Universidad Nacional de Colombia
Departamento de Matemáticas

Análisis Numérico I

Tarea 1 (2024-II)

Beimar José Naranjo Morales

Eduards Alexis Méndez Chipatecua

Julio César Vásquez Arenas

1. Un bloque de masa m desliza bajo la acción de la gravedad por un plano inclinado formando un ángulo θ respecto de la horizontal. Se puede demostrar que, si la fuerza de rozamiento F_r entre el bloque y el plano viene dada por $F_r = -kmv^2$ con v la velocidad del bloque y k un coeficiente de rozamiento, entonces el tiempo T requerido para que el bloque recorra una distancia D partiendo del reposo está relacionado de la forma

$$e^{kD} = \cosh(T\sqrt{kg\sin\theta})$$

siendo $g = 9.8$ la aceleración de la gravedad. Si $\theta = \pi/4, k = 0.5 \pm 0.1$ y $T = 2 \pm 0.2$, hallar la precisión con la que se conoce D.

Solución

En primer lugar, despejando D obtenemos que:

$$D(k,t) = \frac{\ln(\cosh(T\sqrt{kg\sin\theta}))}{k}$$

Calculamos el error relativo para D se mediante series de Taylor, para esto encontramos las siguientes derivadas parciales:

$$\frac{\partial_k D(k,T)}{D(k,T)} = \frac{T\sqrt{gk\sin(\theta)}\tanh\left(T\sqrt{gk\sin(\theta)}\right)}{2k\log\left(\cosh\left(T\sqrt{gk\sin(\theta)}\right)\right)} - \frac{1}{k}$$

$$\frac{\partial_t D(k,T)}{D(k,T)} = \frac{\sqrt{gk\sin(\theta)}\tanh\left(T\sqrt{gk\sin(\theta)}\right)}{\log\left(\cosh\left(T\sqrt{gk\sin(\theta)}\right)\right)}$$

Así,

$$\frac{D(\hat{k},\hat{T})-D(k,T)}{D(k,T)} = \varepsilon_k \frac{\partial_k D(k,T)}{D(k,T)} + \varepsilon_T \frac{\partial_t D(k,T)}{D(k,T)} = 0.0454233 = 4.5\times 10^{-2}$$

Puesto que evaluando la función con ayuda de *Mathematica* obtenemos:

```
theta = Pi/4; g = 9.8; ek = 0.1; et = 0.2;
De[k_, t_] := Log[Cosh[t Sqrt[k g Sin[theta]]]]/k
ek D[De[k, t], k]/De[k, t] + et D[De[k, t], t]/De[k, t] /. {k -> 0.5, t -> 2}
0.0454233
```

2. Utilizando el método de redondeo.

   (a) Hallar el número de máquina mas próximo a 125.6 y a 126 si trabaja con

   - Base 10 y mantisa de 2 dígitos.
   - Base 2 y mantisa de 8 dígitos.

   Solución

   - 125.6 base 10 y mantisa de 2 dígitos:

$$\mathrm{fl}_{10}(12\overline{\mathbf{5}}.6) = 130$$

   - 126 base 10 y mantisa de 2 dígitos:

$$\mathrm{fl}_{10}(12\overline{\mathbf{6}}) = 130$$

   - 125.6 base 2 y mantisa de 8 dígitos:

Cociente	125	62	31	15	7	3	1
Residuo		1	0	1	1	1	1

Cociente	6	2	4	8	6
Parte Entera		1	0	0	1

$$125.6 = 111101.\overline{1001}_2$$

$$\mathrm{fl}_2(1111101.1\overline{\mathbf{0}}) = 1.1111011_2 \times 10_2^6 = 125.5$$

   - 126 base 2 y mantisa de 8 dígitos:

$$126 = 125 + 1 = 1111110_2$$

$$\mathrm{fl}_2(1111110.0\overline{\mathbf{0}}) = 1111110.0_2 = 126$$

(b) Verificar para $x = 125.6$ la cota para el error relativo

$$\left|\frac{x - \text{fl}(x)}{x}\right| \leq \epsilon$$

si $\epsilon = 1/2\beta^{1-d}$ donde β es la base y d la longitud de la mantisa.

