Universidad Nacional de Ingeniería
Facultad de Ciencias
Escuela Profesional de Matemática **Ciclo 2024-2**

[Cod: CM4F1 Curso: Análisis y Modelamiento Numérico I] **Práctica Calificada 1**
[Prof: L. Roca, A. Ramirez, M. Quiñones.] **Duración:** 110 **minutos.**

SOLUCIONARIO

1. Determine if the following statement is true or false. If the statement is true, then prove it. If the statement is false, then give one example where the statement fails, or prove that it is false.

   a) Si se usa aritmética de cuatro dígitos con redondeo, al calcular

   $$\frac{3\sqrt{2}-2{,}445\sqrt{3}}{17}$$

   el error relativo es aproximadamente 2,85 %. (1pts)

   b) Al representar el número $\sqrt{2} \approx 1{,}4142$ en base 10 en un computador que tiene 2 bits para la parte entera y tres para la parte fraccionaria nos da un error absoluto de 0.0392. (Use truncamiento de ser necesario). (1.5pts)

   c) Considere la matriz:

   $$A = \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{1}{\varepsilon} \end{pmatrix}$$

   entonces, su número de condición en la norma infinito tiende para cero cuando ε se aproxima a cero. (1.5pts)

Solución.

a) **Verdadero.** El valor exacto es:

$$\eta = \frac{3\sqrt{2}-2{,}445\sqrt{3}}{17} = 4{,}574389773 \times 10^{-4}.$$

Con aritmética de 4 dígitos resulta:

$$\frac{3\sqrt{2}-2{,}445\sqrt{3}}{17} = \frac{3\times 1{,}414 - 2{,}445\times 1{,}732}{17} = \frac{4{,}242-4{,}234}{17} = 4{,}705\times 10^{-4}.$$

Luego, el error relativo es:

$$ER(\eta) = \frac{4{,}574389773\times 10^{-4} - 4{,}705\times 10^{-4}}{4{,}574389773\times 10^{-4}} \times 100\,\% \approx 2{,}85\,\%$$

b) **Verdadero.** El computador usa 2 dígitos para la parte entera y 3 bits para la parte fraccionaria. Luego, convertimos $\sqrt{2} \approx 1{,}4142$ a base 2:

$$\begin{aligned}
1{,}4142 &= 1 + 0{,}4142 \\
2(0{,}4142) &= 0{,}8284 \Rightarrow d_1 = 0 \\
2(0{,}8284) &= 1{,}6568 \Rightarrow d_2 = 1 \\
2(0{,}6568) &= 1{,}3136 \Rightarrow d_3 = 1 \\
2(0{,}3136) &= 0{,}6272 \Rightarrow d_4 = 0 \\
2(0{,}6272) &= 1{,}2544 \Rightarrow d_5 = 1
\end{aligned}$$

luego: $1{,}4142 = 1{,}01101_{(2)}$ cuya representación en el computador dado es:

$$1{,}011$$

que al ser expresado en base 10 resulta:

$$1{,}011 = 1 + 0 + \frac{1}{4} + \frac{1}{8} = \frac{11}{8} = 1{,}375.$$

Por tanto, el error absoluto cometido es:

$$EA = 1{,}4142 - 1{,}375 = 0{,}0392,$$

es decir, la proposición es verdadera.

c) **Falso.** Observe que en la norma infinito se tiene

$$Cond_\infty(A) = \frac{1}{\varepsilon}$$

entonces $Cond_\infty(A) \to \infty$ cuando $\varepsilon \to 0$.

□

2. Para cada una de la siguientes funciones

$$f(x) = (x-1)^\alpha, \alpha > 0, x > 1$$

$$g(x) = x^{-1}e^x, x > 0$$

a) Calcule el número de condición. (2pts)

b) ¿Qué valores de x con errores relativos pequeños producen errores relativos grandes? (2pts)

Solución.

a)

- Para $f'(x) = \alpha(x-1)^{\alpha-1}$,

$$\kappa(f) = \left|\frac{x\alpha(x-1)^{\alpha-1}}{(x-1)^\alpha}\right| = \frac{|\alpha x|}{|x-1|}$$

