $a_0, a_1, \ldots, a_{m-1}$ y $b_0, b_1, \ldots, b_m$ son constantes y se especifican los valores iniciales

$$x_0 = \alpha,\ x_1 = \alpha_1,\ x_2 = \alpha_2, \ldots, x_{m-1} = \alpha_{m-1} \,. \tag{9.3}$$

Cuando $b_m = 0$, el método es *explícito*, o *abierto* ya que (9.3) da entonces x_{i+1} de manera explícita en término de los valores previamente determinados. Cuando $b_m \neq 0$ el método es *implícito* o *cerrado* ya que x_{i+1} se encuentra en ambos lados de la ecuación.

Por ejemplo, las ecuaciones

$$\begin{cases} x_0 &= \alpha,\ x_1 = \alpha_1,\ x_2 = \alpha_2,\ x_3 = \alpha_3 \\ x_{i+1} &= x_i + \dfrac{h}{24}\,[\,55 f(t_i, x_i) - 59 f(t_{i-1}, x_{i-1}) + 37 f(t_{i-2}, x_{i-2}) - 9 f(t_{i-3}, x_{i-3})\,] \end{cases} \tag{9.4}$$

donde $i = 3, 4, \ldots, N-1$ definen un método explícito de 4 pasos, llamado *método de Adam-Bashforth de cuarto orden.*

Las ecuaciones

$$\begin{cases} x_0 &= \alpha,\ x_1 = \alpha_1,\ x_2 = \alpha_2 \\ x_{i+1} &= x_i + \dfrac{h}{24}\,[\,9 f(t_{i+1}, x_{i+1}) - 19 f(t_i, x_i) - 5 f(t_{i-1}, x_{i-1}) + f(t_{i-2}, x_{i-2})\,] \end{cases} \tag{9.5}$$

donde $i = 2, 3, \ldots, N-1$, definen un método implícito de 3 pasos llamado *método de Adam-Moulton de cuarto orden.*

Tanto en (9.4) o (9.5) deben especificarse los valores iniciales, generalmente suponemos que $x_0 = \alpha$ y se generan los valores residuales por un método de Runge–Kutta o por otro método de 1 paso, por ejemplo, por el método de punto medio

$$\begin{cases} x_0 &= \alpha \\ x_{i+1} &= x_i + h f(t_i + \dfrac{h}{2}, x_i + \dfrac{h}{2} f(t_i, x_i)) \end{cases}$$

donde $i = 0, 1, \ldots, m-1$.

o por el método de Euler modificado

$$\left\{ \begin{array}{rcl} x_0 & = & \alpha \\ x_{i+1} & = & x_i + \dfrac{h}{2}\left[f(t_i, x_i) + f(t_{i+1}, x_i + hf(t_i x_i))\right] \end{array} \right.$$

con $i = 0, 1, \ldots, m-1$,

o por el método de Heun

$$\left\{ \begin{array}{rcl} x_0 & = & \alpha \\ x_{i+1} & = & x_i + \dfrac{h}{4}[f(t_i, x_i) + 3f(t_i + \dfrac{2}{3}h, x_i + \dfrac{2}{3}hf(t_i, x_i))] \end{array} \right.$$

o por el método de Runge–Kutta de orden 4

$$\left\{ \begin{array}{rcl} x_0 & = & \alpha \\ k_1 & = & hf(t_i + \dfrac{h}{2}, x_i + \dfrac{1}{2}k_1) \\ k_3 & = & hf(t_i + \dfrac{h}{2}, x_i + \dfrac{1}{2}k_2) \\ k_4 & = & hf(t_{i+1}, x_i + k_3) \\ x_{i+1} & = & x_i + \dfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \end{array} \right.$$

donde $i = 0, 1 \ldots, m-1$

En el problema (9.8)

$$\left\{ \begin{array}{rcl} x' & = & f(t, x)\,, \quad a \leqslant t \leqslant b \\ x(a) & = & \alpha\,. \end{array} \right.$$

Si

$$\begin{array}{rcl} x_{i+1} & = & a_{m-1}x_i + a_{m-2}x_{i-1} + \cdots + a_0 x_{i-(m-1)} + h[b_m f(t_{i+1}, x_{i+1}) + b_{m-1}f(t_i, x_i) + \\ & & \cdots + b_0 f(t_{i-(m-1)}, x_{i-(m-1)})] \end{array}$$

es el $(i+1)$–paso en un método de multipaso, entonces el error local de truncamiento en este paso es dado por

$$\begin{array}{rcl} \tau_{i+1}(h) & = & \dfrac{x(t_{i+1}) - a_{m-1}x(t_i) - \cdots - a_0 x(t_{i-(m-1)})}{h} \\ & & -b_m[f(t_{i+1}, x(t_{i+1})) + \cdots + b_0 f(t_{i-(m-1)}, x(t_{i-(m-1)}))] \end{array}$$

$i = m-1, m, \ldots, N-1$.

Tenemos los siguientes casos particulares.