Solución

- Base 10 y mantisa de 2 dígitos

$$\left|\frac{125.6 - \text{fl}_{10}(125.6)}{125.6}\right| = \left|\frac{125.6 - 130}{125.6}\right| < 3.5\times 10^{-2} < 5\times 10^{-2} = \frac{1}{20} = \frac{1}{2}10^{1-2}$$

- Base 2 y mantisa de 8 dígitos:

$$\left|\frac{125.6 - \text{fl}_{2}(125.6)}{125.6}\right| = \left|\frac{125.6 - 125.5}{125.6}\right| < 8 \times 10^{-4} < 3 \times 10^{-3} = \frac{1}{2^8} = \frac{1}{2}2^{1-8}$$

(c) ¿Cuál es, en cada caso el valor que da la máquina como resultado de las operaciones $126 + 126.5$ y $126 - 125.6$? ¿Cuál es el error relativo de estos resultados?

Solución

- $126 + 126.5$ en base 10 y mantisa de 2 dígitos:

$$\text{fl}_{10}(\text{fl}_{10}(126) + \text{fl}_{10}(126.5)) = \text{fl}_{10}(130 + 130) = \text{fl}_{10}(260) = 260$$

$$E_R = \left|\frac{251.6 - 260}{251.6}\right| \approx 3.3 \times 10^{-2}$$

- $126 - 125.6$ en base 10 y mantisa de 2 dígitos:

$$\text{fl}_{10}(\text{fl}_{10}(126) - \text{fl}_{10}(125.6)) = \text{fl}_{10}(130 - 130) = \text{fl}_{10}(0) = 0$$

$$E_R = \left|\frac{0.4 - 0}{0.4}\right| = 1$$

- $126 + 126.5$ en base 2 y mantisa de 8 dígitos: Primero calculamos 126.5 base 2 y mantisa de 8 dígitos:

Cociente	126	63	31	15	7	3	1
Residuo		0	1	1	1	1	1

Cociente	5	0
Parte Entera		1

$$126.5 = 1111110.1$$

$$\text{fl}_2(1111110.1_2) = 1.1111101 \times 10_2^6 = 126.5$$

$$\text{fl}_2(\text{fl}_2(126) + \text{fl}_2(126.5)) = \text{fl}_2(126 + 126.5) = \text{fl}_2(252.5)$$

Cociente	252	126	62	31	15	7	3	1
Residuo		0	0	0	1	1	1	1

$$\text{fl}_2(252.5) = \text{fl}_2(11111000.1_2) = 1.1111101 \times 10_2^7 = 253$$

$$E_R = \left|\frac{252.5 - 253}{252.5}\right| \approx 1.9 \times 10^{-3}$$

- $126 - 125.6$ en base 2 y mantisa de 8 dígitos:

$$\text{fl}_2(\text{fl}_2(126) - \text{fl}_2(125.6)) = \text{fl}_2(126 - 125.5) = \text{fl}_2(0.5) = 0.5$$

$$E_R = \left|\frac{0.4 - 0.5}{0.4}\right| = 2.5 \times 10^{-1}$$

3. Suponga que una compañía de computadores está desarrollando un nuevo sistema de punto flotante para usarlo con sus máquinas. Ellos necesitan su ayuda para responder unas preguntas acerca de su sistema. Siguiendo la teminología descrita en clase, el sistema de punto flotante de la compañía se especifica por (β, t, L, U). Usted debe suponer que:

   - Todos los valores de punto flotante son normalizados (excepto la representación de punto flotante de cero).
   - Todos los dígitos en la mantisa de un valor de punto flotante son almacenados explícitamente.
   - El cero se representa con una mantisa y exponente de ceros.

   Preguntas:

   (a) ¿Cuántos valores diferentes de punto flotante no negativos pueden representarse por medio de este sistema de punto flotante ?

   Solución

   En total hay $(U - L + 1)$ exponentes posibles, contando al 0. Para cada uno de estos exponentes el primer dígito de la mantisa debe ser diferente de 0, con lo cual tiene $\beta - 1$ opciones, y los demás dígitos de la mantisa tienen β posibilidades cada uno. Multiplicando estos valores obtenemos todos los valores normalizados posibles. Le agregamos 1 que es la representación del 0.

$$(U-L+1)(\beta-1)\beta^{t-1}+1$$

(b) La misma pregunta para el caso $(\beta,t,L,U)=(8,5,-100,100)$, el cual la compañia esta contemplando en particular.