- Para $g'(x) = e^x(x-1)/x^2$,

$$\kappa(g) = \left|\frac{xx^{-2}e^x(x-1)}{x^{-1}e^x}\right| = |x-1|$$

b) Errores relativos pequeños producen errores relativos grandes si $\kappa > 1$

- $$\kappa(f) = \frac{|\alpha x|}{|x-1|} > 1 \implies \begin{cases} x \in \left\langle \frac{1}{1-\alpha}, \frac{1}{1+\alpha} \right\rangle, & 1 < \alpha \\ \left\langle -\infty, \frac{1}{1+\alpha} \right\rangle \cup \left\langle \frac{1}{1-\alpha}, +\infty \right\rangle & 0 < \alpha < 1 \end{cases}$$

- $$\kappa(g) = |x-1| > 1 \implies x \in \langle -\infty, 0\rangle \cap \langle 2, \infty\rangle$$

□

3. Prove that if $x \in \mathbb{F}(\beta, p, L, U)$ then

$$\beta^{L-1} \leq |x| \leq \beta^U(1 - \beta^{-p})$$

Solución.
Recordemos que

$$|x| = (0.d_1 d_2 \ldots, d_p) \times \beta^e, \beta^{-1} \leq d_1 \leq \beta - 1, L \leq e \leq U$$

por lo tanto el valor mínimo será

$$|x|_{min} = (0{,}100000)_{(\beta)} \times \beta^L = \beta^{L-1}$$

por otro lado

$$|x|_{max} = ((\beta-1)\beta^{-1} + (\beta-1)\beta^{-2} + \ldots + (\beta-1)\beta^{-p}) \times e^U = (\beta-1)\sum_{i=1}^{p}\left(\frac{1}{\beta}\right)^i \times e^U$$

$$x_{max} = (\beta-1)\frac{1}{\beta}\frac{\beta^{-p}-1}{\beta^{-1}-1} \times e^U = (1-\beta^{-p}) \times e^U$$

finalmente

$$\beta^{L-1} \leq |x| \leq (1-\beta^{-p}) \times e^U$$

□

(4pts)

4. Encuentre el error relativo de la siguiente operación $\sum_{i=1}^{10000} 0{,}1$ realizada en una computadora que utiliza 32 bits para los cálculos de punto flotante. (4pts)

Solución.
Sea $S = \sum_{i=1}^{10000} 0{,}1$, $\hat{S}$ su valor aproximado en punto flotante y el algoritmo

$$\begin{cases} s_1 = 0{,}1 \\ s_n = s_{n-1} + 0{,}1, \quad n = 2, \ldots, 10^4 \end{cases}$$

Los cálculos en punto flotante se realizan de la siguiente manera

$$\begin{aligned} \hat{s}_1 &= \text{fl}(0{,}1) \\ \hat{s}_2 &= \text{fl}(\hat{s}_1 + \text{fl}(0{,}1)) \\ \hat{s}_3 &= \text{fl}(\hat{s}_2 + \text{fl}(0{,}1)) \\ &\vdots = \vdots \end{aligned}$$

sabemos que la aritmética de punto flotante sigue la relación $\text{fl}(x) = x(1+\delta)$ entonces

$$\begin{aligned} \hat{s}_1 &\approx 0{,}1 + 0{,}1\delta \\ \hat{s}_2 &\approx (\hat{s}_1 + (0{,}1)(1+\delta))(1+\delta) = 0{,}2(1+\delta)^2 \approx 0{,}2 + 0{,}4\delta \\ \hat{s}_3 &\approx (\hat{s}_2 + (0{,}1)(1+\delta))(1+\delta) \approx 0{,}3 + 0{,}8\delta \\ \hat{s}_4 &\approx 0{,}4 + 1{,}3\delta \\ \hat{s}_5 &\approx 0{,}5 + 1{,}9\delta \\ &\vdots = \vdots \\ \hat{s}_n &\approx (0{,}1)n + (0{,}1)(n^2/2 + 3n/2 - 1)\delta \end{aligned}$$

considerando que el valor exacto $s_n = (0{,}1)n$ entonces el error relativo es

$$\frac{\hat{s}_n - s_n}{s_n} \approx \frac{n^2/2 + 3n/2 - 1}{n}\delta$$

en un computadora de 32 bits, $\delta \approx 2^{-24} \approx 10^{-7}$ entonces para $n = 10^4$

$$\frac{\hat{S} - S}{S} \approx 5 \times 10^{-4}$$

□

5. Trabajo de laboratorio (4pts)

UNI, 11 de setiembre de 2024[1]

[1] Hecho en LaTeX