# Instituto Tecnológico de Costa Rica

Área de Ingeniería en Computadores

Análisis Numérico para Ingeniería

Profesor: Jose Pablo Alvarado Moya

Catálogo: Avance # 2

Alumno:
Isaac Núñez Araya 2013030911

Cartago, 2017

# Índice

# 1. Cálculo de errores

## 1.1. Error verdadero

Este error se obtiene sí y sólo si se conoce el valor teórico (o real) de la medición a realizar.

La fórmula es:

$E_t$= valor verdadero - valor medido

## 1.2. Error relativo porcentual verdadero

Es el error verdadero expresado de manera porcentual mediante la siguiente fórmula:

$$E_{rel} = \frac{E_t}{valor\ verdadero} \cdot 100\%$$

El error relativo sigue la misma fórmula, excepto sin el factor de 100 %, este nos da a conocer cuánto porciento nuestra muestra está errada del valor teórico.

## 1.3. Error porcentual aproximado

En este caso, se obtiene un error cuando el valor teórico es desconocido, entonces el reto es acercarse a ese valor aunque sea un misterio. Este error se obtiene mediante la siguiente fórmula: $\epsilon_a = E_{aprox}/V_{aprox}$, donde E es error y V es valor.

Esta ecuación nos muestra que el reto real es estimar el valor del error aproximado, lo cual nos provee una solución para cuando existe un método iterativo para calcular el valor aproximado y se logra mediante la siguiente fórmula:

$$\epsilon_a = \frac{Aprox_{actual} - Aprox_{anterior}}{Aprox_{actual}} \cdot 100\%$$

Se pone una condición de parada dónde seguirá calculando el valor aproximado mientras $|\epsilon_a| > \epsilon_s$. $\epsilon_s$ se define como la cantidad de cifras significativas que deseo que se cumpla en el pro- ceso iterativo mediante la ecuación $\epsilon_s = (0{,}5 \times 10^{2-n})$, donde n es la cantidad de cifras significativas donde debe ser correcto.

# 2. Errores de redondeo debido a codificación de números

## 2.1. Coma fija

En este caso se representa un número con coma fija, donde el peso de cada bit es constante.

Un número entero sin signo se puede representar de la siguiente manera:

$$\sum_{n=0}^{N-1} b_n 2^n$$

donde b sólo puede tomar valores de 0 ó 1 con un rango de representaciones de 0 a $2^N - 1$.

En el caso de un número con coma fija se tiene una expresión simialr a la anterior pero con una pequeña adición de $1/M$, donde:

$$x = \frac{1}{M}\left(\sum_{n=0}^{N-1} b_n 2^n\right), M = 2^m$$

y m es la cantidad de dígitos que se desea luego de la coma.

En el caso de un número con signo, se añade el bit $b_{N-1}$, el cual codifica el signo y el número entero con signo se obtiene de la siguiente manera:

$$x = -b_{N-1}2^{N-1} + \sum_{n=0}^{N-1} b_n 2^n$$

y el rango que se obtiene de representación numérica va desde $-2^{N-1}$ a $2^{N-1} - 1$.

En el caso que se desee obtener un número con coma fija y signo, se obtiene la representación del número mediante:

$$x = \frac{1}{M}\left(-b_{N-1}2^{N-1} + \sum_{n=0}^{N-1} b_n 2^n\right), M = 2^m$$

## 2.2. Coma flotante

Se tiene un número codificado con tres indicadores: bit de signo, exponente $e$ con E bits y la mantisa que representa la parte fraccionaria y el número algebráico se representa como: $x = (-1)^s \times (1, m) \times 2^{e-bias}$, con $bias = 2^{E-1} - 1$.

## 2.3. Error de redondeo

Hay un rango limitado y fuera de ese rango ocurre un desbordamiento y al utilizar el número más cercano hay un error de cuantificación. El problema de usar coma flotante es que el intervalo entre números aumenta drásticamente.

Si se asigna el número por corte, se escoge un número que cumple con $\epsilon = \frac{|\delta x|}{x}$,si es por redondeo se usa $\frac{\epsilon}{2} = \frac{|\delta x|}{x}$,, con $\epsilon = 2^{1-M}$, donde M es la cantidad de bits en la parte fraccionaria.

Además, si se trabaja con números muy similares, se da la presencia de cifras que no son cifras significativas. En procesos iterativos, el error introducido por cada resultado parcial aumenta.

