Capítulo 2

Erros e aritmética e computacional

2.1 Notação de vírgula flutuante ou científica

Um número real corresponde a uma sucessão de Cauchy de números racionais. Como tal, pode admitir várias representações, mas habitualmente tomamos como representante desse número uma sucessão de racionais que são múltiplos de uma potência de 10 (base decimal).

Definição 2.1.1. Um número real x diz-se representado em **notação de vírgula flutuante** (ou **ponto flutuante**), normalizado na base 10, quando escrito na forma

$$x = m \times 10^p,$$

onde $p \in \mathbb{Z}$, $m = 0$ se $x = 0$, e $0.1 \leq m < 1$ se $x \neq 0$.

Em notação de vírgula flutuante, um número representa-se na base decimal através do sinal + ou −, da mantissa m e do expoente p. Os algarismos da mantissa m variam entre 0 e 9, mas o primeiro algarismo deve ser igual de zero. A forma geral da representação de número em notação de vírgula flutuante é a seguinte,

$$x = \pm 0.a_1 a_2 \ldots a_n \times 10^p.$$

Exemplo 2.1.1. Os números seguintes escrevem-se em notação de vírgula flutuante como indicado:

(a) $425.7 = 0.4257 \times 10^3$;

(b) $-0.0258 = -0.258 \times 10^{-1}$;

(c) $0.35 = 0.35 \times 10^0$;

(d) $12.1(345) = 0.121(345) \times 10^2$.

Definição 2.1.2. Um número real x diz-se representado em **notação científica**, normalizado na base 10, quando escrito na forma

$$x = m \times 10^p,$$

onde $p \in \mathbb{Z}$, $m = 0$ se $x = 0$, e $1 \leq m \leq 9$ se $x \neq 0$.

De igual modo, em notação científica um número representa-se na base decimal através do sinal $+$ ou $-$, da mantissa m e do expoente p. Os algarismos da mantissa m variam entre 0 e 9, mas aqui o primeiro algarismo deve ser diferente de zero. A forma geral da representação de número em notação cinetífica é a seguinte,

$$x = \pm a_0.a_1a_2 \ldots a_n \times 10^p, \qquad a_0 \neq 0.$$

Exemplo 2.1.2. Os números seguintes escrevem-se em notação científica tal como indicado:

(a) $425.7 = 4.257 \times 10^2$;

(b) $-0.0258 = -2.58 \times 10^{-2}$;

(c) $3.5 = 3.5 \times 10^0$;

(d) $12.1(345) = 1.21(345) \times 10^1$.

Na prática, e a menos que estivéssemos na posse de uma máquina com memória infinita, a representação de um número, quer em notação de vírgula flutuante ou em notação científica, deve ser finita, pelo que somos obrigados a considerar um número finito de algarismos na mantissa e uma limitação nos valores dos expoentes admitidos.

2.2 Algarismos posicionais ou relevantes

Definição 2.2.1. Designam-se por **algarismos posicionais** de um número real x, com $0 < x < 1$, aos zeros usados para fixar a vírgula decimal na referida representação. No caso de $|x| \geq 1$ ou $x = 0$, a representação decimal de x não tem algarismos posicionais.

Exemplo 2.2.1. (1) Se $x = 0.004307$, os três primeiros zeros de x (**0.00**4307) são algarismos posicionais.

(2) Se $x = 43.0010$, temos $|x| \geq 1$, pelo que x não tem algarismos posicionais.

Definição 2.2.2. Na representação decimal de um número real x, todos os algarismos não posicionais de x designamos por **algarismos relevantes**.

Exemplo 2.2.2. (1) Se $x = 0.004307$, os três primeiros zeros de x são algarismos posicionais, enquanto os quatro últimos (0.00**4307**) são algarismos relevantes;

(2) Se $x = 43.0010$, todos são algarismos relevantes (**43.0010**).

O algarismo de maior ordem, ou seja, o algarismos colocado mais à esquerda na representação decimal, designa-se por **algarismos mais relevante**. Por sua vez, o algarismos de menor ordem, ou seja, o algarismos colocado mais à direita na representação decimal, designa-se por **algarismo menos relevante**.

Exemplo 2.2.3. (1) Se $x = 0.004307$, **4** é o algarismo mais relevante, enquanto **7** é o algarismo menos relevante.

(2) Se $x = 43.0010$, **4** é o algarismo mais relevante, e o **0** que se situa mais à direita é o menos relevante.

Quando se pretende arredondar um número para, por exemplo, 3 algarismos relevantes, subentende-se que o arredondamento a realizar é para os 3 algarismos mais relevantes.

2.3 Arredondamentos na base 10

No que se segue, consideremos números positivos. Na base decimal, pretendemos arredondar um número positivo x para a ordem 10^k, para determinado $k \in \mathbb{Z}$. Suponhamos que

$$x = a_n \dots a_1 a_0 a_{-1} a_{-2} \dots a_{-m} \dots .$$

No que se segue, quando falarmos de um arredondamento a p algarismos relevantes, pressupõe-se, no caso de existirem outros, que são os p algarismos mais relevantes.

Definição 2.3.1. No **arredondamento por defeito** à ordem 10^k, ou a $n - k$ algarismos relevantes, despreza-se o algarismo da ordem 10^{k-1} e todos os outros algarismos situados à sua direita. Esta técnica é também conhecida por *corte*. Se $k > 0$, multiplica-se o corte obtido por 10^k.

