TEORÍA DE LOS ERRORES

Los métodos numéricos ofrecen soluciones aproximadas muy cercanas a las soluciones exactas; la discrepancia entre una solución verdadera y una aproximada constituye un error, por lo que es importante saber qué se entiende por aproximar y aprender a cuantificar los errores, para minimizarlos.

Los errores numéricos se generan con el uso de aproximaciones para representar las operaciones y cantidades matemáticas.

La relación entre un resultado exacto o verdadero X y el valor aproximado X* está dado por:

$$X = X^* + \text{error} \qquad (1.1)$$

El que un error tenga signo positivo o negativo, generalmente no tiene importancia, de manera que el error absoluto se define como el valor absoluto de la diferencia entre el valor verdadero y el valor aproximado:

$$E = |X - X^*| \qquad (1.2)$$

El error absoluto se expresa en las mismas unidades que X y no toma en cuenta el orden de magnitud de la cantidad que se está midiendo.

El error relativo normaliza el error absoluto respecto al valor verdadero de la cantidad medida:

$$\mathbf{e} = |E/X| = |(X - X^*)/X| \qquad (1.3)$$

El error relativo es adimensional y puede quedar expresado así, en forma fraccional, o se puede multiplicar por 100 para expresarlo en términos porcentuales:

$$\mathbf{e}\ (\%) = |E/X| \times 100 \qquad (1.4)$$

Las ecuaciones (1.1), (1.2), (1.3) y (1.4) suponen que se conoce el valor verdadero de X, lo que hace que los errores absoluto y relativo: E y **e** sean también verdaderos. Pero normalmente X no se conoce; no tendría sentido considerar una aproximación, si se conociese el valor verdadero.

La mejor estimación posible del verdadero valor de X es su aproximación X* y se define entonces una estimación del error relativo como:

$$\mathbf{e}^* = |E/X^*| \qquad (1.5)$$

Pero el problema está en cómo estimar E, en ausencia de conocimiento del verdadero valor de X.

Algunos métodos numéricos usan un esquema iterativo en los que se hace una aproximación con base en la aproximación previa y esto se hace varias veces, para obtener cada vez mejores aproximaciones:

$$\mathbf{e}^* = |\,(\text{valor actual - valor anterior})/\text{valor actual}\,| \qquad (1.6)$$

Los cálculos se repiten hasta que: $\mathbf{e}^* < \mathbf{e}_0$, donde $\mathbf{e}_0$ es un valor prefijado previamente.

Los errores numéricos se clasifican, por su origen, en tres tipos: errores inherentes, errores de redondeo y errores por truncamiento, cada uno de los cuales merece un tratamiento por separado.

Los errores inherentes se producen por la propia variabilidad de los fenómenos; al ser caracterizados a través de cantidades físicas, las mediciones conllevan incertidumbre, pues los instrumentos de medición ofrecen sólo una aproximación numérica del valor verdadero de la magnitud medida, pues se calibran para considerar solamente un determinado número

de cifras significativas. Todas las magnitudes que se manejan en ingeniería son susceptibles a este tipo de errores.
Por ejemplo, cuando se dice que el tirante de agua de una presa es de 123 m, habiendo hecho la medición mediante un dispositivo que ofrece una precisión de tres cifras significativas, el tirante de agua realmente puede fluctuar entre 122.5 y 123.5 m.

$$X\ [122.5, 123.5) \qquad X^* = 123$$

El error inherente absoluto máximo que se puede llegar a cometer cumple con la desigualdad:

$$E_{max} \leq 0.5 \text{ m};$$

y el correspondiente error inherente relativo máximo cumple con la desigualdad:

$$e_{max} \leq 0.5/122.5 = 0.00408$$

El error inherente absoluto medio que se puede cometer cumple con la desigualdad:

$$E_{med} \leq 0.25 \text{ m};$$

y el correspondiente error inherente relativo medio cumple con la desigualdad:

$$e_{med} \leq 0.00204.$$

Algunos autores mencionan dentro de esta clasificación los errores humanos que se cometen al hacer la lectura de una medida, al transmitirla o al transcribirla; pero, en virtud de que estos errores de lectura, transmisión o transcripción pueden constituirse en pifias garrafales que quedan fuera de todo control, no es posible estimarlos en forma sistematizada. Por ejemplo, si al transcribir en un documento la densidad de un producto, se anota 1.381 en vez de 1.831, que es la medida leída, la pifia es imposible de manejar y predecir.
Los errores de redondeo se producen al realizar operaciones aritméticas en las que el resultado produce una mantisa cuyo número de dígitos difiere significativamente del número de dígitos de la mantisa de alguno de los valores numéricos involucrados en la operación. Al manejar un determinado número de cifras significativas en los cálculos, el resultado tiene que ser redondeado de alguna manera, sobrestimando o subestimando el valor resultante verdadero.
Sea X el resultado de una operación aritmética, el cual puede ser expresado mediante notación matemática, en forma normalizada: $F \times 10^n$, donde F está formada por m cifras obtenidas en el resultado, de las cuales, n son enteras. Este valor se puede descomponer en dos sumandos, igualmente normalizados: el primero formado por t cifras significativas, las t primeras cifras del resultado después del punto decimal: $f \times 10^n$, y el segundo formado por las (m-t) cifras no significativas del resultado, $g \times 10^{n-t}$:

$$X = F \times 10^n = f \times 10^n + g \times 10^{n-t}$$

En virtud de que F, f y g son números normalizados, su valor absoluto puede tomar algún valor dentro del intervalo semiabierto [0.1, 1). F está formado por m dígitos, f está formada por t dígitos y g está formada por (m-t) dígitos.

$0.1 \leq \lvert F\rvert < 1$;	$0.1 \leq \lvert f\rvert < 1$;	$0 \leq \lvert g\rvert < 1$
[0.1, 0.999...99]	[0.1, 0.999...99]	[0, 0.999...]
m dígitos	t dígitos	(m-t) dígitos

Al considerar únicamente t cifras significativas, se están despreciando (m-t) cifras del resultado, es decir, se está redondeando el resultado. Ahora bien, hay dos maneras de hacer ese redondeo: la primera consiste en tomar como aproximación numérica X* de la operación realizada el valor $f \times 10^{n}$, haciendo caso omiso del valor de $g \times 10^{n-t}$; la segunda consiste en tomar como aproximación numérica X* el valor $f \times 10^{n}$, pero ajustado conforme al valor que tenga el primer dígito de $g \times 10^{n-t}$.

Redondeo truncado: $X^* = f \times 10^{n}$ (1.7)

El error absoluto que se comete en cada caso particular es:

$$E = |g| \times 10^{n-t}$$

El error absoluto máximo que se puede llegar a cometer, en cualquier caso, es:

$$E_{max} < 1 \times 10^{n-t}$$

Y el error absoluto esperado que se puede cometer, considerando una distribución de probabilidad uniforme para los errores, es:

$$E_{med} < 0.5 \times 10^{n-t}$$

El error relativo que se comete en cada caso particular es:

$$e = |g/F| \times 10^{1-t}$$

El error relativo máximo que se puede llegar a cometer, en todo caso, es:

$$e_{max} < 1 \times 10^{1-t}$$

Y el error relativo esperado o promedio que se puede cometer es:

$$e_{med} < 0.5 \times 10^{1-t}$$

Puesto que X no siempre se puede conocer con exactitud, F tampoco, por lo que es imposible calcular los errores verdaderos, se recurre a sus estimaciones:

El error relativo estimado que se comete en cada caso particular es:

$$e^* = |g/f| \times 10^{1-t}$$

El error relativo máximo estimado que se puede llegar a cometer es:

$$e^*_{max} < 1 \times 10^{1-t}$$

Y el error relativo esperado estimado que se puede cometer es:

$$e^*_{med} < 0.5 \times 10^{1-t}$$

Ejemplo: Efectuar la suma: 162.4 + 1.769, considerando 4 cifras significativas con redondeo truncado.

$$\begin{array}{rcccc} 162.4 & = & 0.1624 \times 10^{3} & = & 0.1624 \times 10^{3} \\ +\ 1.769 & = & 0.1769 \times 10^{1} & = & 0.001769 \times 10^{3} \\ \hline 164.169 & & & & 0.164169 \times 10^{3} \end{array}$$

$X = 0.164169 \times 10^{3} = 0.1641 \times 10^{3} + 0.000069 \times 10^{3}$

$= 0.1641 \times 10^{3} + 0.6900 \times 10^{-1}$; $n = 3$; $t = 4$

$X^* = 0.1641 \times 10^{3}$

$E = |X - X^*| = |0.164169 \times 10^{3} - 0.1641 \times 10^{3}| =$

$E = |0.000069 \times 10^{3}| = |0.69 \times 10^{-1}| = 0.69 \times 10^{-1} < 1 \times 10^{-1}$

