Cálculo Numérico
Módulo III

Erros

Profs.: Bruno Correia da Nóbrega Queiroz
José Eustáquio Rangel de Queiroz
Marcelo Alves de Barros

Erros - Roteiro

- Existência
- Tipos
- Propagação

Erros - Existência I

- **Premissa**
  - **Impossibilidade de obtenção de soluções analíticas para vários problemas de Engenharia.**
  - **Consequência**
    - **Emprego de métodos numéricos na resolução de inúmeros problemas do mundo real.**

Erros - Existência II

- **Erro Inerente**

Erro <u>sempre</u> presente nas soluções numéricas, devido à incerteza sobre o valor real.

Ex. 01: Representação intervalar de dados

(50,3 ± 0,2) cm
(1,57 ± 0,003) ml
(110,276 ± 1,04) Kg

Cada medida é um intervalo e não um número.

Erros - Existência III

- **Método Numérico**

Método adotado na resolução de um problema físico, mediante a execução de uma sequência finita de operações aritméticas.

- **Consequência**

  - **Obtenção de um resultado aproximado, cuja diferença do resultado esperado (exato) denomina-se *erro*.**

Erros - Existência IV

- **Natureza dos Erros I**

  - **Erros inerentes ao *processo de aquisição dos dados***

    - **Relativos à imprecisão no processo de aquisição/entrada, externos ao processo numérico.**

Erros Inerentes aos Dados

- **Proveniência ⇨ Processo de *aquisição*/ *entrada* (medidas experimentais)**

  - **Sujeitos às limitações/aferição dos instrumentos usados no processo de mensuração**

  - **Erros *inerentes* são inevitáveis!**

Erros - Existência V

- **Natureza dos Erros II**
  - **Erros inerentes ao *modelo matemático* adotado**
    - **Relativos à impossibilidade de representação exata dos fenômenos reais a partir de modelos matemáticos**
      - **Necessidade de adotar condições que simplifiquem o problema, a fim de torná-lo numericamente solúvel**

Erros Inerentes ao Modelo

- **Proveniência ⇨ Processo de *modelagem* do problema**

  - **Modelos matemáticos raramente oferecem representações exatas dos fenômenos reais**

  - **Equações e relações, assim como dados e parâmetros associados, costumam ser simplificados**

    - **Factibilidade e viabilidade das soluções**

Erros - Existência VII

- **Natureza dos Erros III**
  - **Erros de *truncamento***
    - **Substituição de um processo infinito de operações por outro finito**

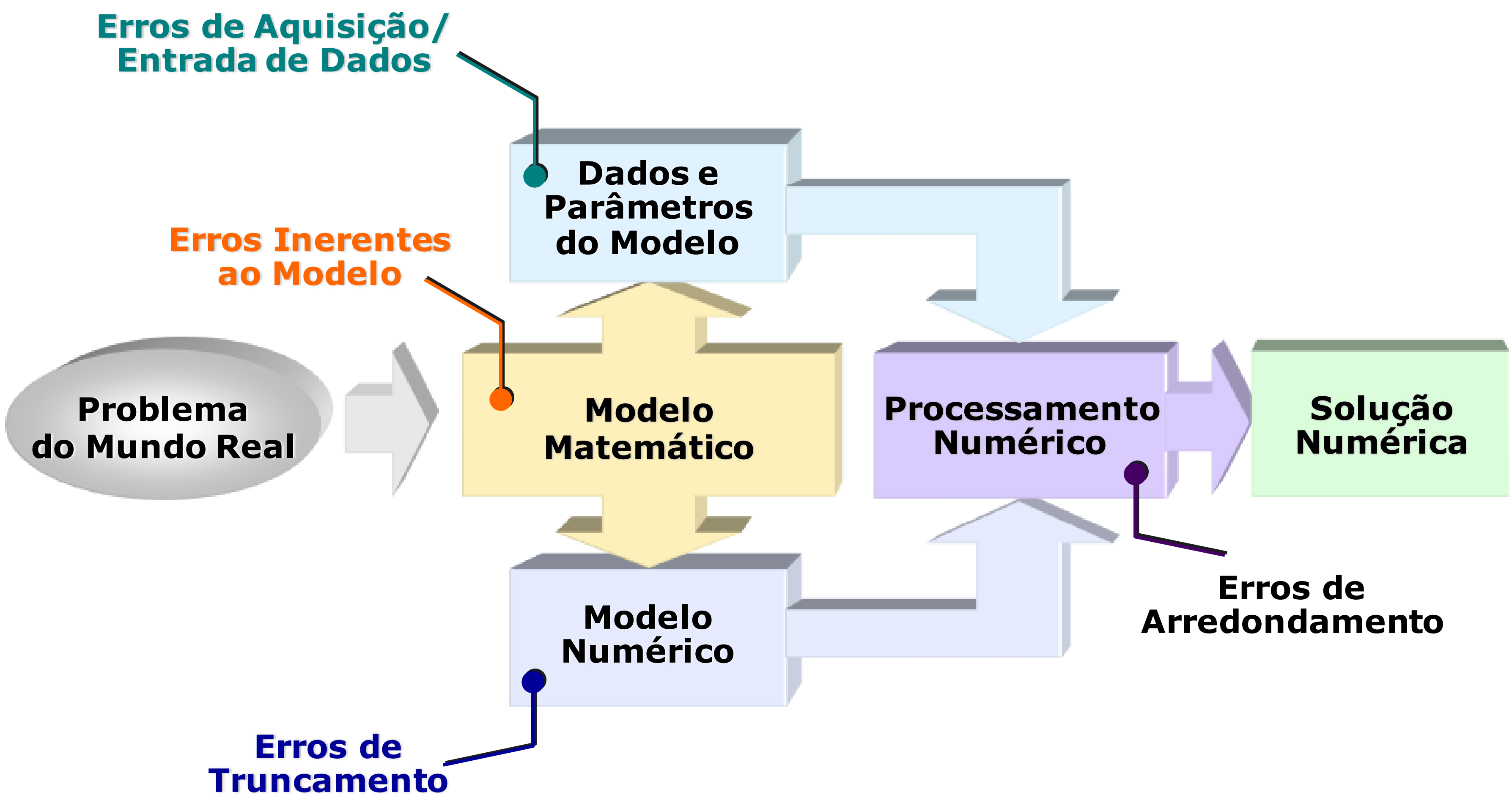
Em muitos casos, o erro de *truncamento* é precisamente a diferença entre o modelo matemático e o modelo numérico.