Solución

$$(100-(-100)+1)(8-1)8^{5-1}+1=201\cdot 7\cdot 8^4+1=5.763.073$$

(c) ¿ Cuál es el valor aproximado (en base 10) del número más grande y el número positivo más pequeño que pueden ser representado en este sistema de punto flotante?

Solución

El más grande se alcanza con el exponente U y la mantisa con todos sus dígitos en $\beta-1$, esto es

$$\beta^U\sum_{k=0}^{t-1}\frac{\beta-1}{\beta^k}=\beta^U(\beta-1)\left(\frac{1-\beta^{-t}}{1-\beta^{-1}}\right)=\beta^U(\beta-1)\left(\frac{\frac{\beta^t-1}{\beta^t}}{\frac{\beta-1}{\beta}}\right)=\beta^{U+1}\left(1-\frac{1}{\beta^t}\right).$$

Reemplazando, $\beta=10$, el número más grande es $10^{U+1}\left(1-\frac{1}{10^t}\right)$. Por otro lado, el valor normalizado más pequeño se alcanza con el exponente L y la mantisa con el primer dígito en 1 y el resto en 0, esto es $\beta^L=10^L$.

(d) Para el caso general, demuestre que los errores relativos producidos por realizar truncamieto y redondeo son

$$\left|\frac{fl(x)-x}{x}\right|=\begin{cases}\beta^{1-t} & \text{cuando se efectúa truncamiento}\\ \frac{1}{2}\beta^{1-t} & \text{cuando se efectúa redondeo}\end{cases}$$

Solución

- Truncamiento

  Supongamos que $x>0$ y calculamos $x-\mathrm{fl}(x)$, entonces

$$\begin{aligned}|x-\mathrm{fl}(x)|&=\left|\beta^e\sum_{k=0}^{\infty}\frac{d_k}{\beta^k}-\beta^e\sum_{k=0}^{t-1}\frac{d_k}{\beta^k}\right|=\beta^e\left|\sum_{k=t}^{\infty}\frac{d_k}{\beta^k}\right|=\beta^{e-t}\left|\sum_{k=0}^{\infty}\frac{d_{k+t}}{\beta^k}\right|\\&\le\beta^{e-t}\sum_{k=0}^{\infty}\frac{\beta-1}{\beta^k}=\beta^{e-t}(\beta-1)\frac{\beta}{\beta-1}=\beta^{e-t+1}.\end{aligned}$$

  Por otra parte, tenemos que

$$x=(d_0.d_1d_2\dots d_{t-1}d_td_{t+1}\dots)\,\beta^e\ge d_0\beta^e\ge\beta^e$$

Luego combinando las desigualdades obtenemos

$$\frac{|fl(x)-x|}{|x|} \leq \frac{\beta^{e-t+1}}{\beta^e} \leq \beta^{1-t}$$

- Redondeo

  Supongamos que $x > 0$ con $d_t \geq \beta/2$ y calculamos $\mathrm{fl}(x) - x$, entonces

$$\begin{aligned}
|\mathrm{fl}(x)-x| &= \left| \left(\beta^e \sum_{k=0}^{t-1} \frac{d_k}{\beta^k} + \beta^e \frac{1}{\beta^{t-1}} \right) - \beta^e \sum_{k=0}^{\infty} \frac{d_k}{\beta^k} \right| \\
&= \beta^e \left| \frac{1}{\beta^{t-1}} - \sum_{k=t}^{\infty} \frac{d_k}{\beta^k} \right| = \beta^e \left(\frac{\beta - d_t}{\beta^t} - \sum_{k=t+1}^{\infty} \frac{d_k}{\beta^k} \right) \\
&\leq \beta^e \frac{\beta - d_t}{\beta^t} \leq \beta^e \frac{1}{2\beta^{t-1}}
\end{aligned}$$

El valor absoluto se puede eliminar puesto que

$$\sum_{k=t}^{\infty} \frac{d_k}{\beta^k} \leq \beta^{-t}(\beta-1) \sum_{k=0}^{\infty} \frac{1}{\beta^k} = \frac{1}{\beta^{t-1}}$$

Por otra parte, tenemos que

$$x = (d_0.d_1 d_2 \ldots d_{t-1} d_t d_{t+1} \ldots)\, \beta^e \geq d_0 \beta^e \geq \beta^e$$

Luego, combinando las desigualdades obtenemos

$$\frac{|fl(x)-x|}{|x|} \leq \frac{1}{2} \beta^{1-t}$$

4. Considere la expansión de Taylor para la función exponencial

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = \sum_{i=0}^{\infty} \frac{x^i}{i!} = \lim_{N \to \infty} S(x, N)$$

donde $S(x, N)$ es la suma parcial con $N + 1$ términos.