9.5.1 Método de Adams–Bashforth de dos pasos

$$\begin{aligned} x_0 &= \alpha,\ x_1 = \alpha_1 \\ x_{i+1} &= x_i + \frac{h}{2}(3f(t_i, x_i) - f(t_{i-1}, x_{i-1})) \end{aligned}$$

$i = 1, 2, \ldots, N-1$. Su error de truncamiento local es

$$\tau_{i+1}(h) = \frac{5}{12}x'''(\mu_i)h^2, \quad \mu_i \in]t_{i-1}, t_{i+1}[\ .$$

9.5.2 Método de Adams–Bashforth de 3 pasos

$$\begin{cases} x_0 &= \alpha,\ x_1 = \alpha_1\ ,\ x_2 = \alpha_2 \\ x_{i+1} &= x_i + \dfrac{h}{12}\left[\, 23f(t_i, x_i) - 16f(t_{i-1}, x_{i-1}) + 5f(t_{i-2}, x_{i-2})\,\right] \end{cases}$$

$i = 2, 3, \ldots, N-1$. Su error de truncamiento local es

$$\tau_{i+1}(h) = \frac{3}{8}x^{(4)}(\mu_i)h^3, \quad \mu_i \in]t_{i-2}, t_{i+1}[\ .$$

9.5.3 Método de Adams–Bashforth de 4 pasos

$$\begin{cases} x_0 &= \alpha, x_1 = \alpha_1,\ x_2 = \alpha_2,\ x_3 = \alpha_3 \\ x_{i+1} &= x_i + \dfrac{h}{24}\left[\, 55f(t_i, x_i) - 59f(t_{i-1}, x_{i-1}) + 37f(t_{i-2}, x_{i-2}) - 9f(t_{i-3}, x_{i-3})\right] \end{cases}$$

$i = 3, 4, \ldots, N-1$. Su error local de truncamiento es dado por

$$\tau_{i+1}(h) = \frac{251}{720}x^{(5)}(\mu_i)h^4\,, \quad \mu_i \in]t_{i-3}, t_{i+1}[\ .$$

9.5.4 Método de Adams–Bashforth de 5 pasos

$$\begin{cases} x_0 &= \alpha,\ x_1 = \alpha_1,\ x_2 = \alpha_2,\ x_3 = \alpha_3,\ x_4 = \alpha_4 \\ x_{i+1} &= x_i + \dfrac{h}{720}[1901f(t_i, x_i) - 2774f(t_{i-1}, x_{i-1}) + 2616f(t_{i-2}, x_{i-2}) - 1274f(t_{i-3}, x_{i-3}) \\ & \quad +251f(t_{i-4}, x_{i-4})] \end{cases}$$

$i = 4, 5, \ldots, N-1$. Su error local de truncamiento es

$$\tau_{i+1}(h) = \frac{95}{288}x^{(6)}(\mu_i)h^5\,, \qquad \mu_i \in]t_{i-4}, t_{i+1}[$$

Los métodos implícitos se derivan usando $(t_{i+1}, f(t_{i+1}, x(t_{i+1})))$ como un nodo de interpolación adicional en el cual se usa una aproximación de la integral

$$\int_{t_i}^{t_{i+1}} f(t, x(t))dt\,,$$

la cual se puede aproximar por diferentes métodos, por ejemplo, regla de los trapecios, regla de Simpson, o regla de Simpson $\left(\frac{3}{8}\right)$, ...

9.5.5 Método de Adams–Moulton de dos pasos

$$\left\{\begin{array}{rcl} x_0 & = & \alpha,\ x_1 = \alpha_1 \\ x_{i+1} & = & x_i + \dfrac{h}{12}\left[\,5f(t_{i+1}, x_{i+1}) + 8f(t_i, x_i) - f(t_{i-1}, x_{i-1})\,\right] \end{array}\right.$$

donde $i = 1, 2, \ldots, N-1$. Su error de truncamiento es dado por

$$\tau_{i+1}(h) = -\frac{1}{24}x^{(4)}(\mu_i)h^3, \quad \mu_i \in]t_{i-1}, t_{i+1}[\,.$$

9.5.6 Método de Adams–Moulton de tres pasos

$$\left\{\begin{array}{rcl} x_0 & = & \alpha,\ x_1 = \alpha_1,\ x_2 = \alpha_2 \\ x_{i+1} & = & x_i + \dfrac{h}{24}[9f(t_{i+1}, x_{i+1}) + 19f(t_i, x_i) - 5f(t_{i-1}, x_{i-1}) + f(t_{i-2}, x_{i-2})] \end{array}\right.$$

donde $i = 2, 3, \ldots, N-1$. Su error local de truncamiento es dado por

$$\tau_{i+1}(h) = -\frac{19}{720}x^{(5)}(\mu_i)h^4, \quad \mu_i \in]t_{i-2}, t_{i+1}[\,.$$

9.6 Sistemas de ecuaciones diferenciales

Un sistema de ecuaciones diferenciales es un sistema del tipo

$$(S1) \qquad \left\{\begin{array}{rcl} \dfrac{du_1}{dt} & = & f_1(t, u_1, \ldots, u_m) \\ \dfrac{du_2}{dt} & = & f_2(t, u_1, \ldots, u_m) \\ & \vdots & \\ \dfrac{du_m}{dt} & = & f_m(t, u_1, \ldots, u_m) \end{array}\right. \tag{9.6}$$

con $a \leqslant t \leqslant b$ y condiciones inicales

$$(CI1) \qquad u_1(a) = \alpha_1,\ u_2(a) = \alpha_2,\ \ldots,\ u_m(a) = \alpha_m\,.$$

Problema. Encontrar las m funciones $u_1, \ldots, u_m$ que satisfacen el sistema $(S1)$.