Erros - Existência VII

- **Natureza dos Erros IV**
  - **Erros de *arredondamento***
    - **Inerentes à estrutura da máquina e à utilização de uma aritmética de precisão finita**

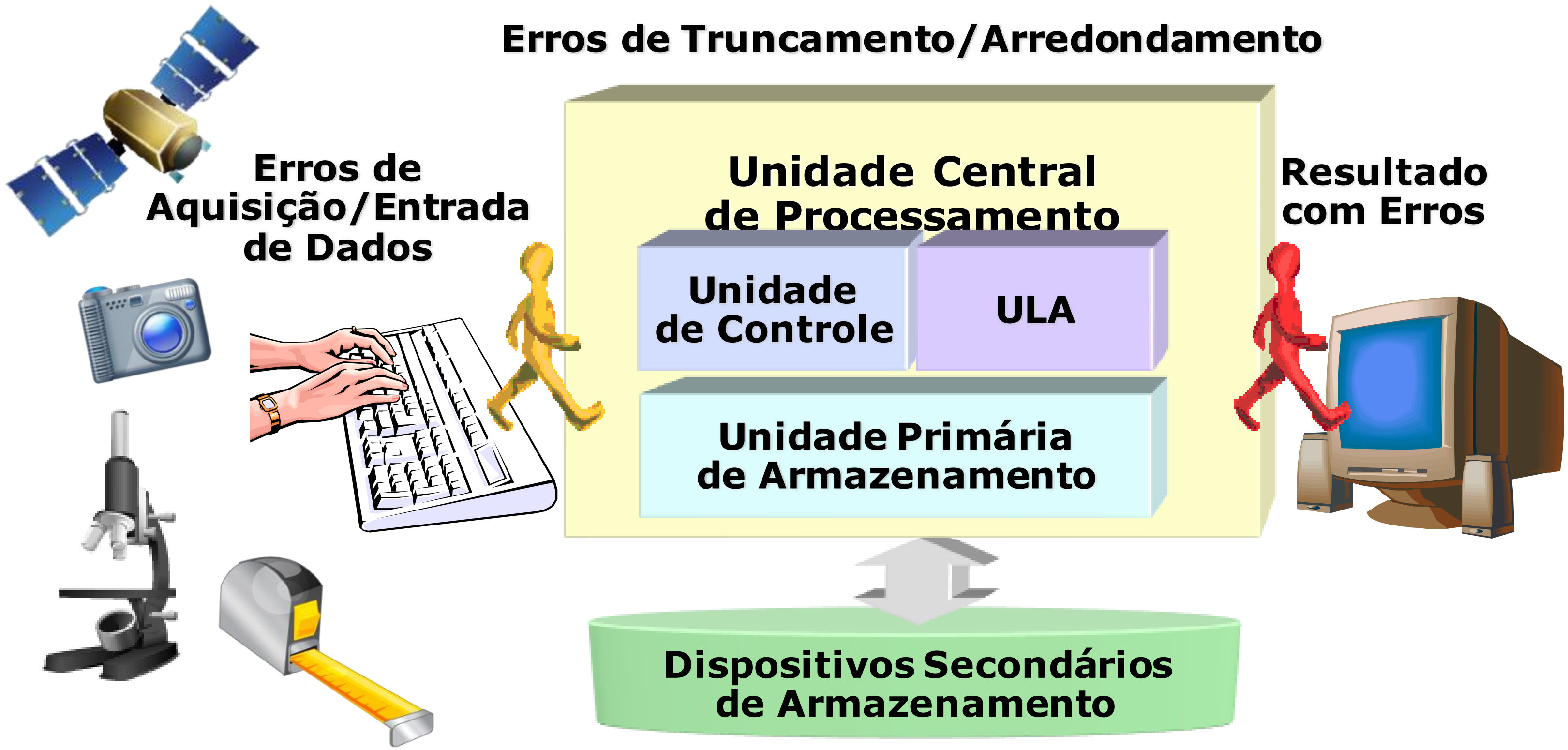
Erros - Existência VIII

- **Fontes de Erros I**



Erros - Existência IX

- **Fontes de Erros II**



Erros - Existência X

- **Representação Numérica em Máquinas Digitais I**
  - **Discreta ⇨ Conjunto finito de números em qualquer intervalo *[a, b]* de interesse**
    - **Implicação imediata ⇨ Possibilidade de comprometimento da precisão dos resultados, mesmo em representações de dupla precisão**

Erros - Existência XI

- **Resultado na Saída**

  - **Incorporação de <u>todos</u> os erros do processo**
  - **<u>Quão</u> confiável é o resultado <u>aproximado</u>?**

    - **<u>Quanto</u> erro está <u>presente</u> no resultado?**
    - **Até que ponto o erro presente no resultado é <u>tolerável</u>?**

Erros - Existência XII

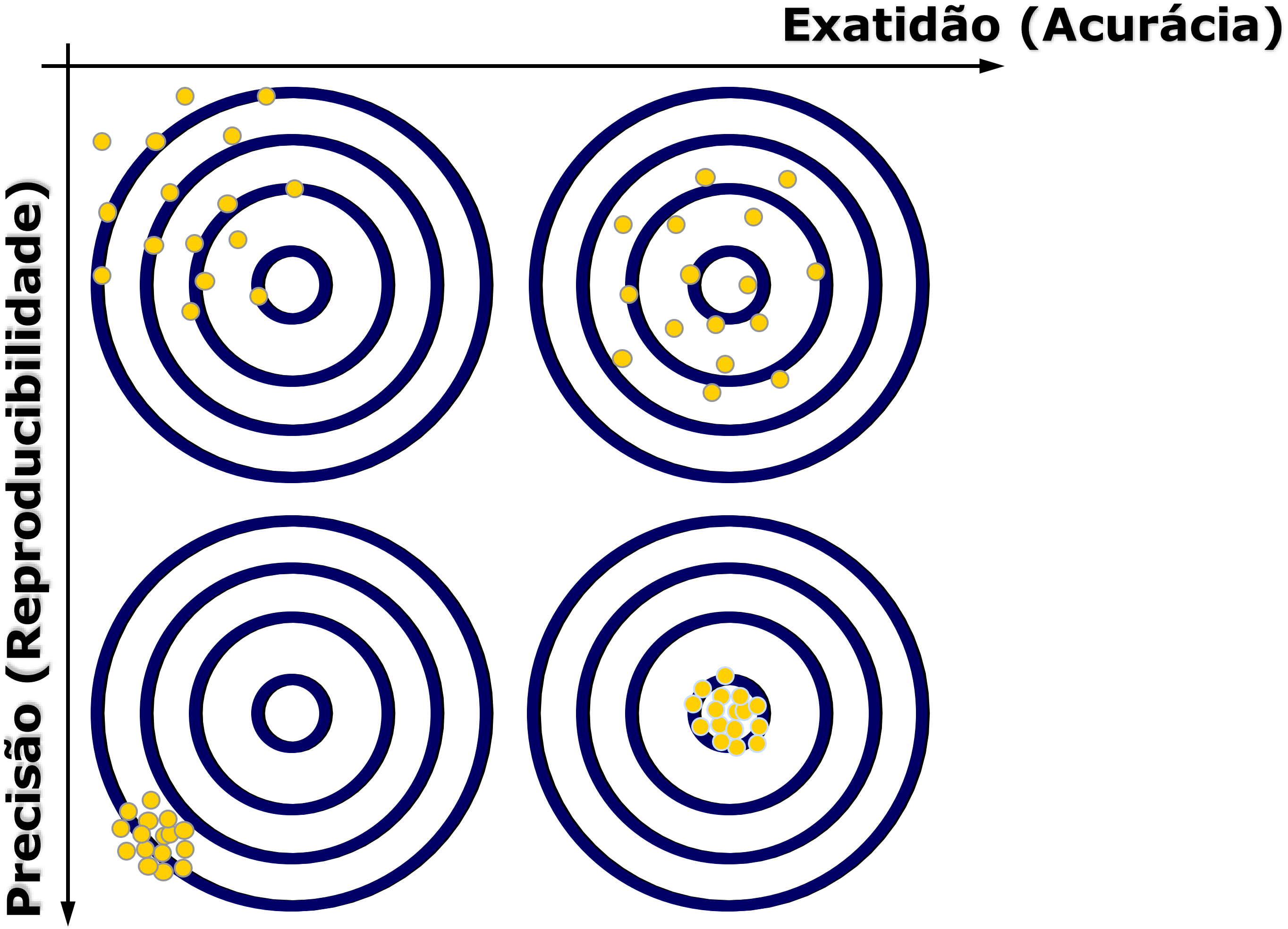
- ***Acurácia*** **(ou *Exatidão*)**
  - **Quão próximo um valor computado/mensurado se encontra do valor real (verdadeiro)**
- ***Precisão*** **(ou *Reprodutibilidade*)**
  - **Quão próximo um valor computado/mensurado se encontra de valores previamente computados/mensurados**