(a) Escriba un programa que grafique el error relativo de la suma, $|S(x, N) - e^x| / e^x$ versus N (hasta N $= 60$) para un valor dado de x. Pruebe su programa para $x = 10, 2, -2 \text{y} -10$. De las gráficas, explique por qué ésta no es una buena manera para evaluar e^x cuando $x < 0$.

Solución

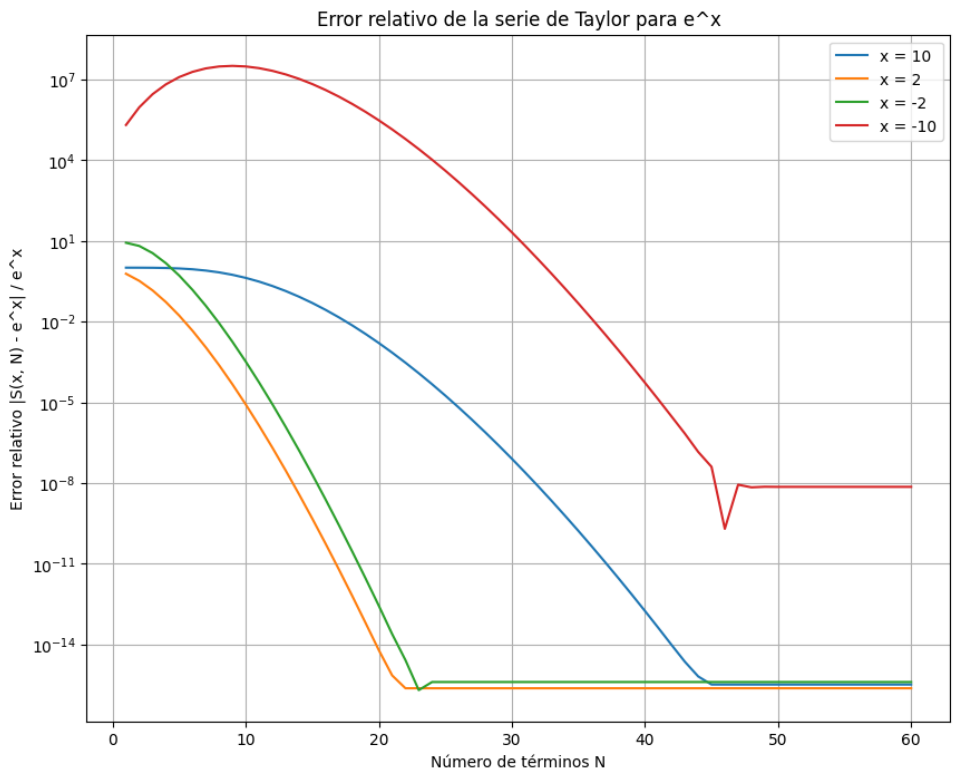

Figure 1: Aproximación con series de taylor hasta un $N = 60$

(b) Modifique su programa tal que use la identidad $e^x = 1/e^{-x} = 1/S(-x, \infty)$ para evaluar la función exponencial cuando x es negativa. Explique por qué esta técnica funciona mejor.