# 3. Errores de truncamiento

## 3.1. Series de Taylor

Se tiene la representación de la Serie de Taylor:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

donde si se trabaja con una representación finita de orden $n$, se obtiene:

$$f(x) \approx \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

Y el residuo se denota como el error de truncamiento, con la siguiente expresión:

$$R_n = \sum_{k=n+1}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

Para funciones contínuas donde no hay cambio de signo y utilizando el Teorema del valor medio se obtiene que: $f(x) \approx g(\xi)(x - x_0)$, con el residuo de la forma:

$$R_n = \frac{f^{n+1}(\xi)}{(n+1)!}(x - x_0)^{n+1}$$

## 3.2. Errores de truncamiento

Existe un problema cuando se desconoce $f^{(n)}(x)$ y $\xi$, pues no se puede calcular el residuo, entonces al truncar una expresión a $n$ términos, se obtiene un error de truncamiento:

$$\frac{R_n}{x_{i+1} - x_i} = \frac{f^{n+1}(\xi)}{(n+1)!}(x_{i+1} - x_i)^{n+1}$$

## 3.3. Diferenciación Numérica

### 3.3.1. Hacia Adelante

Es el cálculo de la derivada mediante la siguiente ecuación:

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h}$$

donde $h$ es el tamaño del paso, dado por: $h = x_{i+1} - x_i$

### 3.3.2. Hacia Atrás

Calcula la derivada tomando puntos atrás del punto donde queremos calcularla mediante:

$$f'(x_i) = \frac{f(x_i) - f(x_{i-1})}{h}$$

donde $h$ es el tamaño del paso, dado por: $h = x_i - x_{i-1}$

### 3.3.3. Centrada

Se calcula la derivada tomando puntos alrededor del punto que buscamos, mediante:

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h}$$

donde $h$ es el tamaño del paso, dado por: $h = x_{i+1} - x_{i-1}$

### 3.3.4. Derivadas superiores

En este caso, se procede con la Serie de Taylor, y para la segunda derivada se utiliza las siguientes ecuaciones:

- **Adelante:**
$$f''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{h^2}$$
- **Atrás:**
$$f''(x_i) = \frac{f(x_i) - 2f(x_{i-1}) + f(x_{i-2})}{h^2}$$
- **Centrada:**
$$f''(x_i) = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{h^2}$$

## 3.4. Propagación del error

Se busca el error entre el valor análito vs la aproximación tomando en cuenta la diferencia $\Delta\tilde{x} = x - \tilde{x}$, por lo tanto $\Delta f(\tilde{x}) \approx |f'(\tilde{x})|\Delta\tilde{x}$, donde $\Delta f(\tilde{x}) = |f(x) - f(\tilde{x})|$.

En el caso de funciones de múltiples variables, se tiene:

$$\Delta f(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) = \left|\frac{\partial f}{\partial x_1}\right| \Delta\tilde{x}_1 + \left|\frac{\partial f}{\partial x_2}\right| \Delta\tilde{x}_2 + \left|\frac{\partial f}{\partial x_3}\right| \Delta\tilde{x}_3 + \cdots + \left|\frac{\partial f}{\partial x_n}\right| \Delta\tilde{x}_n$$

y las operaciones básicas con un error asociado de:

- **Suma:** $\Delta\tilde{u} + \Delta\tilde{v}$
- **Resta:** $\Delta\tilde{u} - \Delta\tilde{v}$
- **Multiplicación** $\Delta|\tilde{v}|\Delta\tilde{u} + |\tilde{u}|\Delta\tilde{v}$
- **División:**
$$\frac{\Delta|\tilde{v}|\Delta\tilde{u} + |\tilde{u}|\Delta\tilde{v}}{|\tilde{v}|^2}$$
Donde $\Delta\tilde{u}$ y $\Delta\tilde{v}$ es la diferencia entre dos puntos de dos funciones diferentes.

### 3.4.1. Estabilidad y condición

El error relativo de:

- **Función:**
$$E_f \approx \frac{f'(\tilde{x})(x - \tilde{x})}{f(\tilde{x})}$$
- **x:**
$$E_x = \frac{x - \tilde{x}}{\tilde{x}}$$

Además el número de condición nos muestra la relación que existe entre $E_f$ y $E_x$, con $N = E_f/E_x$ y nos dice la inexactitud de x con respecto a $f(x)$, donde:

- **N=1:** $E_f = E_x$
- **N>1:** Error relativo crece y si $N \gg 1$ es que $f$ está mal condicionada.
- **N<1:** Error relativo disminuye

  Si la función es mal conficionada, el cálculo se vuelve numéricamente inestable.