Na língua inglesa esta técnica é conhecida por *chopping*.

Exemplo 2.3.1. Arredondar os números seguintes por defeito a 3 algarismos relevantes:

(a) 27.3201;

(b) 45678.9012.

Resolução: (1) Para 3 algarismos relevantes, quer dizer que ficamos com algarismos relevantes somente nas posições a_1, a_0 e a_{-1}. Neste caso, $k = -1$ e desprezamos o algarismo da ordem

$10^{k-1} = 10^{-2}$, ou seja o algarismo a_{-2}, e todos os outros à sua direita. Assim,

$$27.3201 \simeq 27.3.$$

(2) Os três algarismos (mais) relevantes são 456 que correspondem às posições a_4, a_3 e a_2. Por isso, $k = 2$. Assim, desprezamos o algarismo da ordem $10^{k-1} = 10^1$ e todos os outros situados à sua direita. Como neste caso $k = 2$, multiplicamos o *corte* por $10^k = 10^2$ e o arredondamento por defeito será

$$45678.9012 = 456 \times 10^2 = 45600.$$

Definição 2.3.2. No **arredondamento por excesso** à ordem 10^k, ou a $n - k$ algarismos relevantes, realiza-se primeiro o arredondamento por defeito e ao número obtido adiciona-se 10^k.

Exemplo 2.3.2. Arredondar por excesso os números seguintes a 3 algarismos relevantes:

(a) 27.3201; (b) 45678.9012.

Resolução: (1) Como vimos no exemplo anterior, o arredondamento por defeito a 3 algarismos relevantes foi $27.3201 \simeq 27.3$, onde considerámos $k = -1$. Adicionado $10^k = 10^{-1}$, obtemos

$$27.3201 \simeq 27.4.$$

(2) Vimos acima que o arredondamento por defeito a 3 algarismos relevantes foi $45678.9012 = 456 \times 10^2 = 45600$, onde se tomou $k = 2$. Adicionado $10^k = 10^2$, obtemos

$$45678.90120 \simeq 45600 + 10^2 = 45700.$$

Definição 2.3.3. No **arredondamento simétrico** à ordem 10^k, ou a $n - k$ algarismos relevantes, procede-se do modo seguinte:

- Se $a_{k-1} < 5$, fazemos o arredondamos por defeito;
- Se $a_{k-1} \geq 5$, fazemos o arredondamos por excesso

Quando nada se disser sobre o tipo de arredondamento, subentende-se que estamos a falar do arredondamento usual, isto é, do arredondamento simétrico.

Exemplo 2.3.3. Arredondar de modo simétrico os números seguintes a 3 algarismos relevantes:

(a) 27.3201; (b) 45678.9012.

Resolução: (1) Como $a_{-1} = 3$ e $a_{-2} = 2 < 4$, arredondamos por defeito,

$$27.3201 \simeq 27.3.$$

(2) Neste caso, $a_2 = 6$ e $a_1 = 7 \geq 5$, arredondamos por excesso,

$$45678.90120 \simeq 45600 + 10^2 = 45700.$$

No **caso de números negativos**, o arredondamento à ordem 10^k, ou a $n-k$ algarismos relevantes, realiza-se do modo seguinte:

- Arredondamento por defeito: arredonda-se por excesso o valor absoluto do número e apresenta-se o resultado final com sinal negativo;
- Arredondamento por excesso: arredonda-se por defeito o valor absoluto do número e apresenta-se o resultado final com sinal negativo;
- Arredondamento simétrico: arredonda-se de modo simétrico o valor absoluto do número e apresenta-se o resultado final com sinal negativo.

Exemplo 2.3.4. Arredondar o número $x = -27.3201$ a 3 algarismos relevantes:

(a) por defeito; (b) por excesso; (c) de modo simétrico.

Resolução: (1) Temos que $|-27.3201| = 27.3201$, e arredondando por excesso vem $27.3201 \simeq 27.4$. O arredondamento por defeito é então

$$-27.3201 \simeq -27.4.$$

(2) Como $|-27.3201| = 27.3201$, arredondando por defeito vem $27.3201 \simeq 27.3$. O arredondamento por excesso é então

$$-27.3201 \simeq -27.3.$$

(3) Novamente, $|-27.3201| = 27.3201$, e, agora, o arredondando simétrico é $27.3201 \simeq 27.3$. Assim, o arredondamento simétrico é

$$-27.3201 \simeq -27.3.$$

2.4 Arredondamentos numa base $b \neq 10$

Consideremos um algarismo $b \geq 2$, mas tal que $b \neq 10$, e a base correspondente

$$\mathcal{B} = \{0, \ldots, b-1\}, \qquad \#\mathcal{B} = b.$$

Suponhamos primeiramente que o número a arredondar é positivo.

Definição 2.4.1. No **arredondamento por defeito** à ordem b^k, ou a $n-k$ algarismos relevantes, despreza-se o algarismo da ordem b^{k-1} e todos os outros algarismos situados à sua direita. Se $k > 0$, multiplica-se o corte obtido por b^k.