Para resolver este problema, introducimos algunas notaciones. Sea

$$D = \{(t, u_1, \ldots, u_m) \ : \ a \leqslant t \leqslant b, \ -\infty < u_i < \infty, \ i = 1, \ldots, m\}\,.$$

Decimos que una función $f(t, u_1, \ldots, u_m)$ satisface una *condición de Lipschtz* en D en las variables $u_1, \ldots, u_m$ si

$$|f(t, u_1, \ldots, u_m) - f(t, z_1, \ldots, z_m)| \leqslant L \sum_{i=1}^{m} |u_i - z_i|\,,$$

para alguna constante $L \geqslant 0$ y $(t, u_1, \ldots, u_m),\ (t, z_1, \ldots, z_m) \in D$.

Observemos que si

$$\left| \frac{\partial f}{\partial u_i}(t, u_1, \ldots, u_m) \right| \leqslant L$$

para $i = 1, \ldots, m$ y todo $(t, u_1, \ldots, u_m) \in D$. Entonces, por el teorema del valor medio, f satisface una condición de Lipschitz en D.

Tenemos el siguiente teorema.

Teorema 9.4 *Sea* $D = \{(t, u_1, \ldots, u_m) \ : \ a \leqslant t \leqslant b, \ -\infty < u_i < \infty, \ i = 1, \ldots, m\}$ *y sean* $f_i : D \longrightarrow \mathbb{R}$, $i = 1, \ldots, m$. *Supongamos que para cada* $i = 1, \ldots, m$ *la función* f_i *es continua en* D *y satisface una condición de Lipschitz en* D *o* $\frac{\partial f}{\partial u_i}(t, u_1, \ldots, u_m)$ *es continua para* $i = 1, \ldots, m$ *y* $a \leqslant t \leqslant b$. *Entonces el sistema* (9.6) *con las condiciones iniciales* (9.6) *tiene solución única* $u_1(t), \ldots, u_m(t)$, $a \leqslant t \leqslant b$.

Para encontrar aproximaciones a las soluciones del sistema (9.6) numéricamente, podemos utilizar, por ejemplo, el método de Runge–Kutta

$$\begin{cases} x_0 &= \alpha \\ k_1 &= f(t_i, x_i) \\ k_2 &= f\left(t_i + \dfrac{h}{2}, x_i + \dfrac{1}{2}k_1\right) \\ k_3 &= f\left(t_i + \dfrac{h}{2}, x_i + \dfrac{1}{2}k_2\right) \\ k_4 &= f(t_1 + h, x_i + k_3) \end{cases}$$

y entonces

$$x_{i+1} = x_i + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4)\,, \qquad i = 0, 1, \ldots, N-1$$

que resuelve numéricamente el problema

$$\begin{cases} x'(t) &= f(t, x)\,, \qquad a \leqslant t \leqslant b \\ x(a) &= \alpha\,. \end{cases}$$

Este se generaliza como sigue. Sea $N \geqslant 1$ y sea $h = \dfrac{b-a}{N}$. La partición de $[a,b]$ en N subintervalos con nodos

$$t_j = a + jh, \qquad j = 0, 1, \ldots, N$$

es dada.

Usaremos la notación $x_{i,j}$ para denotar una aproximación a $u_i(t_j)$, $j = 0, 1, \ldots, N$, $i = 1, \ldots, m$, es decir, $x_{i,j}$ aproxima la i–ésima solución $u_i(t)$ de (9.6) en el punto t_j de los nodos. Usamos la notación

$$x_{1,0} = \alpha_1, \quad x_{2,0} = \alpha_2, \ldots, x_{m,0} = \alpha_m$$

para las condiciones iniciales.

Si hemos calculado los valores $x_{1,j}$, $x_{1,j}$, $x_{2,j}, \ldots, x_{m,j}$. Obtenemos $x_{1,j+1}$, $x_{2,j+1}, \ldots, x_{m,j+1}$ como sigue. Calculando primero, para $i = 1, \ldots, m$

$$\begin{cases} k_{1,i} &= hf(t_j, x_{1,j}, x_{2,j}, \ldots, x_{m,j}) \\ k_{2,i} &= hf\left(t_j + \dfrac{h}{2}, x_{1,j} + \dfrac{1}{2}k_{1,1}, x_{2,j} + \dfrac{1}{2}k_{1,2}, \ldots, u_{m,j} + \dfrac{1}{2}k_{1,m}\right) \\ k_{3,i} &= hf_i\left(t_j + \dfrac{h}{2}, x_{1,j} + \dfrac{1}{2}k_{2,1}, x_{2,j} + \dfrac{1}{2}k_{2,2}, \ldots, x_{m,j} + \dfrac{1}{2}k_{2,m}\right) \\ k_{4,i} &= hf_i\left(t_j + h, x_{1,j} + k_{3,1}, x_{2,j} + k_{3,2}, \ldots, x_{m,j} + k_{3,m}\right) \end{cases}$$

hacemos entonces, para $, i = 1, \ldots, m$

$$x_{i,j+1} = x_{i,j} + \frac{1}{6}(k_{1,i} + 2k_{2,i} + 2k_{3,i} + k_{4,i}\,.$$

Note que antes de determinar los términos $k_{2,i}$ deben calcularse todos los valores $k_{\ell,1}, k_{\ell,2}, \ldots, k_{\ell,m}$. En general, estos deben calcularse antes que cualquiera de las expresiones $k_{\ell+1,i}$.