Erros - Existência XIII

- ***Inacurácia*** **(ou *Inexatidão*)**
  - **Desvio sistemático do valor real**
- ***Imprecisão*** **(ou *Incerteza*)**
  - **Magnitude do espalhamento dos valores**

Erros - Existência XIV

- *Exatidão* **x** *Precisão*



Erros - Existência XV

- Indicador de *Precisão* de um Resultado
  - Número de algarismos significativos
    - Algarismos significativos (*as*)
      - Algarismos que podem ser usados com *confiança*

Erros - Existência XVI

- ***As*** **de um número I**
  - **Exemplo 02: Considerem-se os seguintes valores de *médias* obtidas em um experimento estatístico**
    - $\mu = 138$ — **0 casas decimais (cd)**
    - $\mu = 138,7$ — **1 cd**
    - $\mu = 138,76$ — **2 cd**
    - $\mu = 138,76875$ — **5 cd**
    - $\mu = 138,\ 7687549$ — **7 cd**
    - $\mu = 138,\ 768754927$ — **9 cd**

Erros - Existência XVII

- *As* de um número II
  - Exemplo 02: Os valores das médias podem ser representadas como:
    - $\mu = 138$ ⇨ $\mu = 0{,}138 \, . \, 10^3$
    - $\mu = 138{,}7$ ⇨ $\mu = 0{,}1387 \, . \, 10^3$
    - $\mu = 138{,}76$ ⇨ $\mu = 0{,}13876 \, . \, 10^3$
    - $\mu = 138{,}76875$ ⇨ $\mu = 0{,}13876875 \, . \, 10^3$
    - $\mu = 138{,}\,7687549$ ⇨ $\mu = 0{,}1387687549 \, . \, 10^3$
    - $\mu = 138{,}\,768754927$ ⇨ $\mu = 0{,}138768754927 \, . \, 10^3$

Erros - Existência XVIII

- ***As*** **de um número III**
  - **Exemplo 02:**
    - $\mu = 0{,}138 \times 10^3$ ⇨ 3 as
    - $\mu = 0{,}1387 \times 10^3$ ⇨ 4 as
    - $\mu = 0{,}13876 \times 10^3$ ⇨ 5 as
    - $\mu = 0{,}13876875 \times 10^3$ ⇨ 8 as
    - $\mu = 0{,}1387687549 \times 10^3$ ⇨ 10 as
    - $\mu = 0{,}138768754927 \times 10^3$ ⇨ 12 as

Erros nos Métodos I

- **Método Numérico**
  - **Aproximação da solução de um problema de Matemática**
    - **Truncamento de uma solução em série, considerando apenas um número finito de termos**
    - **Exemplo 03:** $exp(x)$

$$exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

Erros nos Métodos II

- **Exemplo 03: Determinação do valor de e.**

Lembrar que $e = \sum_{n=0}^{\infty} \frac{1}{n!}$**. Logo:**

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2{,}71828182845905$$

um truncamento no sexto termo gera:

$$e = \sum_{n=0}^{5} \frac{1}{n!} = 2{,}71666666666667$$

Erros nos Métodos III

- **Exemplo 03:**

  Então, o erro de truncamento, E_T, será:

$$E_T = \sum_{n=0}^{\infty} \frac{1}{n!} - \sum_{n=0}^{5} \frac{1}{n!}$$

$$E_T = 2{,}718281828459045 - 2{,}716666666666667$$

$$\Rightarrow \quad E_T = 0{,}001615161792378$$

Erros nos Métodos IV

- **Exemplo 04: Determinação do número de termos para a aproximação de cos(x) com 8 as, considerando $x=\pi/3$.**

  Lembrar que:

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

Erros nos Métodos V

- **Exemplo 04: Então**

$$\frac{x^2}{2} = \frac{(0.3\pi)^2}{2} = 0.4444132198 \qquad \cos x = 1 - \frac{x^2}{2} = 0.555867802$$

$$\frac{x^4}{4!} = \frac{(0.3\pi)^4}{24} = 0.032875568 \qquad \cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} = 0.588743370$$

$$\frac{x^6}{6!} = \frac{(0.3\pi)^6}{720} = 0.000973407 \qquad \cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} = 0.587769964$$

Observe-se que o segundo as não mais se alterará.

Erros nos Métodos VI

- **Exemplo 04: E que o quarto as não mais se alterará a partir de:**

$$\frac{x^8}{8!} = \frac{(0.3\pi)^8}{40320} = 0.00001544 \qquad \cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} = 0.587785404$$

- **nem o sexto as a partir de:**

$$\frac{x^{10}}{10!} = \frac{(0.3\pi)^{10}}{3628800} = 0.000000152387 \qquad \cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{x^{10}}{10!} = 0.587785251$$

- **nem o oitavo as a partir de:**

$$\frac{x^{12}}{12!} = \frac{(0.3\pi)^{12}}{479001600} = 0.000000000102545 \qquad \cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{x^{10}}{10!} = 0.587785251$$

Erros nos Métodos VII

- **Exemplo 04:**

  Assim sendo, o número de termos para a aproximação de $cos(x)$ com 8 *as* é igual a 7 (incluindo o termo de ordem 0, igual a 1)

Erros nos Métodos VIII

- **Exercício 01: Determinar o número de termos para a aproximação de**
  1. $log(1+x)$ **com 8 as, considerando** $x = 0{,}09$
  2. $sen(x)$ **com 6 as, considerando** $x = 4\pi/3$
  3. $exp(x)$ **com 7 as, considerando** $x = 1/3$

  Qual a conclusão a que se chega a partir destes cálculos?