Exemplo 2.4.1. Arredondar por defeito os números seguintes a 3 algarismos relevantes:

(a) $2.022_{(3)}$;

(b) $432241_{(5)}$.

Resolução: (1) Para 3 algarismos relevantes, quer dizer que ficamos com algarismos relevantes somente nas posições a_0, a_{-1} e a_{-2}. Neste caso, $k = -2$ e desprezamos o algarismo da ordem $3^{k-1} = 3^{-3}$, ou seja o algarismo a_{-3}, e todos os outros à sua direita (se existissem). Assim,

$$2.022_{(3)} \simeq 2.02_{(3)}.$$

(2) Os três algarismos (mais) relevantes são 432 que correspondem às posições a_5, a_4 e a_3. Por isso, $k = 3$. Assim, desprezamos o algarismo da ordem $5^{k-1} = b^2$ e todos os outros situados à sua direita. Como neste caso $k = 3 > 0$, multiplicamos o *corte* por $5^k = 5^3$ e o arredondamento por defeito será

$$432241_{(5)} \simeq 432 \times 5^3 = 432000_{(5)}.$$

Definição 2.4.2. No **arredondamento por excesso** à ordem b^k, ou a $n-k$ algarismos relevantes, realiza-se primeiro o arredondamento por defeito e ao número obtido adiciona-se b^k.

Exemplo 2.4.2. Arredondar por excesso os números seguintes a 3 algarismos relevantes:

(a) $2.022_{(3)}$;

(b) $432241_{(5)}$.

Resolução: (1) Vimos acima que o arredondamento por defeito, a 3 algarismos relevantes, à ordem 3^{-2} é dado por $2.022_{(3)} \simeq 2.02_{(3)}$. Adicionando $3^k = 3^{-2}$, temos

$$\begin{aligned} 2.022_{(3)} \simeq & 2.02_{(3)} + b^{-2} = 2 \times 3^0 + 0 \times 3^{-1} + 2 \times 3^{-2} + 1 \times 3^{-2} \\ = & 2 \times 3^0 + 0 \times 3^{-1} + (2+1) \times 3^{-2} \\ = & 2 \times 3^0 + 1 \times 3^{-1} + 0 \times 3^{-2} = 2.10_{(3)}. \end{aligned}$$

(2) Neste caso, tal como vimos acima, o arredondamento por defeito, a 3 algarismos relevantes, à ordem 5^3 foi dado por $432241_{(5)} \simeq 432000_{(5)}$. Como $k = 3$, adicionamos 5^3 e temos

$$\begin{aligned} 432241_{(5)} \simeq & 432000_{(5)} + 5^3 = 4 \times 5^5 + 3 \times 5^4 + 2 \times 5^3 + 1 \times 5^3 \\ = & 4 \times 5^5 + 3 \times 5^4 + (2+1) \times 5^3 = 4 \times 5^5 + 3 \times 5^4 + 3 \times 5^3 = 433000_{(5)}. \end{aligned}$$

Definição 2.4.3. No **arredondamento simétrico** à ordem b^k, ou a $n-k$ algarismos relevantes, procede-se do modo seguinte:

- Se $a_{k-1} < \frac{b}{2}$, arredondamos por defeito;
- Se $a_{k-1} \geq \frac{b}{2}$, arredondamos por excesso

Exemplo 2.4.3. Arredondar de modo simétrico os números seguintes a 3 algarismos relevantes:

(a) $2.022_{(3)}$; (b) $432241_{(5)}$.

Resolução: (1) Neste caso $k = -1$ e $a_{k-1} = a_{-2} = 2 \geq \frac{b}{2} = \frac{3}{2}$, pelo que arredondamos por excesso,

$$2.022_{(3)} \simeq 2.10_{(3)}.$$

(2) Aqui, $k = 3$ e $a_{k-1} = a_2 = 2 < \frac{b}{2} = \frac{5}{2}$. Arredondando, então, por defeito, temos

$$432241_{(5)} \simeq 432 \times 5^3 = 432000_{(5)}.$$

O caso de números negativos é completamente análogo ao da base decimal.

Exemplo 2.4.4. Arredondar o número $x = -1.2111_{(3)}$ a 3 algarismos relevantes:

(a) por defeito; (b) por excesso; (c) de modo simétrico.

Resolução: (1) Temos que $|-1.2111| = 1.2111$, e arredondando por excesso vem $1.2111 \simeq$

$1.21 + 3^{-2} = 1.22$. Então, o arredondamento por defeito é

$$-1.2111 \simeq -1.22.$$

(2) Como $|-1.2111| = 1.2111$, arredondando por defeito vem $1.2111 \simeq 1.21$. O arredondamento por excesso é então

$$-1.2111 \simeq -1.21.$$

(3) Novamente, $|-1.2111| = 1.2111$, e, agora, o arredondando simétrico é $1.2111 \simeq 1.21$. Assim, o arredondamento simétrico é

$$-1.2111 \simeq -1.21.$$

Tal como no caso da base decimal, quando nada se disser sobre o tipo de arredondamento, subentende-se que estamos a falar do arredondamento usual, isto é, do arredondamento simétrico.