9.7 Ecuaciones diferenciales de orden superior

Un problema de valores iniciales

$$(OS1) \qquad \begin{cases} x^{(m)}(t) &= f(t, x, x', \ldots, x^{(m-1)}), \qquad a \leqslant t \leqslant b \\ x(a) &= \alpha_1,\ x'(a) = \alpha_2,\ \ldots\ , x^{(m-1)}(a) = \alpha_m \end{cases}$$

Se transforma en un sistema de ecuaciones diferenciales, como el (9.6) mediante la introducción de nuevas variables. Sean

$$\begin{cases} u_1(t) &= x(t) \\ u_2(t) &= x'(t) \\ & \vdots \\ u_m(t) &= x^{(m-1)}(t) \end{cases}$$

El problema (9.7) se transforma entonces en el siguiente sistema de ecuaciones diferenciales.

$$\begin{cases} \dfrac{du_1}{dt} &= \dfrac{dx}{dt} = u_2 \\ \dfrac{du_2}{dt} &= \dfrac{dx'}{dt} = x'' = u_3 \\ & \vdots \\ \dfrac{du_{m-1}}{dt} &= \dfrac{dx^{(m-2)}}{dt} = x^{(m-1)} = u_m \\ \dfrac{du_m}{dt} &= \dfrac{dx^{(m-1)}}{dt} = x^{(m)} = f(t, u_1, u_2, \ldots, u_m) \end{cases}$$

con condiciones iniciales

$$\begin{cases} u_1(a) &= x(a) = \alpha_1 \\ u_2(a) &= x'(a) = \alpha_2 \\ & \vdots \\ u_m(a) &= x^{(m-1)}(a) = \alpha_m \end{cases}$$

y podemos aplicar lo anterior para resolver este sistema y así obtener la solución del problema (9.7).

9.8 Problemas con valores en la frontera para ecuaciones diferenciales ordinarias

Problema 1 Resolver el siguiente problema con valores en la frontera

$$\begin{cases} x'' &= f(t, x, x')\,, \qquad a \leqslant t \leqslant b \\ x(a) &= \alpha \\ x(b) &= \beta\,. \end{cases} \tag{9.7}$$

Antes de continuar con el problema (9.7), recordemos algunos resultados de ecuaciones diferenciales.

Teorema 9.5 *Supongamos que f es continua en el conjunto $D = \{(t, x, x') \,:\, a \leqslant t \leqslant b\,,\ -\infty < x < \infty\,,\ -\infty < x' < \infty\}$ y que $\dfrac{\partial f}{\partial x}\,,\ \dfrac{\partial f}{\partial x'}$ son continuas en D. Si*

(i) $\dfrac{\partial f}{\partial x}(t,x,x') > 0$ *para todo* $(t,x,x') \in D$, *y*

(ii) existe una constante $M \geqslant 0$ *tal que*

$$\left|\frac{\partial f}{\partial x'}(t,x,x')\right| \leqslant M$$

para todo $(t,x,x') \in D$.

Entonces el problema (9.7) tiene solución única en D .

Ejemplo. El problema con valores en la frontera

$$\begin{cases} x'' + e^{-tx} + \operatorname{sen}(x') &= 0\,, \quad 1 \leqslant t \leqslant 2 \\ x(1) &= 0 \\ x(2) &= 0 \end{cases}$$

tiene solución única en $D = \{(t,x,x') \,:\, 1 \leqslant t \leqslant 2\,,\ -\infty < x < \infty\,,\ -\infty < x' < \infty\}$.

En efecto, tenemos que $f(t,x,x') = -e^{-tx} - \operatorname{sen}(x')$ es continua. Ahora, $\frac{\partial f}{\partial x}(t,x,x') = te^{-tx} > 0$ en D y $\left|\frac{\partial f}{\partial x'}(t,x,x')\right| = |-\cos(x')| \leqslant 1$.

Notemos que en el caso particular en que $f(t,x,x')$ tiene la forma

$$f(t,x,x') = p(t)x' + q(t)x + r(t)\,,$$

la ecuación $x'' = f(t,x,x')$ es llamada *lineal*. Este tipo de ecuaciones ocurre frecuentemente en problemas de la Física.

Corolario 9.1 *Si el problema lineal con valores en la frontera*

$$\begin{cases} x'' &= p(t)x' + q(t)x + r(t) \qquad a \leqslant t \leqslant b \\ x(a) &= \alpha \\ x(b) &= \beta \end{cases}$$

donde las funciones $p(t)$, $q(t)$ *y* $r(t)$ *satisfacen*

(i) $p(t)$, $q(t)$ *y* $r(t)$ *son continuas en* $[a,b]$,

(ii) $q(t) > 0$ *en* $[a,b]$.

Entonces el problema tiene solución única.