Solución

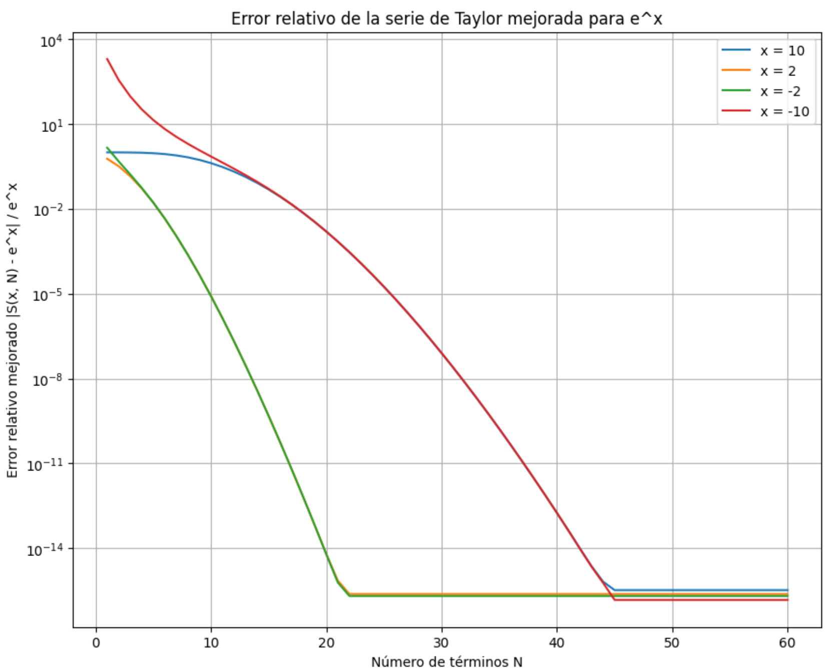

Figure 2: Aproximación teniendo en cuenta la expresión $\frac{1}{e^{-x}}$ para x<0

N	Serie Directa (Suma para e^{-10})	Usando Identidad (1 / Suma para e^{10})	Error Relativo Directo	Error Relativo Identidad
1	-9.0000000000	0.0909090909	1.9823919215e+05	2.0014059813e+03
2	41.0000000000	0.0163934426	9.0308409759e+05	3.6008960319e+02
3	-125.6666666667	0.0043923865	2.7679935349e+06	9.5748751661e+01
4	291.0000000000	0.0015519917	6.4097005463e+06	3.3184892594e+01
5	-542.3333333333	0.0006767426	1.1945687616e+07	1.3906248000e+01
6	846.5555555556	0.0003488507	1.8646625988e+07	6.6839486861e+00
7	-1137.5714285714	0.0002061566	2.5056679161e+07	3.5409002990e+00
8	1342.5873015873	0.0001364100	2.9572452275e+07	2.0046300259e+00
9	-1413.1446208113	0.0000991417	3.1126582653e+07	1.1837412345e+00
10	1342.5873015873	0.0000778676	2.9572452275e+07	7.1514892367e-01

Figure 3: Resultados de la serie para un N especifico

Notebook: si desea ver el notebook de este ejercicio, puede ingresar en: `https://github.com/Things-I-learn/Analisis-Numerico-2024/blob/main/Exponencial_error.ipynb`

¿Por qué esta técnica funciona mejor?

Desgolemos cada operación:

Serie de Taylor Directa para (e^{-10})

Para calcular e^{-10} directamente, la serie de Taylor se escribe como:

$$e^{-10} = 1 - \frac{10^1}{1!} + \frac{10^2}{2!} - \frac{10^3}{3!} + \frac{10^4}{4!} - \cdots$$

Ahora calculemos la serie de Taylor usando la identidad $e^{-10} = \frac{1}{e^{10}}$.

Para mejorar la precisión, podemos calcular e^{-10} usando la identidad:

$$e^{-10} = \frac{1}{e^{10}}$$

Luego, calculamos e^{10} usando la serie de Taylor:

$$e^{10} = 1 + \frac{10^1}{1!} + \frac{10^2}{2!} + \frac{10^3}{3!} + \frac{10^4}{4!} + \cdots$$

Finalmente, al calcular e^{-10} usando la identidad, obtenemos:

$$e^{-10} \approx \frac{1}{1 + 10 + 50 + 166.67 + 416.67 + 833.33 + \cdots}$$

Con esto queremos ilustrar que al tener que alternar de signo, tenemos una cancelación que se va acumulando, el cual hace que nuestra convergencia sea mucho mas lenta como lo muestran las gráficas.

5. (a) De manera similar a la desarrollada en clase, deduzca que una aproximación de $f'(x_0)$ es

$$\frac{f(x_0 + h) - f(x_0 - h)}{2h}$$

Muestre que esta aproximación tiene un error de $\mathcal{O}(h^2)$. Más precisamente, el primer término del error es $-\frac{h^2}{6} f'''(x_0)$ cuando $f'''(x_0) \neq 0$.