2.5 Introdução aos erros

No que se segue, designemos por:

$$\begin{aligned} x &: \text{ valor exacto de um número real;} \\ \overline{x} &: \text{ um valor aproximado de } x. \end{aligned}$$

Definição 2.5.1. (1) O **erro da aproximação** $\overline{x}$ de um número x é dado por

$$E = x - \overline{x}.$$

(2) O **erro absoluto** cometido na aproximação $\overline{x}$ de um número x é dado por

$$E_a = |x - \overline{x}|.$$

(3) Um majorante do **erro absoluto** cometido na aproximação $\overline{x}$ de um número x é um número real Δx tal que

$$E_a \leq \Delta x \Leftrightarrow |x - \overline{x}| \leq \Delta x.$$

Como é óbvio, na majoração do erro absoluto interessa-nos o menor dos majorantes Δx (supremo) que satisfaz $E_a \leq \Delta x$.

Exemplo 2.5.1. Considere o arredondamento simétrico à ordem 10^{-3} da dízima infinita periódica $x = 0.(6)$ dado por

$$\overline{x} = 0.667.$$

Neste caso, o erro e o erro absoluto são dados por:

$$E = 0.(6) - 0.667 = \frac{2}{3} - \frac{667}{1000} = -\frac{1}{3000};$$
$$E_a = |E| = |0.(6) - 0.667| = \left|\frac{2}{3} - \frac{667}{1000}\right| = \left|-\frac{1}{3000}\right| = \frac{1}{3000}.$$

Qualquer valor (finito) não inferior a $\frac{1}{3000}$ é um majorante do erro absoluto. Por exemplo,

$$E_a = \frac{1}{3000} \leq \frac{3}{3000} = 0.001.$$

No arredondamento $\overline{\pi} = 3.14$ do número $\pi = 3.14159265\ldots$, qualquer um dos valores

$$0.0016; \qquad 0.002; \qquad 0.5$$

é um majorante do erro absoluto da aproximação $\overline{\pi}$ considerada, já que

$$\begin{aligned} E_a = |\pi - 3.14| = 0.00159265\ldots &\leq 0.0016 \\ &\leq 0.002 \\ &\leq 0.5. \end{aligned}$$

Para estes majorantes do erro, podemos escrever:

$$\begin{aligned} &\text{com um algarismo relevante em } \Delta x \ : \ x = 3.14 \pm 0.002; \\ &\text{com dois algarismos relevantes em } \Delta x \ : \ x = 3.14 \pm 0.0016. \end{aligned}$$

No exemplo anterior, observe-se que $\Delta\pi = 0.00159$ não é um majorante do erro absoluto, já que $E_a = |\pi - 3.14| = 0.00159265\cdots > 0.00159$. Por este motivo, os majorantes do erro absoluto deverão ser sempre obtidos usando um arredondamento por excesso.

Definição 2.5.2. O erro relativo da aproximação $\overline{x}$ de um número x é dado por

$$E_r = \frac{E_a}{|x|}, \qquad x \neq 0.$$

Se x não for conhecido, uma aproximação do erro relativo é dada por

$$E_r = \frac{E_a}{|\overline{x}|}, \qquad \overline{x} \neq 0.$$

Em qualquer dos casos, o majorante do erro relativo é dado por

$$\varepsilon = \frac{\Delta x}{m}, \qquad m > 0,$$

onde

$$m := \begin{cases} \min\left\{|x| : x \in [\overline{x} - \Delta x, \overline{x} + \Delta x]\right\}, & \text{se } x \text{ não é conhecido,} \\ |x|, & \text{se } x \text{ é conhecido.} \end{cases}$$

Exemplo 2.5.2. Se $x = 73 \pm 2$, então 73 é uma aproximação de x e 2 é um majorante do erro absoluto da aproximação,

$$\overline{x} = 73, \qquad \Delta x = 2.$$

Uma estimativa para o erro relativo do valor aproximado é dada por

$$E_r \simeq \frac{\Delta x}{|\overline{x}|} = \frac{2}{73} \simeq 0.027 = 2.7\%.$$

Para o real valor de x, sabemos apenas que

$$x \in [73 - 2.73 + 2] = [71.75].$$

Assim,

$$m = \min\Big\{|x| : x \in [71.75]\Big\} = 71,$$

e um majorante para o erro relativo é então dado por

$$\varepsilon = \frac{\Delta x}{m} = \frac{2}{71} \simeq 0.029 = 2.9\%.$$

Na proposição seguinte caracterizam-se os erros dos arredondamentos.

Proposição 2.5.1. (1) Um majorante do erro absoluto do arredondamento, quer realizado por defeito ou excesso, é 1 unidade da última ordem decimal do valor arredondado;

(2) Um majorante do erro absoluto do arredondamento simétrico é meia unidade da última ordem decimal do valor arredondado.

Exemplo 2.5.3. Consideremos o arredondamento $\overline{x} = 32.7$ de um número real x.