9.8.1 Método del disparo para el problema lineal

Para aproximar la solución única garantizada bajo las hipótesis del corolario anterior, primero consideramos los problemas de valores iniciales

$$(PVI,1) \quad \left\{ \begin{array}{rcl} x'' & = & p(t)x' + q(t)x + r(t)\,, \quad a \leqslant t \leqslant b \\ x(a) & = & \alpha \\ x'(a) & = & 0 \end{array} \right.$$

y

$$(PVI,2) \quad \left\{ \begin{array}{rcl} x'' & = & p(t)x' + q(t)x\,, \quad a \leqslant t \leqslant b \\ x(a) & = & 0 \\ x'(a) & = & 1\,. \end{array} \right.$$

Bajos las hipótesis del corolario, ambos problemas tienen solución única. Si $x_1(t)$ denota la solucón del $(PVI,1)$ y $x_2(t)$ denota la solución del $(PVI,2)$, entonces si $x(t) = x_1 + Cx_2(t)$ es una solución de la ecuación $x'' = p(t)x' + q(t)x + r(t)$, la verificación de esto es inmediata. Usando las condiciones de frontera, $x(q) = x_1(a) + Cx_2(a) = \alpha + 0 = \alpha$ y $x(b) = x_1(b) + Cx_2(b) = \beta$, por lo tanto $C = \frac{\beta - x_1(b)}{x_2(b)}$. En conclusión, la solución buscada es

$$x(t) = x_1(t) + \frac{\beta - x_1(b)}{x_2(b)}\, x_2(t)\,, \quad \text{si } x_2(b) \neq 0,$$

es la solución del problema con valores en la frontera que estamos estudiando. En las hipótesis del Corolario 9.1 no se da el caso problemático de tener $x_2(t) \equiv 0$,

Ejemplo 74

$$\left\{ \begin{array}{rcll} x'' & = & \frac{2t}{1+t^2}x' - \frac{2}{1+t^2}x + 1 & 0 \leqslant t \leqslant 4 \\ x(0) & = & 1.25 & \\ x(4) & = & -0.95 & \end{array} \right.$$

Otra forma para aproximar la solución del problema con valores en la frontera

$$\left\{ \begin{array}{rcl} -x'' + p(t)x' + q(t)x + r(t) & = & 0\,, \quad a \leqslant t \leqslant b \\ x(a) & = & \alpha \\ x(b) & = & \beta \end{array} \right.$$

es como sigue. Producimos un sistema de ecuaciones de primer orden en el que no figura explicitamente t, para ello definimos

$$\left\{ \begin{array}{rcl} x_0 & = & t \\ x_3 & = & x_1' \\ x_4 & = & x_2', \end{array} \right.$$

donde x_1 es la solución del $(PVI,3)$ y x_2 es la solución del $(PVI,4)$ siguientes

$$(PVI,3) \quad \left\{ \begin{array}{rcl} x'' & = & f(t,x,x')\,, \quad a \leqslant t \leqslant b \\ x(a) & = & \alpha \\ x'(a) & = & 0 \end{array} \right.$$

y

$$(PVI,4) \quad \left\{ \begin{array}{rcll} x'' & = & f(t,x,x')\,, & a \leqslant t \leqslant b \\ x(a) & = & \alpha & \\ x'(a) & = & 1\,. & \end{array} \right.$$

Obtenemos así el sistema de ecuaciones

$$\left\{ \begin{array}{rcll} x_0' & = & 1 & x_0(a) = a \\ x_1' & = & x_3 & x_1(a) = \alpha \\ x_2' & = & x_4 & x_2(a) = \alpha \\ x_3' & = & f(x_0, x_1, x_3) & x_3(a) = 0 \\ x_4' & = & f(x_0, x_2, x_4) & x_4(a) = 1 \end{array} \right.$$

el cual puede resolverse, usando por ejemplo, Runge–Kutta.

9.9 El método del disparo para el problema no lineal de valores en la frontera

Consideremos el problema de valores en la frontera

$$\left\{ \begin{array}{rcll} x'' & = & f(t,x,x')\,, & a \leqslant t \leqslant b \\ x(a) & = & \alpha & \\ x(b) & = & \beta\,. & \end{array} \right.$$

donde f satisface las hipótesis del Teorema 9.5.

En este caso usamos las soluciones de una sucesión de problemas de valor inicial

$$\left\{ \begin{array}{rcll} x'' & = & f(t,x,x')\,, & a \leqslant t \leqslant b \\ x(a) & = & \alpha & \\ x'(a) & = & s & \end{array} \right.$$

donde s es un parámetro. Elegimos $s = s_k$ de modo que $\lim_{k\to\infty} x(b, s_k) = x(b) = \beta$, donde $x(t, s_k)$ denota la solución del problema de valores iniciales asociado en este caso con $s = s_k$ y $x(t)$ denota la solución del problema de valores de frontera.

"Comenzamos con un parámetro s_0 que determina la elevación inicial con la cual se dispara al objetivo desde el punto (a, α) a lo largo de la curva descrita por la solución del problema de valor inicial

$$\left\{ \begin{array}{rcll} x'' & = & f(t,x,x')\,, & a \leqslant t \leqslant b \\ x(a) & = & \alpha & \\ x'(a) & = & s_0\,. & \end{array} \right.$$

Si $x(b, s_0)$ no está suficientemente cerca de β corregimos la elevación, tomando $s_1, s_2, \ldots, ,$ hasta estar suficientemente próximo al blanco".

