Métodos Numéricos

Errores de Redondeo y Truncamiento

Daniel Barragán [1]

[1]Escuela de Ingeniería de Sistemas y Computación
Universidad del Valle

February 13, 2015

Agenda

Errores.

Exactitud y Precisión.

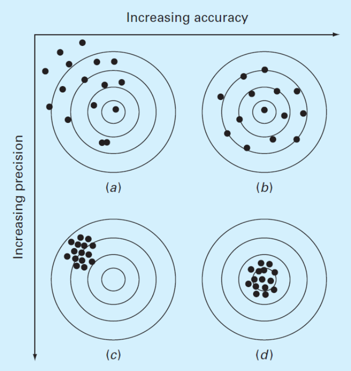
- El computador tiene un límite para representar magnitudes
- Al aproximar la derivada como una diferencia finita, la solución no es exacta. ¿Es posible cuantificar el error?
- Los métodos numéricos deben ser lo suficientemente precisos para solucionar un problema en particular

Errores.
Exactitud y Precisión.

- **Error de Redondeo:** Esta relacionado con el error que ocurre debido a que el computador no puede representar algunas cantidades exactamente
- **Error de Truncamiento:** Esta relacionado con la cantidad de términos que se emplean en una serie para aproximar una solución

Errores.
Exactitud y Precisión.

- Exactitud:accuracy, Precisión:precision



Errores.
Exactitud y Precisión.

- **Error Absoluto:** Diferencia entre el valor exacto y una aproximación

$$E_t = valorverdadero - aproximacion$$

- **Error Relativo:**

$$\varepsilon_t = \frac{valorverdadero - aproximacion}{valorverdadero} \times 100$$

Errores.

Exactitud y Precisión.

- **Problema:** Se mide la longitud de un puente y un tornillo, obteniendo las medidas de 9999 y 9 cm respectivamente, las medidas verdaderas son de 10000 cm y 10 cm, el error en ambos casos es de 1 cm, ¿Cúal es el valor de los errores relativos?

Errores.
Exactitud y Precisión.

- **Solución:**

$$\varepsilon_t = \frac{valorverdadero - aproximacion}{valorverdadero} \times 100$$

$$\varepsilon_t = \frac{10000 - 9999}{10000} \times 100 = 0,01\%$$

$$\varepsilon_t = \frac{10 - 9}{10} \times 100 = 10\%$$

Errores.
Exactitud y Precisión.

- En los problemas donde no hay solución analítica, se estima el error aproximado

$$\varepsilon_a = \frac{aproximacionactual - aproximacionanterior}{aproximacionactual} \times 100$$

- El criterio de parada consiste en encontrar un error de aproximación dentro de un rango ε_s

$$|\varepsilon_a| < \varepsilon_s$$

Errores.
Exactitud y Precisión.

- ¿Qué valor es adecuado para ε_s?

$$\varepsilon_s = (0.5x10^{2-n})\%$$

Donde:
n es la cantidad de cifras significativas

Errores.
Series de Taylor.

- Representación de una función por medio de una suma infinita de términos, los cuales son calculados a partir de los valores de las derivadas de la función en un punto **a**

$$\sum_{n=0}^{\infty} \frac{f^n(a)}{n!}(x-a)^n$$

Cuando $a = 0$, la serie es llamada serie de Maclaurin

Errores.

Serie de Maclaurin.

- **Problema:** Desarrolle la expansion en Series de Maclaurin para e^x hasta el cuarto término y exprese la expansión como una sumatoria

Errores.
Serie de Maclaurin.

- **Solucion:**

$$e^x = \frac{e^0}{0!}x^0 + \frac{e^0}{1!}x^1 + \frac{e^0}{2!}x^2 + \frac{e^0}{3!}x^3 \ldots$$

Expresando la expansion como una sumatoria se tiene:

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Errores.

Serie de Maclaurin.

- **Problema:** Estimar e^x con $x = 0.5$ empleando la expansión en series de Maclaurin para e^x con al menos 3 cifras significativas

$$e_x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + ... + \frac{x^n}{n!}$$

- **Nota**: La solución exacta es $e^{0.5} = 1.648721$.

Errores.

Serie de Maclaurin.

- **Solución:**

$$\varepsilon_s = (0.5x10^{2-n})\%$$
$$\varepsilon_s = (0.5x10^{2-3}) = 0.05\%$$

Se debe llegar a un error menor del 0.05 % para garantizar una correcta aproximación con al menos 3 cifras significativas

Errores.

Serie de Maclaurin.