(1) Se $\overline{x} = 32.7$ for um arredondamento por defeito, então $x \in [32.7, 32.8]$ e um majorante do erro absoluto do arredondamento $\overline{x}$ é 1 unidade da última casa decimal arredondada, neste caso 10^{-1},

$$\Delta x = |32.8 - 32.7| = 0.1 = 1 \times 10^{-1}.$$

(2) Se $\overline{x} = 32.7$ for um arredondamento por excesso, então $x \in [32.6, 32.7]$ e um majorante do erro absoluto do arredondamento $\overline{x}$ é 1 unidade da última casa decimal arredondada, neste caso 10^{-1},

$$\Delta x = |32.7 - 32.6| = 0.1 = 1 \times 10^{-1}.$$

(3) Já se $\overline{x} = 32.7$ for um arredondamento simétrico, então $x \in [32.65, 32.75)$ e um majorante do erro absoluto do arredondamento $\overline{x}$ é meia unidade da última casa decimal arredondada, neste caso 10^{-1},

$$\Delta x = \frac{1}{2} \times |32.75 - 32.65| = \frac{1}{2} \times 0.1 = 0.05 = \frac{1}{2} \times 10^{-1}.$$

Concluímos esta secção abordando os erros de truncatura.

Definição 2.5.3. Entende-se por erro de truncatura qualquer erro introduzido quando se substitui um processo infinito por um processo finito, ou um processo contínuo por um processo discreto.

O exemplo seguinte é sobre a truncatura de uma soma infinita.

Exemplo 2.5.4. Consideremos a série seguinte e a sua soma (número de Euler e),

$$\sum_{n=0}^{\infty} \frac{1}{n!} = e\,.$$

Nos cálculos que habitualmente realizamos com o número de Euler, aproximamos o seu valor pela soma das 5 primeiras parcelas

$$\sum_{n=0}^{\infty} \frac{1}{n!} \simeq \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} = \frac{65}{24} = 2.708(3) \simeq 2.71\,.$$

Nestes cálculos, estamos a cometer um erro de truncatura

$$E_t = \sum_{n=0}^{\infty} \frac{1}{n!} - \sum_{n=0}^{4} \frac{1}{n!} = \sum_{n=5}^{\infty} \frac{1}{n!} \leq 0.00995\,.$$

O próximo exemplo é sobre a truncatura de um integral.

Exemplo 2.5.5. Consideremos o integral

$$\int_0^1 \sqrt{1+x^4}\,dx\,.$$

Por impossibilidade de cálculo da primitiva da função integranda (função contínua), somos forçados a considerar um processo de aproximação. Podemos aproximar o valor do integral dado pela soma de Riemann seguinte

$$\sum_{n=1}^{4} f(\xi_n)\Delta x_n = f\left(\frac{1}{8}\right) \times \frac{1}{4} + f\left(\frac{3}{8}\right) \times \frac{1}{4} + f\left(\frac{5}{8}\right) \times \frac{1}{4} + f\left(\frac{7}{8}\right) \times \frac{1}{4} \simeq 1.0858\,.$$

Sabendo que o valor exacto do integral é dado por $1.089429\ldots$, nestes cálculos cometemos um erro de truncatura

$$E_t = \int_0^1 \sqrt{1+x^4}\,dx - \sum_{n=1}^{4} f(\xi_n)\Delta x_n \leq 0.0037\,.$$

2.6 Algarismos significativos

Sejam:

- $\overline{x} \neq 0$ uma aproximação de um número real x;
- Δx um majorante do erro absoluto da aproximação $\overline{x}$;
- a_i um algarismo relevante da representação decimal de $\overline{x}$.

Definição 2.6.1. O algarismo relevante a_i é algarismo significativo de valor aproximado $\overline{x}$, se Δx não exceder meia unidade da ordem decimal do algarismo a_i em $\overline{x}$.

Na proposição seguinte apresentamos uma primeira caracterização dos algarismos significativos.

Proposição 2.6.1. Sejam $k \in \mathbb{Z}$ e Δx um majorante do erro absoluto da aproximação $\overline{x}$ de um número real x.

(1) Se $\Delta x \leq 0.5 \times 10^k$, então o algarismo relevante da ordem 10^k é algarismo significativo;

(2) Se $\Delta x > 0.5 \times 10^k$, então o algarismo relevante da ordem 10^k não é algarismo significativo.

Exemplo 2.6.1. Para o número dado por

$$x = 1.03789 \pm 0.004$$

determinar os algarismos significativos da aproximação $\overline{x} = 1.03789$.

Resolução: Todos os algarismo de $\overline{x} = 1.03789$ são relevantes.

(1) $a_0 = 1$ é algarismo significativo, porque $\Delta x = 0.004 \leq 0.5 \times 10^0 = 0.5$;

(2) $a_{-1} = 0$ é algarismo significativo, porque $\Delta x = 0.004 \leq 0.5 \times 10^{-1} = 0.05$;

(3) $a_{-2} = 3$ é algarismo significativo, porque $\Delta x = 0.004 \leq 0.5 \times 10^{-2} = 0.005$;

(4) $a_{-3} = 7$ não é algarismo significativo, porque $\Delta x = 0.004 > 0.5 \times 10^{-3} = 0.0005$.

Todos os algarismos (relevantes) de ordem inferior a 10^{-3}, ou sejam $a_{-4} = 8$ e $a_{-5} = 9$, não são algarismos significativos.

Portanto, $\overline{x} = 1.03789$ tem 6 algarismos relevantes, mas apenas 3 algarismos significativos (que são os mais relevantes).

Na proposição seguinte apresentamos outra caracterização dos algarismos significativos. No que se segue designamos os algarismos de Δx por d_i.