9.9.1 Determinación de los parámetros s_k

Problema. Determinar s tal que

$$g(b,s) = x(b,s) - \beta \quad = \quad 0$$

(ecuación no lineal). Podemos usar, por ejemplo el método de la secante, con valores iniciales s_0 y s_1,

$$\begin{aligned} s_k &= s_{k-1} - \frac{g(b,s_{k-1})(s_{k-1}-s_{k-2})}{g(b,s_{k-1}) - g(b,s_{k-2})} \\ &= s_{k-1} - \frac{(x(b,s_{k-1})-\beta)(s_{k-1}-s_{k-2})}{x(b,s_{k-1}) - x(b,s_{k-2})}, \quad k = 2,3,\ldots \end{aligned}$$

o el método de Newton

$$s_k \quad = \quad s_{k-1} - \frac{x(b,s_{k-1})-\beta}{\frac{d}{ds}x(b,s_{k-1})}$$

aquí el problema es obtener una expresión explicíta para $\frac{d}{ds}x(b,s)$.

9.10 Método de diferencias finitas para problemas lineales

Consideremos primero el caso lineal

$$\left\{ \begin{array}{lcl} x'' & = & p(t)x' + q(t)x + r(t), \qquad a \leqslant t \leqslant b \\ x(a) & = & \alpha \\ x(b) & = & \beta \end{array} \right.$$

Dividimos el intervalo $[a,b]$ en $N+1$ puntos $t_i = a+ih$, $i = 0,1,\ldots,N+1$, donde $h = \frac{b-a}{N}$. Para $i = 1,2,\ldots,N$ la ecuación diferencial a aproximar es

$$x''(t_i) = p(t_i)x'(t_i) + q(t_i)x(t_i) + r(t_i).$$

Usando las fórmulas para aproximar derivadas tenemos, por ejemplo,

$$x''(t_i) = \frac{1}{h^2}\left(x(t_{i+1}) - 2x(t_i) + x(t_{i-1})\right) - \frac{h^2}{12}x^{(4)}(\theta_i),$$

donde $\theta_i \in]t_{i-1}, t_{i+1}[$ (fórmula de las diferencias centradas). Analogamente,

$$x'(t_i) = \frac{1}{2h}\left(x(t_{i+1}) - x(t_{i-1})\right) - \frac{h^2}{6}x'''(\xi_i),$$

donde $\xi_i \in]t_{i-1}, t_{i+1}[$. Reemplazando, obtenemos

$$\begin{aligned}\frac{x(t_{i+1})-2x(t_i)+x(t_{i-1})}{h^2} &= p(t_i)\left(\frac{x(t_{i+1})-x(t_{i-1})}{2h}\right)+q(t_i)x(t_i)+r(t_i)\\ &\quad -\frac{h^2}{12}(2p(t_i)x'''(\xi_i)-x^{(4)}(\theta_i))\,.\end{aligned}$$

Un método de diferencias finitas con error de truncamiento de orden $O(h^2)$ y usando la notación $w_j = x(t_j)$, junto con las condiciones de frontera $x(a)=\alpha$ y $x(b)=\beta$, se obtiene

$$\begin{aligned} w_0 &= \alpha\\ \frac{2w_i-w_{i+1}-w_{i-1}}{h^2}+p(t_i)\,\frac{w_{i+1}-w_{i-1}}{2h}+q(t_i)w_i &= -r(t_i)\\ w_{N+1} &= \beta\end{aligned}$$

$i=1,2,\ldots,N$. Esto se escribe como

$$-\left(1+\frac{h}{2}p(t_i)\right)w_{i-1}+(2+h^2q(t_i))w_i-\left(1-\frac{h}{2}p(t_i)\right)w_{i+1}=-h^2r(t_i)$$

y el sistema de ecuaciones lineales de orden $N\times N$ tiene la forma tridiagonal

$$Aw=b$$

donde

$$A=\begin{pmatrix} 2+h^2q(t_1) & -1+\frac{h}{2}p(t_1) & 0 & \cdots & 0 & 0\\ -1-\frac{h}{2}p(t_2) & 2+h^2q(t_2) & -1+\frac{h}{2}p(t_2) & \cdots & 0 & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots\\ & & & & & -1+\frac{h}{2}p(t_{N-1})\\ 0 & \cdots & 0 & 0 & -1-\frac{h}{2}p(t_N)) & 2+h^2q(t_N)\end{pmatrix}$$

y

$$w=\begin{pmatrix} w_1\\ w_2\\ \vdots\\ \vdots\\ w_N\end{pmatrix} \qquad \text{y} \qquad b=\begin{pmatrix} -h^2r(t_1)+\left(1+\frac{h}{2}p(t_1)\right)w_0\\ -h^2r(t_2)\\ \vdots\\ \vdots\\ -h^2t(t_{N-1})\\ -h^2r(t_N)+\left(1-\frac{h}{2}p(t_N)\right)w_{N+1}\end{pmatrix}.$$

El sistema tridiagonal anterior tiene solución única si $h < \frac{2}{L}$, donde $L = \max\{|p(t)| \ : \ a \leqslant t \leqslant b\}$.

9.10.1 Caso no lineal

Consideremos el problema no lineal de valores en la frontera

$$\left\{ \begin{array}{lcll} x'' & = & f(t,x,x')\,, & a \leqslant t \leqslant b \\ x(a) & = & \alpha & \\ x(b) & = & \beta\,. & \end{array} \right.$$

Para garantizar soluciones suponemos que

1.- f, $\dfrac{\partial f}{\partial x}$ y $\dfrac{\partial f}{\partial x'}$ son continuas en $D = \{(t,x,x') \ : \ a \leqslant t \leqslant b\,,\ -\infty < x < \infty\,,\ -\infty < x' < \infty\}$.