- El primer estimado es $e^x = 1$
- El error absoluto para el primer estimado es:

$$\varepsilon_t = \frac{1.648721 - 1}{1.648721} \times 100 = 39.3\%$$

Errores.
Serie de Maclaurin.

- El segundo estimado es $e^x = 1 + x = 1 + 0.5 = 1.5$
- El error absoluto y el error aproximado para el segundo estimado es:

$$\varepsilon_t = \frac{1.648721 - 1.5}{1.648721} \times 100 = 9.02\%$$
$$\varepsilon_a = \frac{1.5 - 1}{1.5} \times 100 = 33.3\%$$

Errores.

Serie de Maclaurin.

- El tercer estimado es $e^x = 1 + x + \frac{x^2}{2} = 1 + 0.5 + 0.125 = 1.625$
- El error absoluto y el error aproximado para el tercer estimado es:

$$\varepsilon_t = \frac{1.648721 - 1.625}{1.648721} \times 100 = 1.44\%$$

$$\varepsilon_a = \frac{1.625 - 1.5}{1.625} \times 100 = 7.69\%$$

Errores.
Serie de Maclaurin.

- **Instrucciones Scilab:**
  - funcionexp.sce

Errores.

Serie de Maclaurin.

Términos	Aproximación	ε_t	ε_a
1	1	39.3	
2	1.5	9.02	33.3
3	1.625	1.44	7.69
4	1.645833333	0.175	1.27
5	1.648437500	0.0172	0.158
6	1.648697917	0.00142	0.0158

Errores de Redondeo.

Definición.

- **Error de Redondeo**: Resultan de la limitación que tienen los computadores para representar magnitudes

Errores de Redondeo.
Conversión Decimal a Binario.

- Representar el número decimal $(173)_{10}$ en formato binario

$$
\begin{gathered}
173/2 = 86 \text{ resto} = 1 \\
86/2 = 43 \text{ resto} = 0 \\
43/2 = 21 \text{ resto} = 1 \\
21/2 = 10 \text{ resto } 1 \\
10/2 = 5 \text{ resto } 0 \\
5/2 = 2 \text{ resto } 1 \\
2/2 = 1 \text{ resto } 0
\end{gathered}
$$

$(10101101)_2$

Errores de Redondeo.

Conversión Binario a Decimal.

- Representar el número binario $(10101101)_2$ en formato decimal

$$\begin{aligned}(10101101)_2\\ 2^7+2^5+2^3+2^2+2^0\\ 128+32+8+4+1\\ (173)_{10}\end{aligned}$$

Errores de Redondeo.
Representación del Computador.

- ¿Cuántos bits son necesarios para representar el número binario $(10101101)_2$
- ¿Como se representa el número binario $(10101101)_2$ en un computador de 16 bits?

Errores de Redondeo.

Representación Signo-Magnitud.

- Representación en Signo-Magnitud de $(10101101)_2$ en un computador de 16 bits

Signo	Magnitud														
0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	1

Errores de Redondeo.

Representación Signo-Magnitud.

- Mayor número positivo que se puede representar con 16 bits

Signo	Magnitud														
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

$$2^{15} - 1$$

Errores de Redondeo.

Representación Signo-Magnitud.

- Menor número negativo que se puede representar con 16 bits

Signo	Magnitud														
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$$-2^{15}$$

Errores de Redondeo.
Representación Punto Flotante.

- Formato de representación de punto flotante

$$\pm s \times b^e$$

Donde:
s es la mantisa
b es la base del sistema numérico empleado
e es el exponente

Errores de Redondeo.

Representación Punto Flotante.

- **Ejemplo:** El número 0.005678 se representa como 5.678×10^{-3} (forma normalizada)
- **Nota:** Se evita el almacenamiento en memoria de ceros no significativos como en el caso de 0.005678×10^{0}

Errores de Redondeo.
Representación Punto Flotante.

- **Problema:** Expresar en forma normalizada:
  - 0.0000000000000395

Errores de Redondeo.
Representación Punto Flotante.

- **Solución:**
  - $0.0000000000000395 = 3.95 \times 10^{-14}$

Errores de Redondeo.

Representación Punto Flotante.

- **Problema:** Suponga que existe un computador que opera en base 10 y puede representar cantidades con 5 dígitos. Se emplea: un dígito para el signo, dos para la mantisa, uno para el signo del exponente y uno para el exponente.
    - ¿Cuál es el valor positivo más grande que se puede representar?
    - ¿Cuál es el valor positivo más pequeño que se puede representar?
    - ¿Qué trae más ventajas al sistema, aumentar un dígito a la mantisa ó un dígito al exponente?

Errores de Redondeo.
Representación Punto Flotante.