#### 3.4.2. Error numérico total (Diferencias centradas)

$$f'(x_i) \approx \frac{f(x_{i+1}) - f(x_{i-1})}{2h} + \frac{e_{i+1} - e_{i-1}}{2h} + \frac{f^{(3)}(x_i)}{3}h^2$$

el primer término es el cálculo, el segundo es el redondeo y el tercero es el error por truncamiento.

Se tienen ciertas consideraciones: si $h$ disminuye, el error por rendonde aumenta y por truncamiento se reduce. Además si se considera $e_i = \epsilon \Rightarrow (e_{i+1} - e_{i-1}) = 2\epsilon$, lo que daría un error total:

$$E_T = \left| f'(x_i) - \frac{f(x_{i+1}) - f(x_{i-1})}{2h} \right| \leq \frac{\epsilon}{h} + \frac{h^2 M}{6}$$

esto si $M$ no es superado por la tercera derivada de $f$, por lo que

$$h_{opt} = \sqrt[3]{\frac{3\epsilon}{M}}$$

# 4. Raíces de ecuaciones

Por métodos gráficos es una aproximación, pues se obtiene directo de la gráfica.

## 4.1. Métodos cerrados

Consideración inicial: $f$ cambia de signo cerca de cierta raíz. (raíces de multiplicidad impar). Además con estos métodos se busca **una** raíz en un intervalo dado.

### 4.1.1. Método de bisección

$$x_r | f(x_r) = 0, x_r \in [x_l, x_u] \Rightarrow f(x_l) \cdot f(x_u) < 0 \; si\ el\ signo\ de\ x_l\ es\ diferente\ al\ de\ x_u$$

Condición de parada:

$$\epsilon_a = \left| \frac{x_r^{(i-1)} - x_r^{(i)}}{x_r^{(i)}} \right| \cdot 100\% < \epsilon_s$$

donde $i$ es debido a que este es un proceso iterativo. Así que reduciendo el error mediante:

$$x_r^{(i)} = \frac{x_l + x_u}{2} \quad y \quad x_r^{(i)} - x_r^{(i-1)} = \frac{x_u - x_l}{2}$$

entonces el error aproximado se expresa como

$$\epsilon_a = \left| \frac{x_u - x_l}{x_u + x_l} \right| \cdot 100\%$$

por lo que el error se reduce con cada iteración tal que:

$$E_a^{(n)} = \frac{\Delta x^{(0)}}{2^n}$$

por lo que si deseamos cierto error, el número de iteraciones está dado por:

$$n = \log_2 \left( \frac{\Delta x^{(0)}}{E_d} \right)$$

El problema que presenta este $\epsilon_a$ es si la raíz es cero o cercana de cero, por lo que se modifca hasta obtener: $\epsilon_a = |x_r^{(i)} - x_r^{(i-1)}|$, donde $\epsilon_s$ se define como:

$$\epsilon_s = \frac{x_u - x_l}{2} \cdot \Phi$$

donde $\Phi$ depende del formato número usado.

### 4.1.2. Método de interpolación lineal

Evita las restricciones del método de bisección mediante tríangulos semejantes tal que:

$$x_l = x_u - \frac{f(x_u)(x_l - x_u)}{f(x_l) - f(x_u)}$$

El problema recae cuando las funciones tienen cambios muy rápidos, por lo cual se debe modificar donde considera que si el cambio en $x$ es muy pequeño, entonces $f(x_i) = f(x_i)/2$.

## 4.2. Métodos abiertos

Posee la ventaja que converge más rápido y necesita pocos valores iniciales; aunque pueden diverger.

### 4.2.1. Iteración de punto fijo

$$\begin{aligned} & f(x) &= 0 \\ \Rightarrow & x &= g(x) \\ \Rightarrow & x_{i+1} &= g(x_i) \\ \therefore & \epsilon_a &= \left|\frac{x_{i+1} - x_i}{x_{i+1}}\right| \cdot 100\% \end{aligned}$$

Por tanto, usando el Teorema del Valor Medio, se establece

$$g'(\xi) = \frac{g(b) - g(a)}{b - a}$$

Ahora, se estudia el cómo converge ese método, tomando la ecuación iterativa y para la raíz verdadera:

$$\begin{aligned} \Rightarrow & x_r - x_{i+1} &= g(x_r) - g(x_{i+1}) \\ \Rightarrow & g'(\xi)(x_r - x_i) &= g(x_r) - g(x_i) \\ \therefore & x_r - x_{i+1} &= g'(\xi)(x_r - x_i) \end{aligned}$$

Por tanto, si el error verdadero en la i-ésima iteración es:

$$\begin{aligned} & E_{t.i} &= x_r - x_i \\ \Rightarrow & E_{t,i+1} &= g'(\xi)E_{t,i} \end{aligned}$$

Donde si $|g'(\xi)| > 1$, este método diverge.

