INSTITUTO SUPERIOR TÉCNICO

Mestrado em Engenharia Electrotécnica e de Computadores

Ano Lectivo: 2007/2008 Semestre: 2º

MATEMÁTICA COMPUTACIONAL

Exercícios[1]

[1.1] Represente x num sistema de ponto flutuante com 4 dígitos na mantissa e arredondamento simétrico, nos seguintes casos:

(a) $x = 1/6$; **(b)** $x = 1/3$; **(c)** $x = -83784$;

(d) $x = -83785$; **(e)** $x = 83798$: **(f)** $x = 0.0013296$.

[1.2] Tomaram-se para valores aproximados de

$$x = 0.3000 \times 10^{-3}, \qquad y = 0.3000 \times 10^{1}, \qquad z = 0.3000 \times 10^{4},$$

respectivamente os valores

$$\tilde{x} = 0.3100 \times 10^{-3}, \qquad \tilde{y} = 0.3100 \times 10^{1}, \qquad \tilde{z} = 0.3100 \times 10^{4}.$$

Determine os respectivos erros absolutos e relativos, bem como as percentagens de erro. Comente sobre os valores obtidos.

[1.3] Considere os números $x = \pi$ e $y = 2199/700$.

(a) Obtenha aproximações $\tilde{x}$ e $\tilde{y}$ de x e y, respectivamente, num sistema de ponto flutuante com 4 dígitos na mantissa e arredondamento simétrico. Obtenha ainda $\tilde{z} = \mathrm{fl}(\tilde{x} - \tilde{y})$.

(b) Calcule os erros absolutos e relativos de $\tilde{x}$, $\tilde{y}$ e $\tilde{z}$, bem como as percentagens de erro. Comente.

(c) Com o objectivo de ilustrar a influência nos resultados da precisão utilizada repita as alíneas(a) e (b) usando um sistema de ponto flutuante com 6 dígitos na mantissa.

[1.4] Considere os valores

$$x = 0.100456683, \qquad y = 0.0995214437.$$

Determine o número de algarismos significativos que se pode garantir a

$$x \times y, \qquad x \div y, \qquad x + y, \qquad x - y,$$

[1] O asterisco * a seguir ao número do exercício indica "exercício recomendado".

ao efectuar as operações num sistema de ponto flutuante FP(10, 7, -38, 38) com arredondamento simétrico.

[1.5]* Considere os números $x = \pi$ e $y = 333/106$.

(a) Obtenha aproximações $\tilde{x}$ e $\tilde{y}$ para x e y, respectivamente, num sistema de ponto flutuante FP(10, 6, -10, 10) com arredondamento simétrico.

(b) Calcule os erros absolutos e relativos de $\tilde{x}$ e $\tilde{y}$.

(c) Calcule, efectuando as operações num sistema FP(10, 6, -10, 10) com arredondamento simétrico, valores aproximados das quantidades

$$x \times y, \qquad x \div y, \qquad x + y, \qquad x - y.$$

(d) Calcule os erros absolutos e relativos as quantidades calculadas na alínea anterior.

(e) Determine o número de algarismos significativos que se pode garantir a cada um das quantidades calculadas na alínea (c).

(f) Repita as alíneas (a) e (b) e a parte respeitante à quantidade $x - y$ das alíneas (c)-(d)-(e) considerando um sistema FP(10, 9, -10, 10) com arredondamento simétrico.

[1.6] Determine o erro absoluto cometido no cálculo do determinante da matriz

$$A = \begin{bmatrix} 5.7432 & 7.3315 \\ 6.5187 & 8.3215 \end{bmatrix}$$

se utilizar um sistema de ponto flutuante com 6 dígitos na mantissa e arredondamento simétrico.

[1.7] Considere os valores

$$A = 0.492, \quad B = 0.603, \quad C = -0.494, \quad D = -0.602, \quad E = 10^{-5}$$

Com a finalidade de calcular

$$F = \frac{A + B + C + D}{E},$$

dois indivíduos, usando uma máquina com 3 dígitos na mantissa e com arredondamento simétrico, efectuaram esse cálculo de forma distinta, mas aritmeticamente equivalente.

O indivíduo X calculou $A + B$, depois $C + D$, somou os valores, e dividiu por E, obtendo $F = 0$.

Por seu turno, indivíduo Y calculou $A+C$, depois $B+D$, somou os valores, e dividiu por E, obtendo $F = -100$.

Verifique os cálculos efectuados pelos dois indivíduos e comente a disparidade de resultados obtidos, atendendo a que se usaram processos matematicamente equivalentes.

[1.8] Sendo x e y números positivos considerados num sistema de ponto flutuante decimal tais que $x > y$ e

$$10^{-q} \leq 1 - \frac{y}{x} \leq 10^{-p},$$

mostre que pelo menos p e no máximo q dígitos significativos são perdidos ao efectuar a diferença $x - y$.