Solución:
Representación: $s_1 d_1 . d2 \times 10^{s_0 d_0}$
Número positivo más grande: $+9.9 \times 10^{+9}$
Número positivo más pequeño: $+1.0 \times 10^{-9}$

Al aumentar un dígito a la mantisa, aumenta la precisión: $+9.99 \times 10^{+9}$
Al aumentar un dígito al exponente, aumenta el rango: $+9.9 \times 10^{+99}$

Errores de Redondeo.

Representación Punto Flotante.

- **Problema:** ¿Cuál es el error de redondeo (error relativo) al expresar el número 0.03125 en el sistema con representación $s_1 d_1.d2 \times 10^{s_0 d_0}$?

Errores de Redondeo.
Representación Punto Flotante.

- **Solución:**
  El número 0.03125 se representa como:
  $+3.1 \times 10^{-2}$

  El error de redondeo (error relativo) corresponde a:

$$\frac{0.03125 - 0.031}{0.03125} \times 100 = 0.8\%$$

Errores de Redondeo.
Representación Punto Flotante.

- Representación de números binarios en punto flotante

$$\pm(1 + f) \times 2^{e}$$

Donde:
f es la mantisa (parte fraccionaria, el 1 no se almacena)

Errores de Redondeo.

Representación Punto Flotante.

- **Problema:** Expresar en formato de punto flotante el número binario 1101.1

Errores de Redondeo.
Representación Punto Flotante.

- **Solución:**

$$1.1011 \times 2^3$$
$$(1 + 0.1011) \times 2^3$$

Errores de Redondeo.

Representación Punto Flotante IEEE-754.

- Representación Punto Flotante IEEE-754

sign (s)	biased exponent (e')	mantissa (m)
1 bit	8 bit	23 bits

$$value = (-1)^s \times (1.m)_2 \times 2^{e'-127}$$

Errores de Redondeo.

Representación Punto Flotante IEEE-754.

- **Problema:** Encontrar el valor decimal que corresponde al siguiente número binario en formato IEEE-754

1	10100010	10100000000000000000000

Errores de Redondeo.

Representación Punto Flotante IEEE-754.

- **Solución:**

$$value = (-1)^s \times (1.m)_2 \times 2^{e'-127}$$

$$value = (-1)^1 \times (1.101)_2 \times 2^{(10100010)_2-127}$$

$$value = (-1)^1 \times (1.625) \times 2^{162-127}$$

$$value = (-1)^1 \times (1.625) \times 2^{35}$$

$$value = -5.5834 \times 10^{10}$$

Errores de Redondeo.

Representación Punto Flotante IEEE-754.

- ¿Cómo es posible pasar de un valor en formato decimal a un valor en formato IEEE-754?

$$-5.5834 \times 10^{10} = (-1)^{1} \times (1.?)_2 \times 2^{\pm ?}$$

Tarea Opcional: Desarrolle un programa en scilab que permita realizar la conversión

Errores de Redondeo.

Representación Punto Flotante IEEE-754.

- Biased Exponent: Es una forma que permite representar cantidades negativas sin tener un bit de signo en el exponente

$$\text{Rango posible con 8bits}: 0 \leq e' \leq 255$$
$$\text{Bias}: -127 \leq e \leq 128$$

$$\text{Rango actual}: 1 \leq e' \leq 254$$
$$\text{Bias}: -126 \leq e \leq 127$$

Errores de Redondeo.
Representación Punto Flotante IEEE-754.

s	e'	m	Representa
0	Todos Ceros	Todos Ceros	0
1	Todos Ceros	Todos Ceros	-0
0	Todos Unos	Todos Ceros	∞
1	Todos Unos	Todos Ceros	$-\infty$
1 ó 0	Todos Unos	Diferente a Cero	NaN

La edición 1984 de IEEE-754 introdujo los números especiales.

Errores de Redondeo.
Representación Punto Flotante IEEE-754.

- **Problema:**
  - ¿Cuál es el valor positivo más grande que se puede representar con el formato IEEE-754?
  - ¿Cuál es el valor positivo más pequeño que se puede representar con el formato IEEE-754?
  - ¿Cuál es el valor del machine epsilon?

Errores de Redondeo.

Representación Punto Flotante IEEE-754.

Solución:

Número positivo más grande:

$+1.1...1 \times 2^{+127} = 3.4 \times 10^{38}$

Número positivo más pequeño:

$+1.0...0 \times 2^{-126} = 1.175 \times 10^{-38}$

Machine epsilon:

$\varepsilon_{mach} = 2^{-23} = 1.19 \times 10^{-7}$

Errores de Redondeo.