Erros - Existência XIX

- Erro de *Representação* x Erro de *Truncamento de Dígitos*
  - Erro de *Representação*
    - Associado à conversão numérica entre bases (representação humana e de máquina) ou à realização de operações aritméticas
  - Erro de *Truncamento de Dígitos*
    - Associado à quantidade de informação que a máquina pode conter sob a forma de um número

Erros - Existência XX

- **Representação dos números reais com um número finito de dígitos (aproximação)**

  Ex. 05: Cálculo da área de uma circunferência de raio 100 m

  Possíveis resultados:

  (1) $A = 31400\ m^2$

  (2) $A = 31416\ m^2$

  (3) $A = 31415{,}92654\ m^2$

Erro de Representação

π não tem representação finita - *3,14* (1), *3,1416* (2) e *3,141592654* (3)

Erros - Existência XXI

- **Representação dos números reais com um número finito de dígitos (aproximação)**
  - **Dependência da representação numérica da máquina utilizada**

$$0{,}1_{10} = 0{,}000110011001100 11\ldots_2$$

Um número pode ter representação **finita** em uma base e **não finita** em outra

Erro de Representação

Operações com dados **imprecisos** ou **incertos** acarretam a **propagação do erro**.

Erros - Existência XXII

- **Ex. 06: Determinar**

$$S = \sum_{i=1}^{3000} x_i$$

a partir de uma calculadora e um computador, para $x_i = 0{,}5$ e $x_i = 0{,}1$

x_i	Calculadora	Computador
0,5	S= 1500	S= 1500
0,1	S= 300	S=300,00909424 (precisão *simples*)
		S=299,999999999999720 (precisão *dupla*)

Erros - Existência XXIII

Ex. 07: **Conversão de $0{,}1_{10}$ para a base 2.**

$$0{,}1_{10} = 0{,}000110011001100011\ldots_2$$

$0{,}1_{10}$ **não tem representação exata na base 2**

> **A representação de um número depende da base em uso e do número máximo de dígitos usados em sua representação.**

Erros - Tipos I

- **Absoluto**

    - **Diferença entre o valor exato de um número e o seu valor aproximado (em módulo)**

$$EA_x = | x - \bar{x} |$$

Erros - Tipos II

- Relativo

  - Razão entre o erro absoluto e o valor exato do número considerado (em módulo)

$$ER_x = \frac{|x - \bar{x}|}{|x|}$$

$$Erro\ Percentual_x = ER_x \, . \, 100\%$$

Erros - Tipos III

- **Relativo**

    - **Este tipo de erro é utilizado em processos iterativos pois, sendo o processo convergente, a cada iteração o valor atual está mais próximo mais do valor exato do que o valor anterior**

$$\bar{x} \equiv valor\ anterior$$

$$x \equiv valor\ atual$$

Erros - Tipos IV

- **Erro Absoluto - Considerações I**

  - EA_x só poderá ser determinado se x for conhecido com exatidão
  - Na prática, costuma-se trabalhar com um limitante superior para o erro, ao invés do próprio erro ($|E| < \varepsilon$, sendo ε é o limitante)

    Ex. 08: Para $\pi \in (3{,}14;\ 3{,}15)$

$$\left|EA_\pi\right| = \left|\pi - \bar{\pi}\right| < 0{,}01$$

Erros – Tipos V

- **Erro Absoluto - Considerações II**

Ex. 08: Sejam $a = 3876{,}373$ **e** $b = 1{,}373$

Considerando-se a parte inteira de a **(**a'**) o erro absoluto será:**

$$EA_a = \left|a - a'\right| = 0{,}373$$

$$EA_b = \left|b - b'\right| = 0{,}373$$

e a parte inteira de b **(**b'**) , o erro absoluto será:**

Erros – Tipos VI

- Erro Absoluto - Considerações III

  - Obviamente, o resultado do erro absoluto é o mesmo nos dois casos

  - Entretanto, o peso da aproximação em b é maior do que em a

Erros – Tipos VII

- **Erro Relativo - Consideração**

O erro relativo pode, entretanto, traduzir perfeitamente este fato, pois:

$$ER_a = \frac{0{,}373}{3876} \cong 0{,}000096 \leq 10^{-4}$$

$$ER_b = \frac{0{,}373}{1} \cong 0{,}373 \leq 5_X 10^0$$

Erros - Tipos VIII

Ex. 09: Cálculo do erro relativo na representação dos números $a = 2112{,}9$ e $e = 5{,}3$, sendo $|EA| < 0{,}1$

$$|ER_a| = |a - \bar{a}|/|a| = 0{,}1/2112{,}9 \cong 4{,}7 \times 10^{-5}$$

$$|ER_e| = |e - \bar{e}|/|e| = 0{,}1/5{,}3 \cong 0{,}02$$

Conclusão: a **é representado com** ***maior*** **precisão do que** e

Erros – Tipos IX

- **Arredondamento**

- **Truncamento de Dígitos**

> Quanto *menor* for o **erro**, maior será a **precisão** do resultado da operação.

Erros – Tipos X

▪Arredondamento I

Ex. 10: Cálculo de $\sqrt{2}$ utilizando uma calculadora digital

Valor apresentado: 1,41421356

Valor real: 1,41421356...

Erros – Tipos XI

- **Arredondamento II**
  - **Inexistência de forma de representação de números irracionais com uma quantidade finita de algarismos**
    - **Apresentação de uma aproximação do número pela calculadora**
      - **Erro de arredondamento**

Erros – Tipos XII

- Truncamento de Dígitos

  - Descarte dos dígitos finais de uma representação exata por limitações de representação em vírgula flutuante $\sqrt{2}$
    - Ex. 11: Representação truncada de em vírgula flutuante com 7 dígitos

      Valor apresentado: $1{,}4142135$

      Valor real: $1{,}41421356\ldots$

Arredondamento e Truncamento I

- **Erros de Truncamento e Arredondamento - Demonstração**
  - **Em um sistema que opera em ponto flutuante de t dígitos na base 10, e seja x:**
    - $x = f_x.10^e + g_x.10^{e-t}$ ($0,1 \leq f_x < 1$ e $0,1 \leq g_x < 1$)
      - **Para** $t = 4$ **e** $x = 234,57$, **então:**

$$x = 0,2345 \; . \; 10^3 + 0,7 \; . \; 10^{-1}$$

$$f_x = 0,2345$$

$$g_x = 0,7$$

Erros - Truncamento

- No **truncamento**, $g_x.10^{e-t}$ é desprezado e

$$\bar{x} = f_x.10^e$$

$$|EA_x| = |x - \bar{x}| = |g_x|.10^{e-t} < 10^{e-t}$$

$$|ER_x| = \frac{|EA_x|}{|x|} = \frac{|g_x|.10^{e-t}}{|f_x|.10^e + |g_x|.10^{e-t}} < \frac{10^{e-t}}{0{,}1.10^e} = 10^{-t+1}$$

,

Erros – Arredondamento I

- No **arredondamento simétrico** (forma mais utilizada):

$$\bar{x} = \begin{cases} f_x \cdot 10^e & \text{, se } |g_x| < \dfrac{1}{2} \\ f_x \cdot 10^e + 10^{e-t} & \text{, se } |g_x| \geq \dfrac{1}{2} \end{cases}$$

(para o primeiro caso: (g_x é desprezado))

(para o segundo caso: (soma 1 ao último dígito f_x))

Erros - Arredondamento II

Se $\left|g_x\right| < \dfrac{1}{2}$ **, então:**

$$\left|EA_x\right| = \left|x - \overline{x}\right| = \left|g_x\right|.10^{e-t} < \frac{1}{2}.10^{e-t}$$

$$\left|ER_x\right| = \frac{\left|EA_x\right|}{\left|x\right|} = \frac{\left|g_x\right|.10^{e-t}}{\left|f_x\right|.10^{e} + \left|g_x\right|.10^{e-t}} < \frac{0{,}5.10^{e-t}}{0{,}1.10^{e}} = \frac{1}{2}.10^{-t+1}$$