$\mathbf{e} = |E/X| = |(0.69 \times 10^{-1})/(0.164169 \times 10^{3})| = 0.4203 \times 10^{-3} < 1 \times 10^{-3}$

$\mathbf{e}^* = |E/X^*| = |(0.69 \times 10^{-1})/(0.1641 \times 10^{3})| = 0.4205 \times 10^{-3} < 1 \times 10^{-3}$

Redondeo simétrico:

$$X^* = \begin{cases} f \times 10^{n} & ; \text{ si } |g| < 0.5 \\ f \times 10^{n} + 1 \times 10^{n-t} & ; \text{ si } |g| \geq 0.5 \end{cases} \qquad (1.8)$$

El error absoluto que se comete en cada caso particular es:

$$E = \begin{cases} |g| \times 10^{n-t} & ; \text{ si } |g| < 0.5 \\ |1 - g| \times 10^{n-t} & ; \text{ si } |g| \geq 0.5 \end{cases}$$

El error absoluto máximo que se puede llegar a cometer, en cualquier caso, es:

$$E_{max} < 0.5 \times 10^{n-t}$$

Y el error absoluto esperado que se puede cometer, considerando una distribución de probabilidad uniforme para los errores, es:

$$E_{med} < 0.25 \times 10^{n-t}$$

El error relativo que se comete en cada caso particular es:

$$e = \begin{cases} |g/F| \times 10^{-t} & ; \text{ si } |g| < 0.5 \\ |(1-g)/F| \times 10^{-t} & ; \text{ si } |g| \geq 0.5 \end{cases}$$

El error relativo máximo que se puede llegar a cometer, en todo caso, es:

$$e_{max} < 0.5 \times 10^{1-t}$$

Y el error relativo esperado o promedio que se puede cometer es:

$$e_{med} < 0.25 \times 10^{1-t}$$

El error relativo estimado que se comete en cada caso particular es:

$$\mathbf{e}^* = \begin{cases} |g/f| \times 10^{-t} & ; \text{ si } |g| < 0.5 \\ |(1-g)/f| \times 10^{-t} & ; \text{ si } |g| \geq 0.5 \end{cases}$$

El error relativo máximo estimado que se puede llegar a cometer es:

$$e^*_{max} < 0.5 \times 10^{1-t}$$

Y el error relativo esperado estimado que se puede cometer es:

$$e^*_{med} < 0.25 \times 10^{1-t}$$

Ejemplo: Efectuar la suma: 162.4 + 1.769, considerando 4 cifras significativas con redondeo simétrico.

$$\begin{array}{rcccl} 162.4 & = & 0.1624 \times 10^{3} & = & 0.1624 \times 10^{3} \\ +\ 1.769 & = & 0.1769 \times 10^{1} & = & 0.001769 \times 10^{3} \\ \hline 164.169 & & & & 0.164169 \times 10^{3} \end{array}$$

$$X = 0.164169 \times 10^{3} = 0.1641 \times 10^{3} + 0.000069 \times 10^{3}$$

$$= 0.1641 \times 10^{3} + 0.6900 \times 10^{-1} \quad ; \quad n = 3 \ ; \ t = 4$$

$X^* = 0.1642 \times 10^{3}$

$E = |X - X^*| = |0.164169 \times 10^{3} - 0.1642 \times 10^{3}| =$

$= |0.000031 \times 10^{3}| = |0.31 \times 10^{-1}| = 0.31 \times 10^{-1} < 0.5 \times 10^{-1}$

$\mathbf{e} = |E/X| = |(0.31 \times 10^{-1})/(0.164169 \times 10^{3})| = 0.1888 \times 10^{-3} < 0.5 \times 10^{-3}$

$\mathbf{e}^* = |E/X^*| = |(0.31 \times 10^{-1})/(0.1642 \times 10^{3})| = 0.1888 \times 10^{-3} < 0.5 \times 10^{-3}$