Operaciones Aritméticas.

- Son operaciones que ocasionan errores de redondeo:
  - Restar cantidades pequeñas muy cercanas (*subtractive cancellation*)

$$0.7642 \times 10^3 - 0.7641 \times 10^3 = 0.0001 \times 10^3$$

  - Adicionar un número grande a uno pequeño

$$0.4000 \times 10^4 - 0.0000001 \times 10^4 = 0.4000001 \times 10^4$$

Errores de Truncamiento.

Definición.

- **Error de Truncamiento**: Resultan de usar una aproximación en lugar de una solución matemática exacta

Errores de Truncamiento.

Definición.

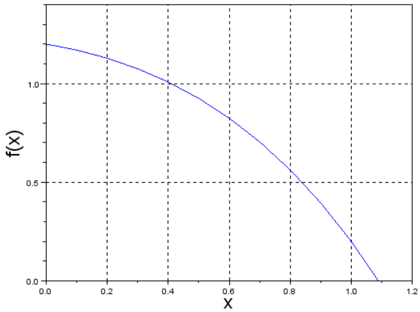
- En la siguiente ecuación se introduce un error de truncamiento, ya se que solo se esta aproximando el valor verdadero de la derivada

$$\frac{dv}{dt} \cong \frac{\Delta v}{\Delta t} = \frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i}$$

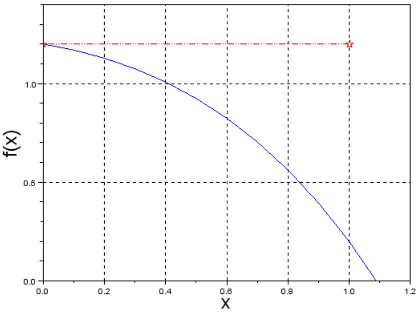
Errores de Truncamiento.

Series de Taylor.

- ¿Cómo estimar el valor de la función en el punto $x_{i+1}(x = 1)$ partiendo de $x_i(x = 0)$?

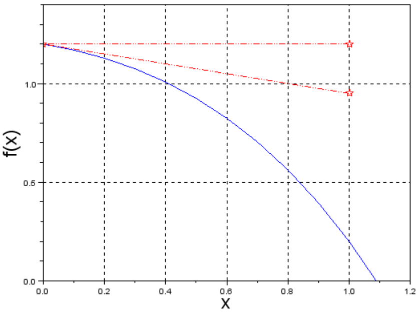


Errores de Truncamiento.
Series de Taylor.



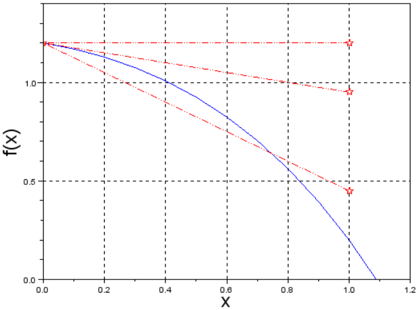
$$f(x_{i+1}) \cong f(x_i)$$

Errores de Truncamiento.
Series de Taylor.



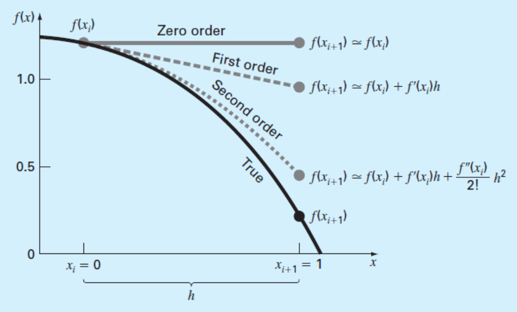
$$f(x_{i+1}) \cong f(x_i) + f'(x_i)h$$

Errores de Truncamiento.
Series de Taylor.



$$f(x_{i+1}) \cong f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2$$

Errores de Truncamiento.
Series de Taylor.



Aproximación de $f(x) = -0.1x^4 - 0.15x^3 - 0.5x^2 - 0.25x + 1.2$ en $x = 1$ en expansión de Series de Taylor

Errores de Truncamiento.

Series de Taylor.

- Expansión completa de los términos de la Serie de Taylor

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f^3(x_i)}{3!}h^3 + ... + \frac{f^n(x_i)}{n!}h^n + R_n$$

$$R_n = \frac{f^{n+1}(\xi)}{(n+1)!}h^{n+1}$$

Donde:
ξ es un valor de x que esta entre x_i y x_{i+1}

Errores de Truncamiento.
Series de Taylor.