Erros – Arredondamento III

Se $\left|g_x\right| \geq \dfrac{1}{2}$**, então:**

$$\left|EA_x\right| = \left|x - \overline{x}\right| = \left|\left(f_x\,.10^{e} + g_x\,.10^{e-t}\right) - \left(f_x\,.10^{e} + 10^{e-t}\right)\right|$$

$$\left|EA_x\right| = \left|g_x.10^{e-t} - 10^{e-t}\right| = \left|\left(g_x - 1\right)\right|.10^{e-t} \leq \frac{1}{2}.10^{e-t}$$

$$\left|ER_x\right| = \frac{\left|EA_x\right|}{\left|x\right|} \leq \frac{1/2.10^{e-t}}{\left|f_x.10^{e} + 10^{e-t}\right|} < \frac{1/2.10^{e-t}}{\left|f_x\right|.10^{e}} \leq \frac{1/2.10^{e-t}}{0{,}1.10^{e}} = \left(\frac{1}{2}.10^{-t+1}\right)$$

Arredondamento e Truncamento I

- **Erros de Truncamento e Arredondamento**

  - **Sistema operando em ponto flutuante - Base 10**

    - **Er** $\left|EA_x\right| < 10^{e-t}$ **nto** e $\left|ER_x\right| < 10^{-t+1}$

    - **Er** **ento** $\left|EA_x\right| \leq \frac{1}{2} \times 10^{e-t}$ e $\left|ER_x\right| < \frac{1}{2} \times 10^{-t+1}$

e - nº de dígitos inteiros
t - nº de dígitos

Arredondamento e Truncamento II

- **Sistema de aritmética de ponto flutuante de 4 dígitos, precisão dupla**

  - **Ex. 12: Seja** $x = 0{,}937.10^4$ **e** $y = 0{,}1272.10^2$**. Calcular** $x+y$**.**
    - **Alinhamento dos pontos decimais antes da soma**

      $x = 0{,}937.\ 10^4$ e $y = 0{,}001272.\ 10^4,$

      $x+y = 0{,}938272.\ 10^4$

    - **Resultado com 4 dígitos**

      Arredondamento: $\overline{x+y} = 0{,}9383.10^4$

      Truncamento: $x+y = 0{,}9382.10^4$

Arredondamento e Truncamento III

- **Sistema de aritmética de ponto flutuante de 4 dígitos, precisão dupla**

  - **Ex. 12: Seja** $x = 0{,}937.10^4$ **e** $y = 0{,}1272.10^2$**. Calcular** $x.y$**.**
    - **Alinhamento dos pontos decimais antes da soma**

      $$x.y = (0{,}937.10^4).(0{,}1272.10^2)$$

      $$x.y = (0{,}937.0{,}1272).10^6 \Rightarrow x.y = 0{,}1191864.10^6$$

    - **Resultado com 4 dígitos**

      Arredondamento: $x.y = 0{,}1192.10^6$

      Truncamento: $x.y = 0{,}1191.10^6$

Arredondamento e Truncamento IV

- **Considerações**
  - **Ainda que as parcelas ou fatores de uma operação possam ser representados exatamente no sistema, não se pode esperar que o resultado armazenado seja exato.**
    - **x e y tinham representação exata, mas os resultados $x+y$ e $x.y$ tiveram representação aproximada.**

Arredondamento e Truncamento V

- **Ex. 13: Seja** $x = 0{,}7237.10^{4}$, $y = 0{,}2145.10^{-4}$ **e** $z = 0{,}2585.10^{1}$. **Efetuar a operação** $x + y + z$ **e calcular o erro relativo do resultado, supondo** x, y **e** z **exatamente representados.**

$$x+y+z = 0{,}7237.10^{4} + 0{,}2145.10^{-4} + 0{,}2585.10^{1}$$
$$= 0{,}7237.10^{4} + 0{,}00000000002145.10^{4} + 0{,}0002585.10^{4} = 0{,}723958502.10^{4}$$

- **Resultado com 4 dígitos**

Arredondamento: $x+y+z = 0{,}7240.10^{4}$

Truncamento: $x+y+z = 0{,}7239.10^{4}$

Arredondamento e Truncamento VI

- **Erro relativo (no arredondamento):**

$$ER_{x+y+z} = \left|\frac{EA_{x+y+z}}{x}\right| = \left|\frac{0{,}723958502.10^4 - 0{,}7240.10^4}{0{,}723958502.10^4}\right| \cong$$

$$5{,}7321.10^{-5} < \frac{1}{2}.10^{-3}$$

Arredondamento e Truncamento VII

- Sistemas de Vírgula Flutuante (*VF*)

  - Um sistema $VF(b, p, q)$ é constituído por todos os números reais X da forma:

$$X = \pm m b^{t}$$

, em que

$$b^{-1} \leq m \leq 1 - b^{-p}$$

e ainda $X = 0$

Arredondamento e Truncamento VIII

- **Sistemas de Vírgula Flutuante (*VF*)**

  - **Portanto,**

$$X = \pm(.d_{-1}d_{-2}d_{-3}\ldots d_{-p})b^{\pm(t_{q-1}\ldots t_1 t_0)}$$

na qual

- p **um número finito de dígitos para a mantissa;**
- q **um número finito de dígitos para o expoente;**
- b **é a base do sistema.**

Arredondamento e Truncamento IX

- **Sistemas de Vírgula Flutuante (*VF*)**

  - **Considera-se que a mantissa é normalizada, i.e., $d \neq 0$, exceto a representação do zero.**

  - **Representam-se na forma $VF(b, p, q, Y)$, onde Y determina qual método o sistema adota:**

    Caso $Y = A$ → Arredondamento;

    Caso $Y = T$ → Truncamento de Dígitos.

Arredondamento e Truncamento X

- **Sistemas de Vírgula Flutuante (*VF*)**

  - **Unidade de arredondamento (u): majorante do erro relativo na representação de um número num dado sistema $VF(b, p, q)$, tal que:**

    - $u = \frac{1}{2} b^{1-p}$ **em** $VF(b, p, q, A)$
    - $u = b^{1-p}$ **em** $VF(b, p, q, T)$,

Arredondamento e Truncamento XI

Ex. 14: Determine as raízes da equação $x^2 + 0{,}7341x + 0{,}600.10^{-4} = 0$ no sistema $VF(10, 4, 2, T)$, considerando que não existem dígitos de guarda no processamento das operações em ponto flutuante.

a) A partir da expressão utilizada na resolução de equações quadráticas, calcule o erros absolutos e relativos (EA_{x1}, EA_{x2}, ER_{x1} e ER_{x2}).

Arredondamento e Truncamento XII

b) Justifique a origem do erro relativo obtido na menor raiz (em módulo), sugerindo uma forma de melhoria numérica para a resolução de tal problema.

