



Teoría de Errores

Los errores numéricos surgen del uso de aproximaciones para representar operaciones y cantidades matemáticas exactas. En particular, los errores de redondeo se producen cuando se usan números que tienen un límite de cifras significativas para representar números exactos. Numéricamente los errores de redondeo se relacionan de manera directa con la forma en que se guardan los números en la memoria de la computadora. Las cantidades fraccionarias generalmente se representan en la computadora usando la forma de punto flotante.

Representación en Punto Flotante:

Cualquier número real p puede escribirse de forma única por medio de la representación

$$p = \pm 0.d_1d_2 \dots d_k d_{k+1} d_{k+2} d_{k+3} \dots \times 10^n$$

donde $0 \leq d_i \leq 9$, $\forall i \in \mathbb{N}$ (con $1 \leq d_1 \leq 9$) y $n \in \mathbb{Z}$.

Definición. La **representación en punto flotante** de un número real p , denotada por $fl(p)$, es la escritura

$$fl(p) = \pm 0.d_1d_2 \dots d_k \times 10^n$$

donde $1 \leq d_1 \leq 9$, $0 \leq d_i \leq 9$, $\forall i \in \{2, 3, \dots, k\}$ y $n \in \mathbb{Z}$.

De la **representación de punto flotante** anterior se define:

- i) \pm : **signo** del número p . Específicamente $+$ si es positivo y $-$ si es negativo (generalmente omitiremos el signo $+$).
- ii) k : **mantisa** o cantidad de dígitos (desde d_1 hasta d_k)
- iii) n : **exponente** (incluido su signo)

Observación. Para representar un número p mediante su representación en punto flotante $fl(p)$, existen dos métodos:

- 1) Método de **truncamiento**
- 2) Método de **redondeo**

Truncamiento y Redondeo:

Sea p y $fl(p)$, un número y su representación en punto flotante, definidos anteriormente. Considere una mantisa de k dígitos.

Definición. Para representar mediante **truncamiento** se eliminan los dígitos $d_{k+1}d_{k+2}d_{k+3} \dots$ del número p original, para así obtener

$$fl(p) = \pm 0.d_1d_2 \dots d_k \times 10^n$$



Definición. Para representar mediante **redondeo** se verifica si el dígito d_{k+1} es menor a 5, en tal caso se procede a truncar el número en el dígito k (coincide con el **truncamiento**). En el caso que d_{k+1} sea mayor o igual a 5, se suma 1 al dígito d_k y luego se trunca en dicho dígito, para obtener

$$fl(p) = \pm 0.\delta_1\delta_2 \dots \delta_k \times 10^m$$

(observar que el nuevo número puede cambiar completamente).

Debemos, además, tener presente que:

- a) La cantidad de dígitos k de una mantisa, que puede almacenar una máquina es siempre fija.
- b) Aunque en nuestra definición no se vea claramente, la cantidad de dígitos para almacenar el exponente y su signo también es fija. Esto trae como consecuencia que exista un **exponente positivo máximo** y un **exponente negativo mínimo** que podemos representar.

La pregunta natural que podemos hacer a esta última observación es:

¿Qué sucede si se supera el exponente positivo máximo o exponente negativo mínimo?

Overflow y Underflow:

Cuando se intenta almacenar un número cuyo exponente positivo es mayor que el máximo representable, se produce un error llamado **desbordamiento** u **overflow**. En tal caso la máquina interpreta que el número en cuestión es **infinito**. En MATLAB el máximo número representable es aproximadamente 2^{1023} .

Por otro lado, cuando se intenta almacenar un número cuyo exponente negativo es menor que el mínimo representable, se produce un error llamado **subdesbordamiento** o **underflow**. En tal caso la máquina interpreta que el número en cuestión es **0**. En MATLAB el mínimo número positivo representable es aproximadamente 2^{-1074} .

Teoría de Errores:

Considere p^* , una aproximación de un número p . Nos interesa averiguar cuán buena es nuestra aproximación, para ello definimos

Definición. Se define el **error absoluto**, denotado por ε_p , a la cantidad

$$\varepsilon_p = |p - p^*|$$

Definición. Se define el **error relativo** (siempre y cuando $p \neq 0$), denotado por δ_p , a la cantidad

$$\delta_p = \frac{|p - p^*|}{|p|}$$

Generalmente denotamos $p = p^* \pm \varepsilon_p$.