- **Problema:** Utilice la expansión en series de Taylor hasta el orden 3, para predecir $f(3)$ en:

$$f(x) = 25x^3 - 6x^2 + 7x - 88$$

Empleando como punto base $x = 1$. Encuentre el error relativo para cada aproximación

Errores de Truncamiento.
Orden del Error.

- Orden del error de truncamiento

$$R_n = \frac{f^{n+1}(\xi)}{(n+1)!} h^{n+1}$$

Se puede expresar como:

$$R_n = O(h^{n+1})$$

Proporciona una idea del error en relación a la cantidad de términos que se emplean para la estimación

Errores de Truncamiento.

Residuo.

- Al emplear una cantidad finita de términos de la Serie de Taylor para estimar una función, una parte infinita de términos es truncada

Errores de Truncamiento.
Residuo.

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f^3(x_i)}{3!}h^3 + ... + \frac{f^n(x_i)}{n!}h^n + R_n$$

- Por ejemplo, al truncar la expansion en Series de Taylor después del término de orden cero, se tiene:

$$f(x_{i+1}) \cong f(x_i)$$

$$R_0 = f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f^3(x_i)}{3!}h^3 + ...$$

Errores de Truncamiento.
Residuo.

$$R_0 = f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f^3(x_i)}{3!}h^3 + \ldots$$

- Puesto que los términos de menor orden son más significativos para la solución final, se suele truncar el residuo

$$R_0 \cong f'(x_i)h$$

Errores de Truncamiento.

Estimación.

- Las Series de Taylor son útiles para calcular los errores de truncamiento

Errores de Truncamiento.

Estimación.

- En el ejemplo de caida libre $v(t)$ puede expandirse en una Serie de Taylor

$$v(t_{i+1}) \cong v(t_i) + v'(t_i)(t_{i+1} - t_i) + \frac{v''(t_i)}{2!}(t_{i+1} - t_i)^2 + ... + R_n$$

Truncando después del término de la primera derivada y despejando $v'(t_i)$:

$$v(t_{i+1}) \cong v(t_i) + v'(t_i)(t_{i+1} - t_i) + R_1$$

$$v'(t_i) = \frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i} - \frac{R_1}{t_{i+1} - t_i}$$

Errores de Truncamiento.

Estimación.

- En la formula se aprecia la aproximación junto con el error de truncamiento

$$v'(t_i) = \frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i} - \frac{R_1}{t_{i+1} - t_i}$$

Empleando la ecuación:

$$R_n = \frac{f^{n+1}(\xi)}{(n+1)!} h^{n+1}$$

Errores de Truncamiento.
Estimación.

Se tiene:

$$R_1 = \frac{v''(\xi)}{2!}(t_{i+1} - t_i)^2$$

$$\frac{R_1}{t_{i+1} - t_i} = O(t_{i+1} - t_i)$$

- La estimación tiene un error de orden $t_{i+1} - t_i$. Al reducir el tamaño del stepsize a la mitad, se reduce el error de truncamiento a la mitad

Redondeo Vs Truncamiento.

- Como puede concluirse de los ejercicios anteriores una disminución en el valor de h disminuye el error de truncamiento, sin embargo, el disminuir h puede ocasionar que el valor de f(x) entre iteraciones sea muy cercano y produzca errores de redondeo en las operaciones aritméticas

Control de Errores I

- Evite realizar restas de cantidades muy cercanas e incurrir en errores de redondeo
- Si los resultados son de uso crítico (medicina, finanzas), asigne un par de grupos para resolver el mismo problema y compare sus resultados.
- Adquiera experiencia resolviendo ejercicios y probando las soluciones para diferentes stepsizes y diferentes métodos

Tipos de Error I

- Error humano
- Error del modelo
- Incertidumbre en los datos

Problemas I

- **Problema:** Encontrar el valor decimal que corresponde al siguiente número binario en formato IEEE-754

0	10101010	10101000000000000000000

- **Problema:** Desarrolle la expansion en Series de Maclaurin para el seno(x) y el coseno(x) y exprese cada expansión como una sumatoria

Problemas I

- **Problema:** Evalue el siguiente polinomio en $x = 1.37$:

$$y = x^3 - 7x^2 + 8x - 0.35$$

Utilice aritmética de 3 cifras significativas con **corte**
(ejemplo de corte: $2.437 \Rightarrow 2.43$)

- **Problema:** Repita el punto anterior expresando y como:

$$y = ((x - 7)x + 8)x - 0.35$$

Encuentre el error relativo para ambos casos y comparelos

Bibliografía I

S. Chapra.
Applied Numerical Methods with MATLAB For Engineers and Scientists, Sixth Edition.
Mac Graw Hill, 2010.