Ejemplo: Efectuar la siguiente suma:

$$0.9999 \times 10^{0}$$
$$0.8888 \times 10^{1}$$
$$0.7777 \times 10^{2}$$
$$0.6666 \times 10^{3}$$
$$0.5555 \times 10^{4}$$
$$0.4444 \times 10^{5}$$
$$0.3333 \times 10^{6}$$
$$0.2222 \times 10^{7}$$
$$0.1111 \times 10^{8}$$

a) primero considerando todas las cifras incluidas en los sumandos.

b) luego efectuando la suma de arriba hacia abajo, con redondeo simétrico y $t = 4$, calcule el error relativo porcentual correspondiente.

c) luego efectuando la suma de abajo hacia arriba, con redondeo simétrico y $t = 4$, calcule el error relativo porcentual correspondiente.

d) compare los errores cometidos en los incisos b y c, y saque conclusiones.

a) la suma exacta es: $X = 13'716,049.2579$

b) la suma descendente da: $X^* = 13'720,000 = 0.1372 \times 10^{8}$

$E = 3960.7421$; $e = 0.2888 \times 10^{-3}$

c) la suma ascendente da: $X^* = 13'710,000 = 0.1371 \times 10^{8}$

$E = 6049.2579$; $e = 0.4410 \times 10^{-3}$

d) Es mayor el error haciendo la suma de abajo hacia arriba que haciéndola de arriba hacia abajo. Aunque teóricamente sabemos que la suma es conmutativa, resulta que al redondear simétricamente a 4 cifras significativas, el orden en que se realiza la suma afecta el resultado.

En ambos casos se cumple que: $e < e_{max} = 0.5 \times 10^{1-t} = 0.5 \times 10^{-3}$

En este caso, $e > e_{med} = 0.25 \times 10^{-3}$, lo que se explica porque el ejemplo está amañado para que dé errores significativos.

Los errores por truncamiento ocurren cuando un número, cuya parte fraccionaria está constituida por un número infinito de dígitos, requiere ser representado numéricamente en forma aproximada, utilizando un determinado número de cifras significativas.

Por ejemplo, 3.1416 es una buena aproximación del número π, pero el valor exacto no puede ser expresado numéricamente por completo, pues consta de un número infinito de dígitos: 3.141592653589793...; lo mismo ocurre con el 2.7183 para el número **e**, el 1.4142 para $\sqrt{2}$, y el 0.33333 para 1/3.

Sin embargo, todos los números, ya sean enteros, racionales o irracionales, pueden ser representados a través de formulaciones matemáticas exactas, utilizando series infinitas; obviamente, las representaciones numéricas acotadas a un determinado número de cifras significativas, son aproximaciones numéricas que llevan implícitos errores por truncamiento.

Por ejemplo, los números 1, 1/3 y **e** pueden expresarse matemáticamente, de manera exacta, a través de las siguientes series infinitas:

$$1 = 1/2 + 1/4 + 1/8 + 1/16 + 1/32 + 1/64 + ...$$
$$1/3 = 3/10 + 3/100 + 3/1000 + 3/10000 + 3/100000 + ...$$
$$\mathbf{e} = 1/0! + 1/1! + 1/2! + 1/3! + 1/4! + 1/5! + ...$$

En este último caso, por ejemplo, la aproximación se puede hacer truncando la serie en cualquier punto, lo que equivale a incluir 1, 2, 3, ..., ó n términos de la serie. Si tomásemos como valor "exacto" de **e,** 2.7182818, tendríamos:

Términos	Aproximación	Error absoluto	Error relativo (%)
1	1.0000000	1.7182818	63.21
2	2.0000000	0.7182818	26.42
3	2.5000000	0.2182818	8.03
4	2.6666667	0.0516151	1.90
5	2.7083333	0.0099485	0.37
6	2.7166667	0.0016152	0.06
7	2.7180556	0.0002263	0.01
8	2.7182540	0.0000279	$\cong$ 0.00

El manejo de series infinitas para aproximar valores específicos de funciones matemáticas es fundamental para comprender a plenitud la mayor parte de los métodos numéricos incluidos en este curso, así como para calcular los errores por truncamiento asociados a esas aproximaciones. Este tema será retomado en otra parte del curso, donde se trata con detalle el uso de la serie de Taylor.

Sean ***X*, *Y*** valores exactos; sean X, Y sus aproximaciones. Sean E_x y E_y los errores absolutos inherentes o por truncamiento, asociados a esas aproximaciones numéricas: sean e_x y e_y los errores relativos correspondientes. Sea E_r el error absoluto de redondeo que se puede cometer al realizar cualquier operación aritmética; y e_r el error relativo de redondeo correspondiente.