Fuente de Errores:

Existen diferentes formas de introducir errores a los problemas



1. **Errores de planteamiento:** El hecho que los fenómenos a estudiar se modelen matemáticamente, generalmente introducen **simplificaciones** que nos llevan a cometer errores.
2. **Errores implicados en procesos infinitos :** Ejemplo

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

no se pueden realizar sumatorias infinitas.

3. **Errores de redondeo:** Estos son debido a que los computadores sólo tienen una aritmética finita, producto de su limitante de almacenamiento.
4. **Errores de propagación:** El hecho de operar con **términos que involucran errores** producen una **propagación** de los mismos.

Muchas veces nuestro interés será realizar alguna operación con dos o más variables que tienen un error asociado. Por ejemplo: sumarlas, multiplicarlas o dividir las, o bien alguna otra operación más compleja. La pregunta es **¿cómo calculamos el error resultante del efectuar dichas operaciones con nuestras variables sometidas a error?**

Proposición. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable. Sean, además, $x_i = x_i^* \pm \varepsilon_{x_i}$, $\forall i \in \{1, \dots, n\}$ un conjunto de n datos y sus respectivos errores absolutos. Luego, el **error absoluto** ε_f y **error relativo** δ_f , están dados por las expresiones

$$\varepsilon_f \approx \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(x_1^*, x_2^*, \dots, x_n^*) \right| \varepsilon_{x_i} \quad \text{y} \quad \delta_f \approx \frac{\sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(x_1^*, x_2^*, \dots, x_n^*) \right| \varepsilon_{x_i}}{|f(x_1^*, x_2^*, \dots, x_n^*)|}$$

Observación.

1. Las operaciones de producto y cociente en punto flotante siempre dan resultados **confiables**, en el sentido que todos los dígitos del resultado que se obtienen son correctos (siempre que no se produzca **overflow** o **underflow**).
2. Las operaciones de suma y resta usualmente dan resultados con todos sus dígitos correctos. Sin embargo, esto no es así cuando se restan **números muy parecidos**. Este fenómeno se denomina **cancelación excesiva**.

Rapidez de Convergencia de Sucesiones:

Definición. Supongamos que $\{\beta_n\}_{n=1}^{\infty}$ es una sucesión cuyo valor de convergencia es 0 y que $\{\alpha_n\}_{n=1}^{\infty}$ converge a un número α . Si existe una constante positiva K tal que, para n suficientemente grande, se tenga

$$|\alpha_n - \alpha| \leq K|\beta_n|$$

entonces se dice que $\{\alpha_n\}_{n=1}^{\infty}$ converge a α con **rapidez de convergencia** $\mathcal{O}(\beta_n)$ (y se lee **orden** β_n). Escribimos $\alpha_n = \alpha + \mathcal{O}(\beta_n)$.



Aunque β_n es una sucesión arbitraria, generalmente buscaremos aquellas de la forma $\beta_n = \frac{1}{n^p}$ para algún p en particular (nos interesa el mayor que satisfaga la condición). Mientras más grande es p , decimos que nuestra sucesión **converge con mayor rapidez**. Así, escribimos $\alpha_n = \alpha + \mathcal{O}(\frac{1}{n^p})$.

Rapidez de Convergencia de Funciones:

Definición. Supongamos que $\lim_{h \rightarrow 0} G(h) = 0$ y que $\lim_{h \rightarrow 0} F(h) = L$. Si existe una constante positiva K tal que, para h suficientemente pequeño, se tenga

$$|F(h) - L| \leq K|G(h)|$$

entonces se dice que $F(h)$ converge a L con **rapidez de convergencia** $\mathcal{O}(G(h))$ (y se lee **orden** $G(h)$). Escribimos $F(h) = L + \mathcal{O}(G(h))$.

Aunque $G(h)$ es una sucesión arbitraria, generalmente buscaremos aquellas de la forma $G(h) = h^p$ para algún $p > 0$ en particular (nos interesa el mayor que satisfaga la condición). Mientras más grande es p , decimos que nuestra sucesión **converge con mayor rapidez**. Así, escribimos $F(h) = L + \mathcal{O}(h^p)$.

Orden de Convergencia de Sucesiones:

Definición. Supongamos que $\{p_n\}_{n=0}^{\infty}$ es una sucesión que converge a p , con $p_n \neq p$ para todo n . Si existen constantes positivas λ y α con

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

entonces se dice que $\{p_n\}_{n=0}^{\infty}$ converge a p con orden α y **constante de error asintótica** λ .

Ejemplos. Los ejemplos más comunes:

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

DMH/VAD.