Proposição 2.6.2. Consideremos um número real dado por

$$x = \overline{x} \pm \Delta x.$$

(1) Se Δx tem apenas um algarismo relevante d_k (na ordem 10^k), então:

(i) Se $d_k \leq 5$, o algarismo de $\overline{x}$ da ordem 10^{k+1} é o menos significativo;

(ii) Se $d_k > 5$, o algarismo de $\overline{x}$ da ordem 10^{k+2} é o menos significativo;

(2) Se Δx tem mais do que um algarismo relevante, arredonda-se Δx por excesso a um algarismo relevante apenas e procede-se como em (1)

Na proposição anterior está subentendido que se a_k for o algarismo menos significativo, então todos os algarismos de ordem superior são também significativos, tanto mais significativos quanto maior for a ordem correspondente.

Exemplo 2.6.2. Indique quais os algarismos significativos das aproximações $\overline{x}$ dos números reais representados por

(a) $x = 6.0004321 \pm 0.0002$;

(b) $x = 6.0004321 \pm 0.00051$.

Resolução: (a) Temos aqui $\Delta x = 0.0002$ tem apenas um algarismo relevante, neste caso na ordem 10^{-4} : $d_{-4} = 2$. Como $d_{-4} = 2 \leq 5$, o algarismo menos significativo de $\overline{x} = 6.0004321$ é o da ordem $10^{k+1} = 10^{-4+1} = 10^{-3}$: $a_{-3} = 0$.

Por consequência, $\overline{x} = 6.0004321$ tem 4 algarismos significativos, neste caso $a_0 = 6$, $a_{-1} = 0$, $a_{-2} = 0$ e $a_{-3} = 0$:

$$x = \mathbf{6.000}4321 \pm 0.0002$$

(b) Neste caso, $\Delta x = 0.00051$ tem dois algarismos relevantes, pelo que arredondamos por excesso a apenas um algarismo relevante: $\Delta x \simeq 0.0006$. Para esta aproximação de Δx, o único algarismo relevante é o da ordem 10^{-4} : $d_{-4} = 6$. Como $d_{-4} = 6 > 5$, o algarismo menos significativo de $\overline{x}$ é o da ordem $10^{k+2} = 10^{-4+2} = 10^{-2}$: $a_{-2} = 0$.

Por consequência, $\overline{x} = 6.0004321$ tem 4 algarismos significativos, neste caso $a_0 = 6$, $a_{-1} = 0$ e $a_{-2} = 0$:

$$x = \mathbf{6.00}04321 \pm 0.00051$$

Os resultados anteriores contidos nesta secção podem facilmente generalizar-se a uma base não decimal qualquer

$$\mathcal{B} = \{0, \ldots, b-1\}, \qquad \#\mathcal{B} = b \geq 2.$$

Na proposição que se segue, consideremos um número real arbitrário x representado por

$$x_{(b)} = \overline{x} \pm \Delta x$$

numa base $\mathcal{B}$.

Proposição 2.6.3. Sejam $k \in \mathbb{Z}$ e Δx um majorante do erro absoluto da aproximação $\overline{x}$ de um número real x escrito na base $\mathcal{B}$.

(1) Se $\Delta x \leq \frac{1}{2} \times b^k$, então o algarismo relevante da ordem b^k é algarismo significativo;

(2) Se $\Delta x > \frac{1}{2} \times b^k$, então o algarismo relevante da ordem b^k não é algarismo significativo.

Tal como no caso da base decimal, a proposição anterior é por vezes mais fácil de aplicar.

Proposição 2.6.4. Consideremos um número real dado por

$$x_{(b)} = \overline{x} \pm \Delta x.$$

(1) Se Δx tem apenas um algarismo relevante d_k (na ordem b^k), então:

(i) Se $d_k \leq \frac{b}{2}$, o algarismo de $\overline{x}$ da ordem b^{k+1} é o menos significativo;

(ii) Se $d_k > \frac{b}{2}$, o algarismo de $\overline{x}$ da ordem b^{k+2} é o menos significativo;

(2) Se Δx tem mais do que um algarismo relevante, arredonda-se Δx por excesso a um algarismo relevante apenas e procede-se como em (1).

Exemplo 2.6.3. Para o número dado por

$$x_{(7)} = 0.123 \pm 0.003$$

determinar os algarismos significativos da aproximação $\overline{x} = 0.123_{(7)}$.

Resolução: Temos aqui $\Delta x = 0.003$ o qual tem apenas um algarismo relevante, neste caso na ordem $7^{-3} : d_{-3} = 3$. Como $d_{-3} = 3 \leq \frac{b}{2} = \frac{7}{2}$, o algarismo menos significativo de $\overline{x} = 0.123_{(7)}$ é o da ordem $b^{k+1} = 7^{-3+1} = 7^{-2} : a_{-2} = 2$.

Por consequência, $\overline{x} = 0.123_{(7)}$ tem 2 algarismos significativos, neste caso $a_{-1} = 1$ e $a_{-2} = 2$:

$$x = 0.\mathbf{12}3 \pm 0.003$$

Voltemos à base decimal.

Observação. De ora em diante, e sempre que não seja dito nada em contrário, quando se apresentar uma aproximação $\overline{x}$, sem indicação explícita do majorante do erro absoluto Δx, devemos assumir que todos os algarismos relevantes de $\overline{x}$ são significativos.