[1.9] Considere um sistema de ponto flutuante FP(10, 7, -38, 38) com arredondamento simétrico. Sendo $u = 0.5 \times 10^{-6}$ a unidade de arredondamento do sistema e $v = 0.9u$ calcule $\mathrm{fl}(1+u)$ e $\mathrm{fl}(1+v)$.

[1.10] Considere a função $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 1 - \cos x$, e os seguintes dois algoritmos para o cálculo de $z = f(x)$:

$$\begin{aligned} &(1)\ \ u = \cos x, \qquad z = 1 - u; \\ &(2)\ \ u_1 = \frac{x}{2}, \qquad u_2 = \sin u_1, \qquad u_3 = u_2^2, \qquad z = 2u_3. \end{aligned}$$

(a) Determine para que valores de x o cálculo de $f(x)$ conduz a um problema mal posto.

(b) Determine as expressões dos erros relativos dos dois algoritmos.

(c) Determine para que valores de x os algoritmos são numericamente instáveis.

[1.11] Sejam $\tilde{x}, \tilde{y}, \tilde{z}$ valores aproximados de x, y, z, respectivamente, com erros relativos $\delta_{\tilde{x}}, \delta_{\tilde{y}}, \delta_{\tilde{z}}$. Determine uma estimativa do erro relativo cometido no cálculo de $v = xy + z$ num sistema de ponto flutuante com unidade de arredondamento u e usando os valores aproximados.

[1.12]* Considere a função $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^2 - y^2$ e os três algoritmos seguintes para o cálculo de $z = f(x, y)$:

$$\begin{aligned} &(1)\ \ z = x \times x - y \times y; \\ &(2)\ \ z = (x + y) \times (x - y); \\ &(3)\ \ z = (x + y) \times x - (x + y) \times y. \end{aligned}$$

(a) Determine as expressões dos erros relativos dos três algoritmos.

(b) Supondo que x e y são representados exactamente no sistema de ponto flutuante utilizado, determine para cada algoritmo condições para as quais este algoritmo é *numericamente de mais confiança* que os outros.

[1.13] Considere o seguinte algoritmo para o cálculo de $z = \phi(x)$, $\phi : \mathbb{R}^n \to \mathbb{R}$):

$$\begin{cases} z = \psi(u), & \psi : \mathbb{R}^p \to \mathbb{R}, \qquad u = (u_1, u_2, \ldots, u_p) \\ u_i = \theta_i(x), & \theta_i : \mathbb{R}^n \to \mathbb{R}, \qquad i = 1, 2, \ldots, p, \end{cases}$$

onde $\theta_1, \theta_2, \ldots, \theta_p, \psi$ são $p+1$ funções elementares. Determine a expressão do erro relativo de $\tilde{z}$ em termos dos erros relativos das componentes de x e dos erros de arredondamento no cálculo dos valores das funções $\theta_1, \ldots, \theta_p, \psi$.

[1.14] Considere o método iterativo a um passo com função iteradora $g : \mathbb{R} \to \mathbb{R}$:

$$z_{n+1} = g(z_n), \quad n \in \mathbb{N}, \qquad z_0 \text{ dado.}$$

Determine o erro relativo de $\tilde{z}_{n+1}$ expresso em termos do erro relativo do valor inicial $\tilde{z}_0$ e dos erros relativos de arredondamento no cálculo dos sucessivos valores da função g.

[1.15]* Suponha que pretende calcular a soma de três números reais a, b, c, $S = a+b+c$, usando os dois seguintes algoritmos:

$$(1)\ S_1 = (a+b)+c; \qquad (2)\ S_2 = a+(b+c).$$

(a) Para

$$a = 0.33678429 \times 10^2, \qquad b = -0.33677811 \times 10^2, \qquad c = 0.23371258 \times 10^{-4},$$

calcule o valor exacto de S.

(b) Para os valores de a, b, c da alínea (a), e supondo que efectua os cálculos num sistema FP(10, 8, -10, 10), com arredondamento simétrico, calcule valores aproximados de S usando os dois algoritmos indicados.

(c) Determine os erros relativos dos valores obtidos na alínea (b).

(d) Determine a expressão do erro relativo do algoritmo (1) em termos dos erros relativos das parcelas e dos erros de arredondamento das duas operações. Utilize este resultado para concluir qual a ordem por que deve proceder à soma por forma a minimizar os efeitos dos erros de arredondamento.

[1.16]* Considere o polinómio definido por

$$f(x) = x^3 + ax^2 + bx + c$$

e os dois seguintes algoritmos para o cálculo de $f(x)$:

$$(1)\ \ f_1(x) = x \times (x \times x) + a \times (x \times x) + b \times x + c;$$
$$(2)\ \ f_2(x) = \big((x+a) \times x + b\big) \times x + c.$$

O algoritmo (2) é designado por algoritmo de Horner.