### 4.2.2. Método Newton-Raphson

Se debe encontrar $x_r$ tal que $f(x_r) = 0$, por lo que:

$$\begin{aligned} & f'(x_i) &\approx \frac{f(x_i)}{x_i - x_{i+1}} \\ \Rightarrow & x_{i+1} &= x_i - \frac{f(x_i)}{f'(x_{i+1})} \end{aligned}$$

La estimación del error se obtiene mediante la Serie de Taylor, de manera que:

$$\begin{aligned} 0 &= f(x_i) + f'(x_i)(x_r - x_i) + \frac{f''(\xi)}{2!}(x_r - x_i)^2 \\ 0 &= f(x_i) + f'(x_i)(x_{i+1} - x_i) \\ \Rightarrow \quad 0 &= f'(x_i)(x_r - x_{i+1}) + \frac{f''(\xi)}{2!}(x_r - x_i)^2 \end{aligned}$$

Tomando $E_{t,i} = x_r - x_i$ y $E_{t,i+1} = x_r - x_{i+1}$

$$\Rightarrow \quad 0 = f'(x_i) \cdot E_{t,i+1} + \frac{f''(\xi)}{2!} \cdot E_{t,i}^2$$

Esto anterior significa que si converge, lo hace de manera cuadrática, por lo que las cifras significativas se duplica con cada iteración, por tanto converge si

$$\left| \frac{f''(\xi)}{2f'(x_i)} \right| < 1$$

Al final, se debe asegurar que $f(x_r) \approx 0$ pues este método puede converger a puntos que no son raíz. También se debe considerar que este método puede oscilar debido a puntos de inflexión.

### 4.2.3. Método de la secante

En vez de usar la derivada, utiliza una aproximación hacia atrás, por lo que

$$x_{i+1} = x_i - \frac{f(x_i)(x_{i-1} - x_i)}{f(x_{i-1}) - f(x_i)}$$

Este método cambia pues no es necesario encerrar a $x_r$, además cambia, pues la convergencia es subcuadrática:

$$E_{t.i+1} = const \times E_{t,i}^{1,618}$$

## 4.3. Métodos mixtos

### 4.3.1. Interpolación cuadrática

Dado tres puntos y sus valores evaluados, se asume que:

$$\begin{aligned} y_i &= ax_i^2 + bx_i + c \\ y_{i-1} &= ax_{i-1}^2 + bx_{i-1} + c \\ y_{i-2} &= ax_{i-2}^2 + bx_{i-2} + c \end{aligned}$$

Donde es posible despejar $a, b$ y $c$

### 4.3.2. Interpolación inversa cuadrática

Dado

$$\begin{aligned} x_i &= ay_i^2 + by_i + c \\ x_{i-1} &= ay_{i-1}^2 + by_{i-1} + c \\ x_{i-2} &= ay_{i-2}^2 + by_{i-2} + c \end{aligned}$$

Es posible despejar $a, b$ y $c$. Es importante considerar que lo anterior funciona cuando no hayan dos o más valores iguales para $y$. Por tanto.

$$x = \frac{(y - y_{i-1})(y - y_i)}{(y_{i-2} - y_{i-1})(y_{i-2} - y_i)} \cdot x_{i-2} + \frac{(y - y_{i-2})(y - y_i)}{(y_{i-1} - y_{i-2})(y_{i-1} - y_i)} \cdot x_{i-1} + \frac{(y - y_{i-2})(y - y_{i-1})}{(y_i - y_{i-2})(y_i - y_{i-1})} \cdot x_i$$

Y como se busca cuando $y = 0$, se tiene que:

$$x = \frac{y_{i-1} y_i}{(y_{i-2} - y_{i-1})(y_{i-2} - y_i)} \cdot x_{i-2} + \frac{y_{i-2} y_i}{(y_{i-1} - y_{i-2})(y_{i-1} - y_i)} \cdot x_{i-1} + \frac{y_{i-2} y_{i-1}}{(y_i - y_{i-2})(y_i - y_{i-1})} \cdot x_i$$