Por exemplo, se considerarmos a aproximação

$$\overline{e} = 2.718$$

do número de Euler

$$e = 2.718281828\ldots,$$

está subentendida a representação de e na forma

$$e = 2.718 \pm 0.0005\,.$$

Na verdade, para Δe poderíamos considerar qualquer valor $\Delta e = 0.000p$, com $1 \leq p \leq 5$. De facto, teríamos sempre $d_{-4} = p \leq 5$, pelo que o algarismo menos significativo de $\overline{e}$ é o da ordem $10^{k+1} = 10^{-4+1} = 10^{-3}$: $a_{-3} = 8$.

Na proposição seguinte apresentamos uma relação entre o erro relativo de um valor aproximado $\overline{x}$ de um número real x e o número de algarismo significativo.

Proposição 2.6.5. Seja $\overline{x}$ um valor aproximado (em base 10) de um número real x com n algarismos significativos. Então o erro relativo da aproximação $\overline{x}$ é majorado da forma seguinte

$$E_r \leq 0.5 \times 10^{1-n}.$$

Exemplo 2.6.4. Seja $\overline{x} = 24.76$ uma aproximação (com 4 algarismos significativos) de um número real x na base 10. Está então subentendido que $\Delta x = 0.5 \times 10^{-3}$, pelo que uma aproximação (por excesso) para o erro relativo é dada por

$$E_r \simeq \frac{E_a}{\overline{x}} = \frac{|\Delta x|}{\overline{x}} = \frac{0.005}{24.76} \simeq 0.00021 = 0.21\%.$$

E pela proposição anterior,

$$E_r \leq 0.5 \times 10^{1-n} = 0.5 \times 10^{1-4} = 0.5 \times 10^{-3} = 0.0005 = 0.05\%.$$

Observe-se que, de facto, $0.0005 \geq 0.00021$.

A proposição anterior generaliza-se a uma base arbitrária b do modo seguinte.

Proposição 2.6.6. Seja $\overline{x}$ um valor aproximado (numa base b qualquer) de um número real x com n algarismos significativos. Então o erro relativo da aproximação $\overline{x}$ é majorado da forma seguinte:

$$E_r \leq \frac{1}{2} \times b^{1-n}.$$

Exemplo 2.6.5. Seja $\overline{x} = 2B9.A$ uma aproximação (com 4 algarismos significativos) de um número real x na base 13. Está então subentendido que $\Delta x = \frac{1}{2} \times 13^{-3}$, pelo que uma aproximação (por excesso) para o erro relativo é dada por

$$E_r \simeq \frac{E_a}{\overline{x}} = \frac{|\Delta x|}{\overline{x}} = \frac{\frac{1}{2} \times 13^{-3}}{2B9.A_{(13)}}.$$

Convertendo para base 10 para facilitar a compreensão:

$$2B9.A_{(13)} = 2 \cdot 13^2 + 11 \cdot 13^1 + 9 \cdot 13^0 + 10 \cdot 13^{-1} = 338 + 143 + 9 + 0.769 \simeq 490.769_{(10)}.$$

O erro absoluto:

$$\frac{1}{2} \times 13^{-3} = 0.5 \cdot \frac{1}{2197} \simeq 0.0002275.$$

Logo o erro relativo aproximado:

$$E_r \simeq \frac{0.0002275}{490.769} \simeq 4.64 \times 10^{-7} = 0.000046\%.$$

Pela proposição anterior (para base 13),

$$E_r \leq \frac{1}{2} \times 13^{1-4} = \frac{1}{2} \times 13^{-3} = 0.0002275 \simeq 0.02275\%.$$

Observe-se que, de facto, $0.0002275 \geq 4.64 \times 10^{-7}$.

2.7 Propagação de erros

Consideremos primeiramente o caso de propagação de erros em funções reais de uma variável real.

Proposição 2.7.1. Consideremos uma função real de variável real $f : \mathbb{D} \subseteq \mathbb{R} \longmapsto \mathbb{R}$, a qual supomos ser diferenciável em $\mathbb{D}$. Sejam $x \in \mathbb{D}$, $\overline{x}$ um valor aproximado de x e Δx um majorante da aproximação $\overline{x}$ tais que

$$[\overline{x} - \Delta x, \overline{x} + \Delta x] \subset \mathbb{D}\,.$$

Uma aproximação da propagação do erro absoluto por meio da função f é dada por

$$E_y \simeq |f'(x)| E_x\,.$$

Uma majoração da propagação do erro absoluto é dada por

$$E_y \leq \max_{x \in [\overline{x} - \Delta x, \overline{x} + \Delta x]} |f'(x)| E_x\,.$$

O resultado da proposição anterior significa que um majorante do erro de arredondamento da variável x é propagado para a variável $y = f(x)$ por um factor multiplicativo relacionado com a derivada da função numa vizinhança de $\overline{x}$.

Exemplo 2.7.1. Estude a propagação do erro da aproximação $\overline{x} = 2.7$ por meio da função $f(x) = 3x^2 + 1$.

