Elementos de Cálculo Numérico - Cálculo Numérico
Primer Cuatrimestre de 2021
Entrega n°1

---

1. Definimos una operación $*$ en $\mathbb{R}$, $x * y := -2xy$. Se quiere calcular $x * y$ con aritmética de punto flotante. Mostrar que el error relativo de esta operación es $O(\varepsilon)$ donde ε es el épsilon de la máquina.

Elementos de Cálculo Numérico - Cálculo Numérico
Primer Cuatrimestre de 2021
Entrega n°1 - Resolución del ejercicio

---

1. Queremos estudiar el error relativo de la operación $*$ y mostrar que es $O(\varepsilon)$. Primero, asumimos que -2 es un número de máquina. Por otro lado, si $x = 0$ o $y = 0$, entonces $x * y$ también dará cero, y el error (tanto absoluto como relativo) en ese caso será cero. Si $x, y \neq 0$, lo que debemos acotar entonces es el siguiente cociente:

$$\frac{|fl(fl(-2fl(x))fl(y)) - (-2xy)|}{|-2xy|}.$$

Es muy importante notar que, además de no olvidar que x e y no necesariamente son números de máquina, cada operación que se realiza introduce un error que se propaga. La notación que conviene usar para esta cuenta es la siguiente: $fl(z) = z(1 + \delta_z)$. Veamos entonces cómo es la cuenta que debemos realizar:

$$\frac{|fl(fl(-2fl(x))fl(y)) - (-2xy)|}{|-2xy|} = \frac{|((-2x(1+\delta_x))(1+\delta_{\times x})y(1+\delta_y))(1+\delta_{\times y}) + 2xy|}{|2xy|},$$

donde $\delta_{\times x}$ es el δ de multiplicar por x, y $\delta_{\times y}$ es el δ de multiplicar por y.

Reacomodamos:

$$((-2x(1+\delta_x))(1+\delta_{\times x})y(1+\delta_y))(1+\delta_{\times y}) = -2xy(1+\delta_x)(1+\delta_{\times x})(1+\delta_y)(1+\delta_{\times y})$$

Si sacamos de factor común $-2xy$ del numerador, cancelamos con el denominador y obtenemos:

$$\frac{|fl(fl(-2fl(x))fl(y)) + 2xy|}{|2xy|} = |(1+\delta_x)(1+\delta_{\times x})(1+\delta_y)(1+\delta_{\times y}) - 1|.$$

Distribuimos:

$$\begin{aligned}(1+\delta_x)(1+\delta_y)(1+\delta_{\times x})(1+\delta_{\times y}) = & 1 + \delta_x + \delta_y + \delta_{\times x} + \delta_{\times y} + \delta_x\delta_y + \delta_x\delta_{\times x} + \delta_x\delta_{\times y} + \\ & + \delta_y\delta_{\times x} + \delta_y\delta_{\times y} + \delta_{\times x}\delta_{\times y} + \delta_x\delta_y\delta_{\times x} + \delta_x\delta_y\delta_{\times y} + \\ & + \delta_x\delta_{\times x}\delta_{\times y} + \delta_y\delta_{\times x}\delta_{\times y} + \delta_x\delta_y\delta_{\times x}\delta_{\times y}\end{aligned}$$

Luego, el cociente que queremos acotar es

$$\begin{aligned}\frac{|fl(fl(-2fl(x))fl(y)) + 2xy|}{|-2xy|} &= |\delta_x + \delta_y + \delta_{\times x} + \delta_{\times y} + \cdots + \delta_x\delta_y\delta_{\times x}\delta_{\times y}| \\ &\leq |\delta_x| + |\delta_y| + |\delta_{\times x}| + |\delta_{\times y}| + \cdots + |\delta_x\delta_y\delta_{\times x}\delta_{\times y}|.\end{aligned}$$

Como cada δ se acota en módulo por el épsilon de máquina, tenemos:

$$\frac{|fl(fl(-2fl(x))fl(y)) + 2xy|}{|-2xy|} \leq 4\,\varepsilon + 6\,\varepsilon^2 + 4\,\varepsilon^3 + \varepsilon^4$$

Veamos que $4\ \varepsilon + 6\ \varepsilon^2 + 4\ \varepsilon^3 + \varepsilon^4 = O(\varepsilon)$:

$$\lim_{\varepsilon \to 0} \frac{4\ \varepsilon + 6\ \varepsilon^2 + 4\ \varepsilon^3 + \varepsilon^4}{\varepsilon} = 4.$$

Luego, el cociente está acotado y por lo tanto existen $C > 0$ y $\delta > 0$ tales que si $|\varepsilon| < \delta$ entonces $|4\ \varepsilon + 6\ \varepsilon^2 + 4\ \varepsilon^3 + \varepsilon^4| < C\varepsilon$, por lo que $4\ \varepsilon + 6\ \varepsilon^2 + 4\ \varepsilon^3 + \varepsilon^4 = O(\varepsilon)$. Y así el error relativo de la operación $*$ también es es $O(\varepsilon)$.