Solución Usando el teorema de Taylor,

$$f(x_0 - h) = f(x_0) - hf'(x_0) + \frac{h^2}{2} f''(x_0) - \frac{h^3}{6} f'''(x_0) + O(h^4),$$

y

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \frac{h^3}{6} f'''(x_0) + O(h^4).$$

De modo que,

$$f(x_0+h)-f(x_0-h)=2hf'(x_0)+2\frac{h^3}{6}f'''(x_0)+O(h^5),$$

despejando $f'(x_0)$,

$$f'(x_0)=\frac{f(x_0+h)-f(x_0-h)}{2h}-\frac{h^2}{6}f'''(x_0)+O(h^4).$$

(b) Adapte el programa de Matlab hecho en clase para visualizar el comportamiento del error de aproximación a medida que el paso h decrece desde $h=10^{-1},\ldots,10^{-16}$.

Solución

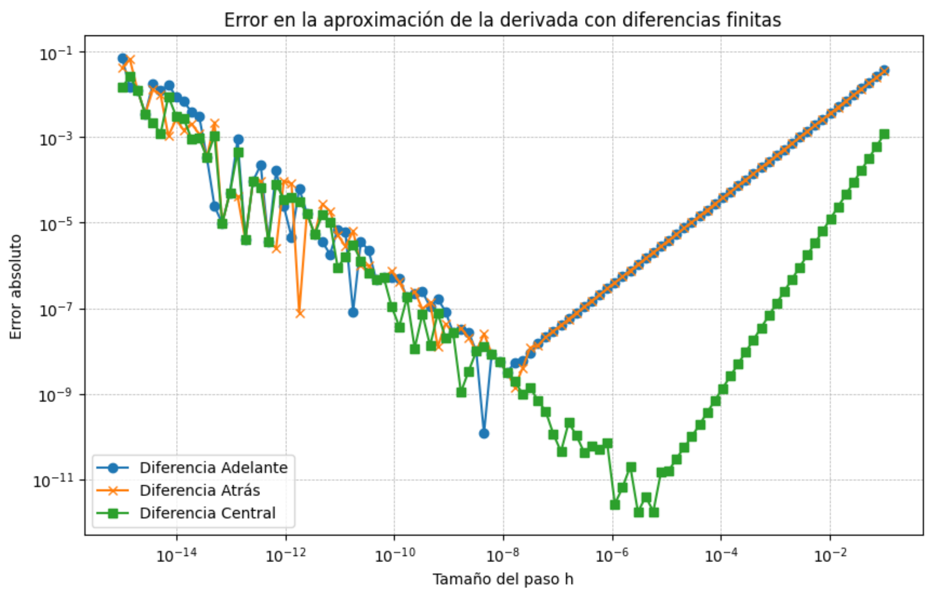

Figure 4: Error en la aproximación de la derivada

Notebook: si desea ver el notebook de este ejercicio a y b, puede ingresar en: `https://github.com/Things-I-learn/Analisis-Numerico-2024/blob/main/derivada.ipynb`

(c) Muestre que

$$f'\left(x_0\right)=\frac{-f\left(x_0+2h\right)+8f\left(x_0+h\right)-8f\left(x_0-h\right)+f\left(x_0-2h\right)}{12h}+\mathcal{O}\left(h^4\right)$$

De nuevo adapte su programa para visualizar el comportamiento del error de aproximación a medida que el paso h decrece desde $h=10^{-1},\ldots,10^{-16}$.

Solución

Usando el teorema de Taylor,

$$f(x_0 - 2h) = f(x_0) - 2hf'(x_0) + \frac{4h^2}{2}f''(x_0) - \frac{8h^3}{6}f'''(x_0) + O(h^4),$$

$$f(x_0 + 2h) = f(x_0) + 2hf'(x_0) + \frac{4h^2}{2}f''(x_0) + \frac{8h^3}{6}f'''(x_0) + O(h^4).$$

Operando,

$$\begin{aligned} -f\left(x_0+2h\right)+f\left(x_0-2h\right) &= -4hf'(x_0) - \frac{16h^3}{6}f'''(x_0) + O(h^5) \\ 8f\left(x_0+h\right)-8f\left(x_0-h\right) &= 16hf'(x_0) + 16\frac{h^3}{6}f'''(x_0) + O(h^5) \end{aligned}$$

Así,

$$12hf'\left(x_0\right) = -f\left(x_0+2h\right)+8f\left(x_0+h\right)-8f\left(x_0-h\right)+f\left(x_0-2h\right)+O(h^5).$$

Despejando $f'(x_0)$ se obtiene la igualdad requerida.