Resolução: Considerando significativos todos os algarismos de $\overline{x}$, temos

$$x = 2.7 \pm 0.05 \Rightarrow \Delta x = 0.05\,.$$

Um valor aproximado de $y = f(x)$ é dado por

$$\overline{y} = 3 \times 2.7^2 + 1 = 22.87\,.$$

Uma aproximação do erro absoluto propagado é dada por

$$E_y \simeq |f'(\overline{x})|E_x = |6 \times 2.7| \times |x - \overline{x}| = 16.2 \times 0.05 = 0.81\,.$$

Um majorante do erro propagado da aproximação $\overline{y} = 22.87$ é dado por

$$\begin{aligned} E_y = & \max_{x \in [\overline{x}-\Delta x, \overline{x}+\Delta x]} |f'(x)|\Delta x = \max_{x \in [2.7-0.05, 2.7+0.05]} |6x| \times 0.05 \\ = & 6 \times (2.7 + 0.05) \times 0.05 = 0.825 \leq 0.83\,. \end{aligned}$$

Deste modo,

$$x = 2.7 \pm 0.05 \quad \Rightarrow \quad y = 22.87 \pm 0.83\,.$$

O resultado anterior pode ser extendido a funções reais com mais de uma variável real.

Proposição 2.7.2. Consideremos uma função real de duas variáveis reais $f : \mathbb{D} \subseteq \mathbb{R}^2 \longmapsto \mathbb{R}$, a qual supomos ser diferenciável nas variáveis x e y no domínio $\mathbb{D}$. Sejam $(x, y) \in \mathbb{D}$, $(\overline{x}, \overline{y})$ um valor aproximado de (x, y) e Δx, Δy majorantes das aproximações $\overline{x}$ e $\overline{x}$ tais que

$$[\overline{x} - \Delta x, \overline{x} + \Delta x] \times [\overline{y} - \Delta y, \overline{y} + \Delta y] \subset \mathbb{D}\,.$$

Uma aproximação da propagação do erro absoluto por meio da função f é dada por

$$E_z \simeq \left|\frac{\partial f}{\partial x}(\overline{x}, \overline{y})\right| E_x + \left|\frac{\partial f}{\partial y}(\overline{x}, \overline{y})\right| E_y\,.$$

Uma majoração da propagação do erro absoluto é dada por

$$E_z \leq \max_{\substack{x \in [\overline{x} - \Delta x, \overline{x} + \Delta x] \\ y \in [\overline{y} - \Delta y, \overline{y} + \Delta y]}} \left|\frac{\partial f}{\partial x}(x, y)\right| E_x + \max_{\substack{x \in [\overline{x} - \Delta x, \overline{x} + \Delta x] \\ y \in [\overline{y} - \Delta y, \overline{y} + \Delta y]}} \left|\frac{\partial f}{\partial y}(x, y)\right| E_y$$

Exemplo 2.7.2. Estude a propagação do erro da aproximação $(\overline{x}, \overline{y})$ por meio da função $f(x, y) = xy - x^2 + 3$, onde

$$x = 2.3 \pm 0.02, \qquad y = 15.2 \pm 0.3\,.$$

Resolução: Neste caso, temos

$$\overline{x} = 2.3, \qquad \Delta x = 0.02\,,$$
$$\overline{y} = 15.2, \qquad \Delta y = 0.3\,.$$

Um valor aproximado de $\overline{z} = f(\overline{x}, \overline{y})$ é dado por

$$\overline{z} = f(\overline{x}, \overline{y}) = 2.3 \times 15.2 - 2.3^2 + 3 = 32.67\,.$$

A aproximação da propagação do erro absoluto por meio da função $z = f(x, y)$ é dada por

$$\begin{aligned} E_z &\simeq \left|\frac{\partial f}{\partial x}(\overline{x}, \overline{y})\right| E_x + \left|\frac{\partial f}{\partial y}(\overline{x}, \overline{y})\right| E_y \\ &= |15.3 - 2 \times 2.3|\,|2.3 \pm 0.02 - 2.3| + |2.3|\,|15.2 \pm 0.3 - 15.2| = 0.902\,. \end{aligned}$$

Uma majoração da propagação do erro absoluto é dada por

$$\begin{aligned} E_z \leq & \max_{\substack{x \,\in\, [\overline{x} - \Delta x, \overline{x} + \Delta x] \\ y \,\in\, [\overline{y} - \Delta y, \overline{y} + \Delta y]}} \left|\frac{\partial f}{\partial x}(x, y)\right| E_x + \max_{\substack{x \,\in\, [\overline{x} - \Delta x, \overline{x} + \Delta x] \\ y \,\in\, [\overline{y} - \Delta y, \overline{y} + \Delta y]}} \left|\frac{\partial f}{\partial y}(x, y)\right| E_y \\ = & \max_{\substack{x \,\in\, [2.3 - 0.02, 2.3 + 0.02] \\ y \,\in\, [15.2 - 0.3, 15.2 + 0.3]}} |y - 2x|\,|2.3 \pm 0.02 - 2.3| + \max_{\substack{x \,\in\, [2.3 - 0.02, 2.3 + 0.02] \\ y \,\in\, [15.2 - 0.3, 15.2 + 0.3]}} |x|\,|15.2 \pm 0.3 - 15.2| \\ = & |15.2 + 0.3 - 2 \times (2.3 - 0.02)| \times 0.02 + |2.3 + 0.02| \times 0.3 \\ = & 0.9148 \leq 0.92 = \Delta z\,. \end{aligned}$$