(a) Para $a = -6.1,\ b = 3.2,\ c = 1.5$, calcule o valor exacto de $f(4.71)$.

(b) Para $a = -6.1,\ b = 3.2,\ c = 1.5$, e supondo que efectua os cálculos no sistema FP(10, 3, -10, 10), com arredondamento simétrico, calcule valores aproximados de $f(4.71)$ usando os dois algoritmos indicados.

(c) Determine os erros relativos dos valores obtidos na alínea (b).

(d) Determine a expressão do erro relativo do algoritmo de Horner em termos dos erros relativos de a, b, c, x e dos erros de arredondamento das operações efectuadas.

[1.17]* Considere a equação quadrática

$$x^2 + 2bx + c = 0,$$

com coeficientes b e c reais positivos. Considere os dois seguintes algoritmos para o cálculo das raízes x_1 e x_2 da equação:

$$\begin{aligned} &(1)\ \ x_1 = -b - \sqrt{b^2 - c}, \qquad && x_2 = -b + \sqrt{b^2 - c}; \\ &(2)\ \ x_1 = -b - \sqrt{b^2 - c}, \qquad && x_2 = \frac{c}{x_1}. \end{aligned}$$

(a) Para $b = 34.56,\ c = 1$, verifique que as raízes têm os valores $x_1 = -69.105529\ldots$ e $x_2 = -0.014470622\ldots$

(b) Para $b = 34.56$, $c = 1$, e supondo que efectua os cálculos num sistema FP(10, 4, -10, 10), com arredondamento simétrico, obtenha valores aproximados para as raízes usando os algoritmos indicados.

(c) Determine os erros relativos dos valores obtidos na alínea (b).

(d) Determine as expressões dos erros relativos dos dois algoritmos indicados em termos dos erros relativos dos coeficientes b, c e dos erros de arredondamento das operações efectuadas. Suponha que a raiz quadrada é uma operação elementar.

[1.18] Considere o sistema linear

$$\begin{bmatrix} 10^{-6} & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix}.$$

(a) Resolva o sistema pelo método da eliminação de Gauss.

(b) Suponha que o sistema é resolvido numa calculadora onde os números são representados num sistema de ponto flutuante com 6 dígitos na mantissa. Que solução obteria nesse caso? Compare com a solução exacta.

(c) Suponha que o sistema é resolvido na mesma máquina, mas usando pesquisa parcial de pivot. Qual é o resultado nestas condições? Compare com o resultado da alínea anterior e comente.

[1.19] Considere o sistema linear

$$\begin{bmatrix} 1 & 10^6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0.5 \times 10^6 \\ 1.0 \end{bmatrix}.$$

(a) Verifique que este sistema é equivalente ao do exercício anterior.

(b) Será que, neste caso, a pesquisa parcial de pivot permite superar os efeitos dos erros de arredondamento, como acontecia no exercício anterior? Justifique.

(c) Resolva o sistema, utilizando o método da pesquisa total de pivot. Comente.

[1.20]* Considere o sistema linear

$$A \begin{bmatrix} x \\ y \end{bmatrix} = b,$$

onde A é uma matriz 2×2 não singular de elementos reais e b é um vector de $\mathbb{R}^2$, ambos supostos conhecidos.

(a) Para

$$A = \begin{bmatrix} 0.003000 & 59.14 \\ 5.291 & -6.130 \end{bmatrix}, \qquad b = \begin{bmatrix} 59.17 \\ 46.78 \end{bmatrix},$$

verifique que a solução exacta do sistema é $x = 10.00,\ y = 1.000$.

(b) Supondo que efectua os cálculos num sistema FP(10, 4, -10, 10), com arredondamento simétrico, determine as soluções aproximadas do sistema pelo método de eliminação de Gauss, sem e com pesquisa parcial de pivot.

(c) Determine os erros relativos das soluções aproximadas obtidas na alínea (b).

(d) Para A e b com componentes arbitrárias, sem erros inerentes e com representação exacta no sistema de ponto flutuante utilizado, determine a expressão dos erros relativos dos valores aproximados $\tilde{x}, \tilde{y}$ de x, y, obtidos pelo método de eliminação de Gauss, em termos dos erros de arredondamento das operações utilizadas.

[1.21] Considere o sistema linear

$$A \begin{bmatrix} x \\ y \\ z \end{bmatrix} = b,$$

onde

$$A = \begin{bmatrix} 10^{-6} & 0 & 1 \\ 1 & 10^{-6} & 2 \\ 1 & 2 & -1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}.$$

Representando os números com 6 dígitos na mantissa, resolva este sistema pelo método da eliminação de Gauss, sem e com pesquisa parcial de pivot. Compare os resultados e comente.