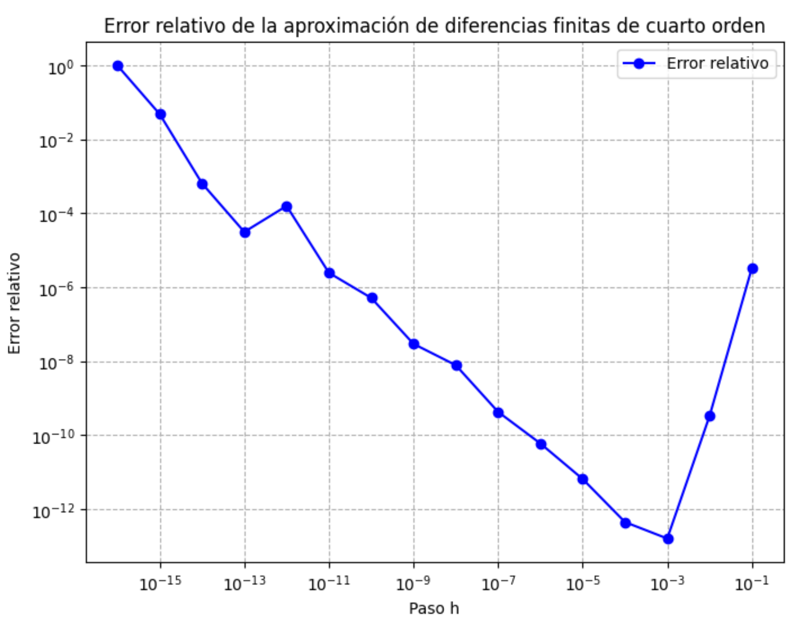

Figure 5: Error relativo usando la expresión: $f'\left(x_0\right) = \frac{-f(x_0+2h)+8f(x_0+h)-8f(x_0-h)+f(x_0-2h)}{12h} + \mathcal{O}\left(h^4\right)$

Notebook: si desea ver el notebook de este ejercicio c, puede ingresar en: `hhttps://github.com/Things-I-learn/Analisis-Numerico-2024/blob/main/Exponencial_error.ipynb` **Nota: este ejercicio esta al final del notebook.**

6. Repaso O mayúscula y o minúscula de una función. Si $f(x) = O(x^2)\,, g(x) = O(x^3)$ y $h(x) = O(x^3)$ cuando $x \to 0$. ¿Cuáles de las siguientes afirmaciones son verdaderas?. Justifique sus respuestas

(a) $f(x) = o(x)$ cuando $x \to 0$.
Verdadero. Por definición, $|f(x)| \leq kx^2$ para alguna constante k cuando $x \to 0$. Por tanto,
$$0 = \lim_{x\to 0} \frac{-kx^2}{x} \leq \lim_{x\to 0} \frac{f(x)}{x} \leq \lim_{x\to 0} \frac{kx^2}{x} = 0.$$
Por el teorema del sandwich queda demostrado.

(b) $f(x) = o(x^2)$ cuando $x \to 0$.
Falso. Si $f(x) = x^2 + x^3$, $f(x) = O(x^2)$ cuando $x \to 0$, y
$$\lim_{x\to 0} \frac{f(x)}{x^2} = 1.$$

(c) $f(x) \cdot g(x) = O(x^5)$ cuando $x \to 0$.
Verdadero. Tenemos que $|f(x)| \leq k_1 x^2$ y $|g(x)| \leq k_2 x^3$ para algunas constantes k_1 y k_2 cuando $x \to 0$, entonces $|f(x) \cdot g(x)| \leq k_1 k_2 x^5$.

(d) $f(x) + g(x) = O(x^3)$ cuando $x \to 0$.
Falso. Si $f(x) = x^2 - x^3$ y $g(x) = x^3$, $f(x) + g(x) = x^2$. De modo que para cualquier constante $k > 0$, si $0 < x < 1/k$, se tiene que $1 > kx$ y así $|f(x) + g(x)| = x^2 > kx^3$.

(e) $g(x) - h(x) = 0$.
Falso. Si $g(x) = x^3 + x^4$ y $h(x) = x^3$, $g(x) - h(x) = x^4 \neq 0$.

(f) $g(x)/h(x) = O(1)$ cuando $x \to 0$.
Falso. Si $g(x) = x^3$ y $h(x) = x^4$, entonces $g(x)/h(x) = 1/x$, pero la función $1/x$ tiende a infinito cuando $x \to 0$, y por lo tanto, $1/x$ no puede ser acotada por una constante.