2.- $\dfrac{\partial f}{\partial x}(t,x,x') \geqslant \delta$ en D para algún $\delta > 0$.

3.- Existen

$$\begin{aligned} k &= \max\left\{ \left| \frac{\partial f}{\partial x}(t,x,x') \right| \ : \ (t,x,x') \in D \right\}, \\ L &= \max\left\{ \left| \frac{\partial f}{\partial x'}(t,x,x') \right| \ : \ (t,x,x') \in D \right\}. \end{aligned}$$

Como antes, usando fórmulas para $x''(t_i)$ y $x'(t_i)$, obtenemos

$$\frac{x(t_{i+1}) - 2x(t_i) + x(t_{i-1})}{h^2} = f\left(t_i, x(t_i), \frac{x(t_{i+1}) - x(t_{i-1})}{2h} - \frac{h^2}{6}x'''(\xi_i)\right) + \frac{h^2}{12}x^{(4)}(\theta_i)$$

donde $\xi_i, \theta_i \in]t_{i-1}, t_{i+1}[$.

Poniendo

$$\begin{aligned} w_0 &= \alpha \\ -\frac{w_{i+1} - 2w_i + w_{i-1}}{h^2} + f\left(t_i, w_i, \frac{w_{i+1} - w_{i-1}}{2h}\right) &= 0 \\ w_{N+1} &= \beta \end{aligned}$$

$i = 1, 2, \ldots, N$, obtenemos un sistema de ecuaciones no lineales

$$\begin{aligned} 2w_1 - w_2 + h^2 f(t_1, w_1, \frac{w_2 - \alpha}{2h}) - \alpha &= 0 \\ -w_1 + 2w_2 - w_3 + h^2 f(t_2, w_2, \frac{w_3 - w_1}{2h}) &= 0 \\ \vdots \quad \vdots \quad \vdots & \\ -w_{N-2} + 2w_{N-1} - w_N + h^2 f(t_{N-1}, w_{N-1}, \frac{w_N - w_{N-2}}{2h}) &= 0 \\ -w_{N-1} + 2w_N + h^2 f(t_N, w_N, \frac{\beta - w_{N-1}}{2h}) &= 0 \end{aligned}$$

y tiene solución única si y sólo si $h < 2/L$. Para resolver este sistema podemos usar, por ejemplo, el método de Newton en varias variables.

9.11 Problemas resueltos

Ejemplo 75 Usando los métodos de aproximación de soluciones de ecuaciones diferenciales de Euler, Euler Mejorado, y de Runge–Kutta, encuentre una solución aproximada del problema

$$\begin{cases} x' = \text{sen}(tx) \\ x(0) = 3 \end{cases}$$

con $h = 0.1$ y $n = 10$.

Solución. Usando las fórmulas de aproximación de Euler, Euler Mejorado, y de Runge–Kutta, tenemos la siguiente tabla para los valores de (t_n, x_n), los cuales unidos por segmentos de rectas nos dan la aproximación poligonal deseada.

		Euler	Euler Mejorado	Runge–Kutta
n	t_n	x_n	x_n	x_n
0	0	3	3	3
1	0,1	3	3,01494	3,01492
2	0,2	3,02955	3,05884	3,05874
3	0,3	3,08050	3,12859	3,12829
4	0,4	3,16642	3,21812	3,21744
5	0,5	3,26183	3,31761	3,31637
6	0,6	3,36164	3,41368	3,41185
7	0,7	3,45185	3,49163	3,48947
8	0,8	3,51819	3,53946	3,53746
9	0.9	3,55031	3,5514	3,55003
10	1	3,54495	3,52867	3,52803

9.12 Ejercicios

Problema 9.1 El método de Adams–Bashforth de segundo orden para aproximar soluciones de un problema de valor inicial

$$\left\{ \begin{array}{rcl} x' & = & f(t,x)\,, \quad a \leqslant t \leqslant b \\ x(a) & = & x_0 \end{array} \right.$$

es dado por

$$x_{n+1} = x_n + h\left(\frac{3}{2}f(t_n,x_n) - \frac{1}{2}f(t_{n-1},x_{n-1})\right)$$

$n \geqslant 1$, donde para $n = 1$, $t_0 = a$ y $t_1 = t_0 + h = a + h$ y x_1 puede calcularse con cualquier método de 1 paso, por ejemplo, Runge–Kutta, de de orden 4. Conocidos (t_0, x_0) y $t_1, x_1)$, calculamos (t_2, x_2), y así sucesivamente usando el método de Adams–Bashforth. Use lo anterior para aproximar la solución al problema de valor inicial

$$\left\{ \begin{array}{rcl} x' & = & -2tx + \dfrac{1}{10^2}\displaystyle\int_0^t x(s)ds \\ x(0) & = & 1\,, \end{array} \right.$$

donde para aproximar la integral del lado derecho se usa la regla de los trapecios.

Problema 9.2 Determine una solución aproximada del problema

$$-\frac{d^2x(t)}{dt^2} = -e^{-x(t)}\,, \quad t \in]0,1[\,, \quad x(0) = x(1) = 0$$

mediante el método de elementos finitos.