Solução:

a) $$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$fl(b) = 0{,}7341.10^0$$

$$fl(b^2) = (0{,}7341.0{,}7341)(10^0.10^0) =$$

$$0{,}5389028.10^0 \Rightarrow \quad fl(b^2) = 0{,}5389.10^0$$

Arredondamento e Truncamento XIII

Solução:

a) $$fl(c) = (0.6000)10^{-4}$$

$$fl(4) = (0.4000)10^{1}$$

$$fl(2) = (0.2000)10^{1}$$

$$fl(4c) = (0,4000.0,6000)(10^{-4}.10^{1})$$

$$\Rightarrow fl(4c) = 0,2400.10^{-3}$$

$$fl(b^2 - 4c) = 0,5389.10^{0} - 0,2400.10^{-3} =$$

$$(0,5389 - 0,0002400).10^{0} =$$

$$\Rightarrow fl(b^2 - 4c) = 0,5387.10^{0}$$

$$fl(\sqrt{(b^2 - 4c})) = (0,5387.10^{0})^{1/2} = 0,7339.10^{0}$$

Arredondamento e Truncamento XIV

Solução:

a) Primeira raiz:

$$fl(-b-\sqrt{b^2-4c}) = -0{,}7341.10^0 - 0{,}7339.10^0$$

$$\Rightarrow fl(x_1) = fl\left(\frac{-b-\sqrt{b^2-4c}}{2}\right) = \frac{-0{,}1468.10^1}{0{,}2000.10^1}$$

$$\Rightarrow fl(x_1) = -0{,}7340.10^0$$

Arredondamento e Truncamento XV

Solução:

a) Segunda raiz:

$$fl\left(-b+\sqrt{b^2-4c}\right) = -0{,}7341.10^{0} + 0{,}7339.10^{0}$$

$$\Rightarrow fl(x_1) = fl\left(\frac{-b+\sqrt{b^2-4c}}{2}\right) = \frac{-0{,}0002.10^{1}}{0{,}2000.10^{1}}$$

$$\Rightarrow fl(x_1) = -0{,}1000.10^{-3}$$

O cancelamento subtrativo (ou catastrófico) ocorre quando se subtraem números muito próximos em sistemas de vírgula flutuante.

Arredondamento e Truncamento XVI

Solução:

a) **Para calcular os erros cometidos em *FP*, é necessário conhecer os valores exatos das raízes.**

Considerando um dígito a mais do que a representação da mantissa no sistema, i.e., 5 dígitos, obtém-se:

$x_1 = -0,73402.10^{0}$ e $x_2 = -0,81742.10^{-4}$

Arredondamento e Truncamento XVII

Solução:

a) Assim sendo, os erros absolutos e relativos serão:

$$\left|EA_{x1}\right| = \left|-0{,}73402.10^{0} - (-0{,}7340.10^{0})\right| = 0{,}20000.10^{-4}$$

$$\left|EA_{x1}\right| = \left|-0{,}81742.10^{-4} - (-0{,}1000.10^{-3})\right| = 0{,}18258.10^{-4}$$

$$\left|ER_{x1}\right| = \left|\frac{EA_{x1}}{x_1}\right| = \left|\frac{0{,}2000.10^{-4}}{-0{,}73402.10^{0}}\right| = 0{,}27247.10^{-4} \Rightarrow \left|ER_{x1}\right|_{\%} \cong 0{,}003\%$$

$$\left|ER_{x1}\right|_{\%} \cong 0{,}0\%$$

$$\left|ER_{x2}\right| = \left|\frac{EA_{x2}}{x_2}\right| = \left|\frac{0{,}18258.10^{-4}}{-0{,}81742.10^{-4}}\right| = 0{,}22336.10^{0} \Rightarrow \left|ER_{x2}\right|_{\%} \cong 22{,}3\%$$

Arredondamento e Truncamento XVIII

Solução:

a) Constatação:

Apesar dos erros absolutos serem praticamente iguais, a segunda raiz apresenta um erro relativo quatro ordens de grandeza maior do que o erro relativo cometido no cálculo da primeira raiz.

Arredondamento e Truncamento XIX

Solução:

b) O problema do erro relativo cometido no cálculo da segunda raiz deve-se ao cancelamento subtrativo, verificado quando números muito próximos se subtraem em aritmética de vírgula flutuante.

Arredondamento e Truncamento XX

Solução:

b) Para evitar o cancelamento subtrativo, 2 opções conduzem ao mesmo resultado, a saber:

1. Manipulação da fórmula para a determinação dos zeros

$$x_2 = \frac{-b + \sqrt{b^2 - 4c}}{2} = \frac{-b + \sqrt{b^2 - 4c}}{2} \cdot \frac{-b - \sqrt{b^2 - 4c}}{-b - \sqrt{b^2 - 4c}} =$$

$$\frac{(-b)^2 + \left(\sqrt{b^2 - 4c}\right)^2}{2.\left(-b - \sqrt{b^2 - 4c}\right)} = \frac{2c}{-b - \sqrt{b^2 - 4c}} = \frac{2c}{2x_1} = \frac{c}{x_1}$$

Arredondamento e Truncamento XXI

Solução:

1. **Manipulação da fórmula para a determinação dos zeros**

   Assim:

$$fl(x_2) = fl\left(\frac{c}{x_1}\right) = \frac{0{,}6000.10^{-4}}{-0{,}7340.10^{0}} = -0{,}8174.10^{-4}$$

2. **Manipulação simbólica da equação genérica de segundo grau**

$$ax^2 + bx + c = a(x - x_1)(x - x_2) =$$

$$a(x_2 - x_1x - x_2x + x_1x_2) =$$

$$ax^2 - a(x_1 + x_2)x + ax_1x_2 \Rightarrow c = ax_1x_2$$

ou

$$x_2 = \frac{c}{ax_1}$$

Erros – Propagação I

- Propagação dos Erros
  - Durante as operações aritméticas de um método, os erros dos operandos produzem um erro no resultado da operação
    - Propagação ao longo do processo
    - Determinação do erro no resultado final obtido

Erros – Propagação II

- **Ex. 14:** Sejam as operações a seguir, processadas em uma máquina com 4 dígitos significativos e fazendo-se: $a = 0,3491.10^4$ e $b = 0,2345.10^0$.

  $(b+a)-a=(0,2345.10^0+0,3491.10^4)-0,3491.10^4=0,3491.10^4-0,3491.10^4$

  $= 0,0000$

  $b+(a-a)=0,2345.10^0+(0,3491.10^4-0,3491.10^4)=0,2345+0,0000$

  $= 0,2345$

Erros – Propagação III

- **Os dois resultados são diferentes, quando não deveriam ser.**

  $(b + a) - a = 0{,}0000$ **e** $b + (a - a) = 0{,}2345$

  - **Causa**
    - **Arredondamento da adição** $(b + a)$**, a qual tem** 8 **dígitos ⇨ Cancelamento subtrativo de** $(b + a) - a$ **devido à representação de máquina com** 4 **dígitos**

A distributividade é uma propriedade da adição.