## 4.4. Raíces de polinomios

### 4.4.1. Defliación polinomial

Reducción del grado del polinomio dividíendolo por $(x - t_i)$, donde $t_i$ es la raíz conocida del polinomio. Tal que

$$\sum_{i=o}^{n} a_i x^i = a_n \prod_{i=1}^{n} (x - t_i)$$

Ese método posee ciertas fallas, debido a que las raíces son aproximaciones, se ve sensible a los errores de redondeo. Por tanto si se hace hacia adelante, es preferible dividir por las raíces de valor absoluto pequeño; si es hacia atrás se usan las raíces de valor absoluto grande.

La ténica de pulir es usar las raíces obtenidas para encontrar nuevas raíces a partir del polinomio original.

### 4.4.2. Método de Müller

$$\tilde{f}(x) = a(x - x_i)^2 + b(x - x_i) + c$$

Usando tres puntos, $x_{i-2}, x_{i-1}$ y $x_i$:

$$
\begin{aligned}
h_{i-2} &= x_{i-1} - x_{i-2} \\
h_{i-1} &= x_i - x_{i-1} \\
\delta_{i-2} &= \tfrac{f(x_{i-1}) - f(x_{i-2})}{x_{i-1} - x_{i-2}} \\
\delta_{i-1} &= \tfrac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}
\end{aligned}
$$

Por lo que:

$$
\begin{aligned}
a &= \frac{\delta_{i-1} - \delta_{i-2}}{h_{i-1} - h_{i-2}} \\
b &= a \cdot h_{i-1} + \delta_{i-1} \\
c &= f(x_i) \\
\Rightarrow \quad x_{i+1} &= x_i - \frac{2c}{b \pm \sqrt{b^2 - 4ac}}
\end{aligned}
$$

Este método permite reducir el error de redondeo potencial, aunque las raíces pueden ser complejas. Una de las dificultades es que produce dos raíces, entonces se escoge el signo igual a $b$.

En cuanto a iteración de este método, del conjunto $\{x_i, x_{i-1}, x_{i-2}\}$ se escoge las dos que estén más cerca de $x_i + 1$, esto si se busca raíces puramente reales. Si se busca complejas, se utiliza un método secuencial con $\{x_i, x_{i-1}\}$ y $x_{i+1}$.

# 5. Optimización unidimensional

## 5.1. Método cerrado

### 5.1.1. Sección dorada

Es muy similar a la bisección, se da un límite inferior y superior y dentro de él debe estar un máximo o mínimo. Por lo que se busca un razón de ir reduciendo el intevalo y que me queden valores calculados para rehacer los que son verdaderamente necesarios, por tanto:

$$\begin{aligned}
& l_0 && = l1 + l_2 \\
& \tfrac{l_1}{l_o} && = \frac{l_2}{l_1} \\
\Rightarrow \quad & \tfrac{l_1}{l_1+l_2} && = \frac{l_2}{l_1} \\
& R && = \frac{l_2}{l_1} \\
\Rightarrow \quad & \tfrac{1}{1+\frac{l_2}{l_1}} && = \frac{l_2}{l_1} \\
\Rightarrow \quad & \tfrac{1}{1+R} && = R \\
\Rightarrow \quad & R^2 + R - 1 && = 0 \\
\Rightarrow \quad & R && = \frac{\sqrt{5}-1}{2} \approx 0{,}61803
\end{aligned}$$

Por tanto, tomando un intervalo inicial, se crearán dos nuevos puntos $x_1, x_2$. por tanto

$$\begin{aligned}
d & = R(x_u - x_l) \\
x_1 & = x_l + d \\
x_2 & = x_u - d
\end{aligned}$$

Si $f(x_1) > f(x_2)$, entonces el máx está entre $[x_2, x_u]$ sino en $[x_l, x_1]$; donde el peor caso de iteración es cuando el máximo está en los extremos $x_{l,u}$ tal que:

$$\Delta x = (1 - R)(x_u - x_l)$$

De esta manera el error se aproxima con el intervalo dado y valor más óptimo calculado hasta el momento, tal que

$$\epsilon_a = (1 - R)\left|\frac{x_u - x_l}{x_{opt}}\right| \cdot 100\%$$