Problema 9.3 Considere la función

$$f(t) = \int_0^t \sqrt{1 - k\,\mathrm{sen}^2(\theta)}\,d\theta$$

donde k es un parámetro, con $0 \leqslant k \leqslant 1$. Usando un problema de valor inicial adecuado y el método de Runge–Kutta de orden 4, con una subdivisión del intervalo $[0, \pi/2]$ en 10 subintervalos, encuentre una aproximación al valor de la integral.

Problema 9.4 En un estudio de una población de ratones de campo se encuentra que ella evoluciona de acuerdo a la siguiente ecuación diferencial

$$\frac{dN(t)}{dt} = AN(t) - B(t)N(t)^{1.1}$$

$N(t)$ indica la cantidad de inviduos en el tiempo t. Se determina experimentalmente que $A = 2$ y se conocen las siguientes mediciones para $B(t)$ (t medido en alguna unidad razonable de tiempo)

t	0	4	8	12
B	0.70	0.04	0.56	0.70

(a) Usando interpolación de Lagrange, encuentre una aproximación polinomial para $B(t)$.

(b) Usando la aproximación polinomial para $B(t)$ encontrada en la parte (a), encuentre una aproximación en el intervalo $[0, 0.6]$ de la solución de la ecuación diferencial usando el método de Euler con $h = 0.2$ y dada la población inicial de ratones $N(0) = 100$.

Problema 9.5 Para resolver el siguiente problema con valor en la frontera:

$$\begin{aligned} x''(t) &= 4(x(t) - t) \quad , \ 0 \leqslant t \leqslant 1 \\ x(0) &= 0 \quad , \ x(1) = 2 \end{aligned}$$

se propone usar el método de diferencias finitas sobre la malla $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$, es decir, el tamaño de paso es $h = \frac{1}{4}$. Se les recuerda que las fórmulas de diferencias finitas centradas son:

$$\begin{aligned} x''(t_i) &= \frac{x(t_{i+1}) - 2x(t_i) + x(t_{i-1})}{h^2} \\ y'(t_i) &= \frac{x(t_{i+1}) - x(t_{i-1})}{2h} \quad , \ x(t_i) = \omega_i \qquad \text{donde } i = 1, 2, 3 \end{aligned}$$

(a) Escriba el sistema lineal que se obtiene al aplicar estas fórmulas en la ecuación diferencial con valor en la frontera dada anteriormente.

(b) Para resolver dicho sistema se propone usar el método iterativo de Gauss-Seidel, escriba cómo quedaría dicho método para el sistema obtenido en la parte (a) y diga si el método converge o no. Justifique su respuesta. Partiendo del punto $\omega^{(0)} = (\omega_1^{(0)}, \omega_2^{(0)}, \omega_3^{(0)}) = (0, 0, 0)$, obtenga los puntos $\omega^{(1)}$, $\omega^{(2)}$ y $\omega^{(3)}$ es decir, las tres primeras iteraciones del método de Gauss-Seidel.

(c) Tomando $\omega^{(3)}$ como la solución del sistema lineal obtenido en la parte (a), obtenga un polinomio de interpolación de grado 3 que aproxime la solución de la ecuación diferencial $x(t)$ en el intervalo $[\frac{1}{4}, \frac{1}{2}]$.

Índice Alfabético

Referencias

[1] Kendall Atkison, *Elementary numerical analysis*. John Willey & Sons, Inc. 1985.

[2] Richard L. Burden, J. Douglas Faires, *Análisis numérico*. (Séptima edición) International Thomson Editores, 2002.

[3] Steven C. Chapra, Raymond P. Canale, *Métodos numéricos para ingenieros*. McGraw–Hill, 1999.

[4] Lars Elden, Linde Wittmeyer–Koch, Hans Bruun Nielsen, *Introduction to numerical computation–analysis and matlab illustrations*. Studentlitteratur AB, 2004.

[5] Curtis F. Gerald, Patrick O. Wheatley, *Análisis numérico con aplicaciones*. (Sexta edición). Pearson Educación, 2000.

[6] Desmond J. Higham, Nicholas J. Higham, *Matlab guide*. SIAM, 2000.

[7] David Kincaid, Ward Cheney, *Análisis numéricos: las matemáticas del cálculo científico*. Addison–Wesley Iberoamericana, 1994.

[8] John Mathews, Kurtis D. Fink, *Métodos Numéricos con MatLab*. Prentice Hall, 2003.

[9] Clever B. Moler, *Numerical computing with matlab*. SIAM, 2004.

[10] Shoichiro Nakamura, *Análisis numérico y visualización con Matlab*. Pearson Educación, 1997.

[11] James M. Ortega, *Numerical analysis: a second course*. SIAM 1990.

[12] W. Allen Smith, *Análisis numérico*. Prentice–Hall Hispanoamericana S.A., 1988.

[13] J. Stoer, R. Burlirshc, *Introduction to numerical analysis*. (Third edition). Text in Applied Mathematics 12, Springer, 2002.

[14] G. W. Stewart, Afternotes on numerical analysis. SIAM, 1996.

[15] G. W. Stewart, *Afternotes goes to graduate school: lectures on advanced numerical analysis*. SIAM, 1998.