Erros – Propagação IV

- Resolução numérica de um problema
  - Importância do conhecimento dos efeitos da propagação de erros
    - Determinação do erro final de uma operação
    - Conhecimento da sensibilidade de um determinado problema ou método numérico

Erros – Propagação V

- **Ex. 15: Dados $a = 50 \pm 3$ e $b = 21 \pm 1$, calcular $a + b$.**
  - **Variação de a ⇨ 47 a 53**
  - **Variação de b ⇨ 20 a 22**
  - **Menor valor da soma ⇨ $47 + 20 = 67$**
  - **Maior valor da soma ⇨ $53 + 22 = 75$**
  - **$a + b = (50 + 21) \pm 4 = 71 \pm 4$ ⇨ 67 a 75**

Erros – Propagação VI

- **Ex. 16: Dados** $a = 50 \pm 3$ **e** $b = 21 \pm 1$**, calcular** $a - b$**.**
  - **Variação de** a ⇨ **47 a 53**
  - **Variação de** b ⇨ **20 a 22**
  - **Menor valor da diferença** ⇨ $47 - 20 = 25$
  - **Maior valor da diferença** ⇨ $53 - 22 = 33$

  - $a - b = (50 - 21) \pm 4 = 29 \pm 4$ ⇨ **25 a 33**

Na subtração, os erros absolutos se somam, pois sempre se admite o pior caso.

Erros – Propagação VII

- Ex. 17: Dados $a = 50 \pm 3$ e $b = 21 \pm 1$, calcular $a.b$.
  - Variação de a ⇨ 47 a 53
  - Variação de b ⇨ 20 a 22
  - Menor valor do produto ⇨ $47 . 20 = 940$
  - Maior valor da produto ⇨ $53 . 22 = 1166$
  - $a . b = (50 \pm 3) \text{ x } (21 \pm 1)$

    $\approx 1050 \pm (3.21 + 50.1)$

    $\approx 1050 \pm 113$ ⇨ 937 a 1163

Erros – Propagação VII

- Ex. 18: Dados $a = 50 \pm 3$ e $b = 21 \pm 1$, calcular $a.b$.
  - Considerações
    - Despreza-se o produto 3.1, por ser muito pequeno diante de $(3.21 + 50.1) = 113$
    - Ligeiramente diferente do verdadeiro intervalo, por conta da desconsideração do produto 3.1, assumido como desprezível

Erros – Propagação X

- Análise dos Erros **Absoluto** e **Relativo**
  - Expressões para o determinação dos erros nas operações aritméticas
  - Erros presentes na representação das **parcelas** ou **fatores**, assim como no **resultado** da operação
    - Supondo um **erro final arredondado**, sendo x e y, tais que:

$$x = \bar{x} + EA_x \text{ e } y = \bar{y} + EA_y$$

Erros – Propagação XI

- **Adição**
  - **Erro Absoluto**

$$
\begin{aligned}
x + y = (\bar{x} + EA_x) + (\bar{y} + EA_y) = \\
(\bar{x} + \bar{y}) + (EA_x + EA_y)
\end{aligned}
$$

  - **Erro Relativo**

$$
ER_{x+y} = \frac{EA_{x+y}}{\bar{x} + \bar{y}} = ER_x\left(\frac{\bar{x}}{\bar{x} + \bar{y}}\right) + ER_y\left(\frac{\bar{y}}{\bar{x} + \bar{y}}\right)
$$

Erros – Propagação XII

- **Subtração**
  - **Erro Absoluto**

$$x - y = (\bar{x} + EA_x) - (\bar{y} + EA_y) = (\bar{x} - \bar{y}) + (EA_x - EA_y)$$

  - **Erro Relativo**

$$ER_{x-y} = \frac{EA_{x-y}}{\bar{x} - \bar{y}} = ER_x\left(\frac{\bar{x}}{\bar{x} - \bar{y}}\right) - ER_y\left(\frac{\bar{y}}{\bar{x} - \bar{y}}\right)$$

Erros – Propagação XIII

- **Multiplicação**
  - **Erro Absoluto**

$$x.y = (\bar{x} + EA_x).(\bar{y} + EA_y) = \bar{x}.\bar{y} + \bar{y}.EA_x + \bar{x}EA_y + \underbrace{(EA_x.EA_y)}_{\text{muito pequeno}}$$

$$x.y \approx (\bar{x} + EA_x).(\bar{y} + EA_y) = \bar{x}.\bar{y} + \bar{y}.EA_x + \bar{x}EA_y$$

  - **Erro Relativo**

$$ER_{x.y} = ER_x + ER_y$$

Erros – Propagação XIII

- **Divisão**
  - **Erro Absoluto**

$$\frac{x}{y} = \frac{\left(\bar{x} + EA_x\right)}{\left(\bar{y} + EA_y\right)} = \frac{\left(\bar{x} + EA_x\right)}{\bar{y}} \cdot \left(\frac{1}{1 + \frac{EA_y}{\bar{y}}} \right)$$

Simplificação:

$$\frac{1}{1 + \frac{EA_y}{\bar{y}}} = 1 - \frac{EA_y}{\bar{y}} + \left(\frac{EA_y}{\bar{y}}\right)^2 - \left(\frac{EA_y}{\bar{y}}\right)^3 + \ldots$$

(desprezam-se os termos de potência >1)

$$\frac{x}{y} \approx \frac{\bar{x}}{\bar{y}} + \frac{EA_x}{\bar{y}} - \frac{\bar{x}EAy}{\bar{y}^2} = \frac{\bar{y}.EA_x - \bar{x}EA_y}{\bar{y}^2}$$

  - **Erro Relativo**

$$ER_{x/y} = ER_x - ER_y$$

Erros – Análise I

Ex. 19: Cálculo de $ER(x+y)$

$$ER_{x+y} = \frac{EA_{x+y}}{\bar{x} + \bar{y}} + RA$$

$$ER_{x+y} = RA$$

$EA_x = EA_y = 0,$
$\therefore \quad EA_{x+y} = 0$

$$\left|ER_{x+y}\right| = \left|RA\right| < \frac{1}{2} \times 10^{-t+1}$$

Como x e y são exatamente representados, ER_{x+y} se resume ao *Erro Relativo de Arredondamento* (RA) no resultado da soma.

Erros – Análise II

- **Sistema de aritmética de ponto flutuante de 4 dígitos, precisão dupla I**
  - **Ex. 20: Seja $x = 0{,}937.10^4$, $y = 0{,}1272.10^2$ e $z = 0{,}231.10^1$, calcular $x+y+z$ e $ER_{(x+y+z)}$, sabendo que x, y e z estão exatamente representados.**

    Solução:

    Alinhando as vírgulas decimais:

    $x = 0{,}937000.10^4$

    $y = 0{,}001272.10^4$ e

    $z = 0{,}000231.10^4$

Erros – Análise III

- Ex. 20: Seja $x = 0{,}937.10^4$, $y = 0{,}1272.10^2$ e $z = 0{,}231.10^1$, calcular $x+y+z$ e $ER_{(x+y+z)}$, sabendo que x, y e z estão exatamente representados.