Suma: $\boldsymbol{X} + \boldsymbol{Y} = X + E_x + Y + E_y + E_r = (X + Y) + (E_x + E_y) + E_r$

$$E_{x+y} = E_x + E_y + E_r$$
$$\mathbf{e}_{x+y} = (E_x + E_y + E_r)/(X + Y)$$

$$= [X/(X + Y)](E_x/X) + [Y/(X + Y)](E_y/Y) + e_r$$

$$\mathbf{e}_{x+y} = [X/(X + Y)]\mathbf{e}_x + [Y/(X + Y)]\mathbf{e}_y + e_r \qquad (1.9)$$

Resta: $\boldsymbol{X} - \boldsymbol{Y} = (X + E_x) - (Y + E_y) + E_r = X + E_x - Y - E_y + E_r$

$$= (X - Y) + (E_x - E_y) + E_r$$

$$E_{x-y} = E_x - E_y + E_r$$

$$\mathbf{e}_{x-y} = (E_x - E_y + E_r)/(X - Y) =$$

$$= [X/(X - Y)](E_x/X) - [Y/(X - Y)](E_y/Y) + e_r =$$

$$\mathbf{e}_{x-y} = [X/(X - Y)]\mathbf{e}_x - [Y/(X - Y)]\mathbf{e}_y + e_r \qquad (1.10)$$

Producto: $\boldsymbol{X \cdot Y} = (X + E_x)(Y + E_y) + E_r = X \cdot Y + XE_y + YE_x + E_x E_y + E_r$

$$= X \cdot Y + X E_y + Y E_x + E_r$$

$$E_{xy} = X E_y + Y E_x + E_r$$

$$\mathbf{e}_{xy} = (X E_y + Y E_x + E_r)/X Y = E_x/X + E_y/Y + e_r$$

$$\mathbf{e}_{xy} = \mathbf{e}_x + \mathbf{e}_y + e_r \qquad (1.11)$$

Cociente: $\boldsymbol{X}/\boldsymbol{Y} = (X + E_x)/(Y + E_y) + E_r = (X + E_x)/[Y(1 + E_y/Y)] + E_r$

$$= [(X + E_x)/Y] [1/(1 + E_y/Y)] + E_r$$

$$= [(X + E_x)/Y] [1 - E_y/Y + (E_y/Y)2 - (E_y/Y)3 + ...] + E_r$$

$$= [(X + E_x)/Y] (1 - E_y/Y) + E_r$$

$$= [(X + E_x)/Y] + [(X + E_x)E_y/Y2] + E_r$$

$$= X/Y + E_x/Y - XE_y/Y2 - {}_{ExEy/Y2} + Er$$

$$= X/Y + Ex/Y - XEy/Y2 + Er$$

$$E_{x/y} = Ex/Y - XEy/Y2 + Er = (YEx - XEy)/Y2 + E_r$$

$$\mathbf{e}_{x/y} = [(YE_x - XE_y)/Y2 + E_r]/(X/Y) = (YE_x - XE_y)/XY + e_r$$

$$= E_x/X - E_y/Y + e_r$$

$$\mathbf{e}_{x/y} = \mathbf{e}_x - \mathbf{e}_y + e_r \qquad (1.12)$$

Las ecuaciones (1.10), (1.11), (1.12) y (1.13) muestran como se propagan los errores al efectuar las operaciones aritméticas de suma, resta, producto y cociente, respectivamente. La propagación de los errores crece en la medida que se efectúan más y más operaciones, aunque eventualmente llegan a disminuir por efecto de compensación, cuando éstas se combinan.

Cada una de las operaciones vistas anteriormente puede ser representada a través de un gráfica de proceso, como se muestra a continuación:

a) Suma: $\mathbf{e}_{x+y} = [X/(X + Y)]\mathbf{e}_x + [Y/(X + Y)]\mathbf{e}_y + e_r$

b) Resta: $\mathbf{e}_{x-y} = [X/(X - Y)]\mathbf{e}_x - [Y/(X - Y)]\mathbf{e}_y + e_r$

c) Producto: $\mathbf{e}_{xy} = \mathbf{e}_x + \mathbf{e}_y + e_r$

c) Cociente: $\mathbf{e}_{x/y} = \mathbf{e}_x - \mathbf{e}_y + e_r$