  Solução:

  A soma é feita por partes: $(x+y)+z$

  $$x+y = 0{,}9383 \,.\, 10^4$$

  $$x+y+z = 0{,}9383 \,.\, 10^4 + 0{,}000231 \,.\, 10^4$$

  $$x+y+z = 0{,}938531.\, 10^4$$

  $$x+y+z = 0{,}9385.\, 10^4$$

  (após o arredondamento)

Erros – Análise IV

Solução:

$$ER_{x+y+z} = ER_s\left(\frac{\overline{x+y}}{\overline{x+y}+\overline{z}}\right) + ER_z\left(\frac{EA_z}{\overline{x+y}+\overline{z}}\right) + RA$$

$$ER_{x+y+z} = ER_s\left(\frac{\overline{x+y}}{\overline{x+y}+\overline{z}}\right) + RA$$

$$ER_{x+y+z} = RA_s\left(\frac{\overline{x+y}}{\overline{x+y}+\overline{z}}\right) + RA = RA\left(\frac{\overline{x+y}}{\overline{x+y}+\overline{z}} + 1\right)$$

$EA_z = 0,$
$\therefore \quad ER_z = 0$

$$\left|ER_{x+y+z}\right| < \left(\frac{\overline{x+y}}{\overline{x+y}+\overline{z}} + 1\right)\frac{1}{2}\times 10^{-t+1}$$

Erros – Análise V

Solução:

$$\left|ER_{x+y+z}\right| < \left(\frac{\overline{x+y}}{\overline{x+y+z}}+1\right)\frac{1}{2}.10^{-t+1}$$

$$\left|ER_{x+y+z}\right| < \left(\frac{0{,}9383.10^{4}}{0{,}9385.10^{4}}+1\right)\frac{1}{2}.10^{-3}$$

$$\left|ER_{x+y+z}\right| < 0{,}9998.10^{-3}$$

Erros – Análise VI

- **Ex. 21:** Supondo que u é representado em um computador por $\bar{u}$, que é obtido por arredondamento. Obter os limites superiores para os erros relativos de $v = 2.\ \bar{u}$ e $w = \bar{u} + \bar{u}$.

Erros – Análise VII

- **Ex. 21:**

  Solução:

$$v = 2.\overline{u}$$

$$ER_{2.\overline{u}} = ER_2 + ER_{\overline{u}} + RA = RA + RA = 2.RA$$

$$\left|ER_{2.\overline{u}}\right| < 2.\frac{1}{2}.10^{-t+1}$$

$$\boxed{\left|ER_v\right| < 10^{-t+1}}$$

Erros – Análise VIII

- **Ex. 21:**

Solução: $w = \bar{u} + \bar{u}$

$$ER_w = ER_{\bar{u}}\left(\frac{\bar{u}}{\bar{u}+\bar{u}}\right) + ER_{\bar{u}}\left(\frac{\bar{u}}{\bar{u}+\bar{u}}\right) + RA$$

$$ER_w = 2.RA\left(\frac{\bar{u}}{\bar{u}+\bar{u}}\right) + RA = 2.RA$$

$$|ER_w| = 2.|RA| < 2.\frac{1}{2}.10^{-t+1} = 10^{-t+1}$$

$$|ER_w| = |ER_v| < 10^{-t+1}$$

Erros – Sumário I

1. **Erro Relativo da Adição** ⇨ Soma dos erros relativos de cada parcela, ponderados pela participação de cada parcela no total da soma.

2. **Erro Relativo da Subtração** ⇨ Soma dos erros relativos do minuendo e do subtraendo, ponderados pela participação de cada parcela no resultado da subtração.

Erros – Sumário II

3. **Erro Relativo da Multiplicação** ⇨ **Soma dos erros relativos dos fatores.**

4. **Erro Relativo da Divisão** ⇨ **Soma dos erros relativos do dividendo e do divisor.**

Erros – Exercício I

- **Seja um sistema de aritmética de ponto flutuante de 4 dígitos, base decimal e com acumulador de precisão dupla. Dados os números** $x = 0{,}7237.10^4$, $y = 0{,}2145.10^{-3}$ **e** $z = 0{,}2585.10^1$, **efetuar as seguintes operações e obter o erro relativo nos resultados, supondo que** x, y, **e** z **estão exatamente representados.**

a) $x+y+z$ **b)** $x-y-z$ **c)** x/y

d) $(x.y)/z$ **e)** $x.(y/z)$ **f)** $(x+y).z$

Erros – Exercício II

- Supondo que $\overline{x}$ é representado num computador por x e obtido por arredondamento, determinar os limites superiores para os erros relativos de:

a) $u = 3.\overline{x}$

b) $w = \overline{x} + \overline{x} + \overline{x}$

c) $u = 4.\overline{x}$

d) $w = \overline{x} + \overline{x} + \overline{x} + \overline{x}$

Erros – Exercícios III

- Sejam $\bar{i}$ e $\bar{u}$ as representações de i e u obtidas em um computador por arredondamento. Deduzir expressões de limitante de erro, a fim de mostrar que o limitante de erro relativo de $u = 3.\bar{x}.\bar{y}$ é

$$v = (\bar{x} + \bar{x} + \bar{x}).\bar{y}$$

Erros – Exercício IV

- Um computador armazena números reais utilizando 1 bit para o sinal do número, 7 bits para o expoente e 8 bits para a mantissa. Admitindo que haja truncamento, como ficarão armazenados os seguintes números decimais?

a) $n_1 = 25{,}5$ b) $n_2 = 120{,}25$ c) $n_3 = 2{,}5$

d) $n_4 = 460{,}25$ e) $n_5 = 24{,}005$

Erros – Exercícios V

- **Considerando o sistema de vírgula flutuante $F(10, 4, 2, T)$:**

$$1{,}023x^2 + 0{,}3714x + 0{,}5999.10^{-2} = 0$$

e a inexistência de dígitos de guarda (o processador pode ter mais dígitos do que a memória, sendo os dígitos adicionais denominados *dígitos de guarda*) no processamento das operações em ponto flutuante.

Erros – Exercícios VI

a) Determinar os zeros da equação a partir da fórmula resolvente;

b) Calcular os erros absolutos cometidos nos cálculos dos dois zeros;

c) Explicar a origem do erro relativo resultante do cálculo da menor raiz (em módulo), sugerindo uma forma de melhoria numérica para a resolução deste problema.

Erros - Bibliografia

- **Ruggiero, M. A. Gomes & Lopes, V. L. da R.** ***Cálculo Numérico: Aspectos teóricos e computacionais*****. MAKRON Books, 1996, 2ª ed.**
- **Asano, C. H. & Colli, E.** ***Cálculo Numérico: Fundamentos e Aplicações*****. Departamento de Matemática Aplicada – IME/USP, 2007.**
- **Sanches, I. J. & Furlan, D. C.** ***Métodos Numéricos*****. DI/UFPR, 2006.**
- **Paulino, C. D. & Soares, C. Erros e Propagação de Erros,** ***Notas de aula*****, SE/ DM/ IST [Online]** ***http://www.math.ist.utl.pt/stat/pe/qeb/semestre_1_2004-2005/PE_erros.pdf*** **[Último acesso 07 de Junho de 2007].**

Erros - Bibliografia

- Paulino, C. D. & Soares, C. Erros e Propagação de Erros, *Notas de aula*, SE/ DM/ IST [Online] *http://www.math.ist.utl.pt/stat/pe/qeb/semestre_1_2004-2005/PE_erros.pdf* [Último acesso 08 de Setembro de 2011].