Métodos Numéricos 1 (MN1)

Unidade 1: Teoria dos Erros
Parte 3: Tipos de Erros Numéricos

Joaquim Bento Cavalcante Neto
joaquimb@lia.ufc.br

Grupo de Computação Gráfica, Realidade Virtual e Animação (CRAb)
Departamento de Computação (DC)
Universidade Federal do Ceará (UFC)

CRAb

UFC

Erro absoluto

- É a diferença entre o valor exato de um número x e de seu valor aproximado $\bar{x}$:

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

- Normalmente o valor exato não é disponível
  - Obtém-se um limitante superior para o erro ou uma estimativa para o módulo do erro absoluto

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

$$-\varepsilon < x - \bar{x} < +\varepsilon$$

$$\bar{x} - \varepsilon < x < \bar{x} + \varepsilon$$

Erro absoluto: Exemplos

- Sabe-se que o valor para $\pi \in (3.14, 3.15)$:

$$|EA_\pi| = |\pi - \overline{\pi}| < 0.01$$

- Erro absoluto é insuficiente para descrever a precisão de um cálculo (depende da grandeza):
  - x, representado por $\overline{x} = 2112.9$, onde $| EA_x | < 0.1$
  - y, representado por $\overline{y} = 5.3$, onde $| EA_y | < 0.1$
  - x e y não são representados com a mesma precisão

Erro relativo

- É erro absoluto dividido pelo valor aproximado:

$$|ER_x| = \left|\frac{EA_x}{\bar{x}}\right| = \frac{|x-\bar{x}|}{|\bar{x}|}$$

- Exemplos:
  - $\bar{x} = 2112.9$, $|EA_x| < 0.1 \Rightarrow |ER_x| = \left|\frac{EA_x}{\bar{x}}\right| = \frac{0.1}{2112.9} \approx 4.7 \times 10^{-5}$
  - $\bar{y} = 5.3$, $|EA_y| < 0.1 \Rightarrow |ER_y| = \left|\frac{EA_y}{\bar{y}}\right| = \frac{0.1}{5.3} \approx 0.02$

Truncamento e Arredondamento

- Seja um sistema que opera em aritmética de ponto flutuante de t dígitos na base 10, e seja x escrito na forma mostrada abaixo:
  - $x = f_x \times 10^e + g_x \times 10^{e-t}$ onde $0.1 \leq f_x < 1$ e $0 \leq g_x < 1$
- Por exemplo, se t=4 e x = 234.57:
  - $x = 0.2345 \times 10^3 + 0.7 \times 10^{-1}$, onde $f_x = 0.2345$ e $g_x = 0.7$
- A parcela dada por $g_x \times 10^{e-t}$ não pode ser incorporado totalmente à mantissa de x:
  - erros absoluto e relativo máximos cometidos?

Erro de truncamento

- $g_x \times 10^{e-t}$ é desprezado e $\bar{x} = f_x \times 10^e$:
  - $|EA_x| = |\, x - \bar{x} \,| = |g_x| \times 10^{e-t} < 10^{e-t}$

    visto que $|g_x| < 1$

  - $$|ER_x| = \frac{|EA_x|}{|\bar{x}|} = \frac{|g_x| \times 10^{e-t}}{|f_x| \times 10^{e}} < \frac{10^{e-t}}{0.1 \times 10^{e}} = 10^{-t+1}$$

    visto que 0.1 é o menor valor possível para f_x

Erro de arredondamento

- fx é modificado para considerar g_x :
  - Arredondamento simétrico:

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

  - Portanto se $|g_x| < 1/2$, g_x é desprezado, caso contrário, somamos 1 ao último dígito de f_x

Erro de arredondamento

- Se |gx| < ½ :

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

visto que $|g_x| < ½$

$$|ER_x| = \frac{|EA_x|}{|\overline{x}|} = \frac{|g_x| \times 10^{e-t}}{|f_x| \times 10^{e}} < \frac{0.5 \times 10^{e-t}}{0.1 \times 10^{e}} = \frac{1}{2} \times 10^{-t+1}$$

visto que 0.1 é o menor valor possível para f_x

Erro de arredondamento

- Se $|gx| \geq ½$:

$$|EA_x| = |x - \bar{x}| = |(f_x \times 10^e + g_x \times 10^{e-t}) - (f_x \times 10^e + 10^{e-t})|$$

$$= |g_x \times 10^{e-t} - 10^{e-t}| = |(g_x - 1)| \times 10^{e-t} \leq \frac{1}{2} \times 10^{e-t}$$

visto que $(g_x – 1) < ½$ pois $|g_x| \geq ½$

$$|ER_x| = \frac{|EA_x|}{|\bar{x}|} \leq \frac{\frac{1}{2}\times 10^{e-t}}{|f_x \times 10^e + 10^{e-t}|} < \frac{\frac{1}{2}\times 10^{e-t}}{|f_x| \times 10^e} < \frac{\frac{1}{2}\times 10^{e-t}}{0.1 \times 10^e} = \frac{1}{2} \times 10^{-t+1}$$

visto que $|f_x \times 10^e + 10^{e-t}| > |f_x \times 10^e|$ (denominador)

- Portanto, em qualquer caso teremos:

$$|EA_x| \leq \frac{1}{2} \times 10^{e-t} \quad |ER_x| < \frac{1}{2} \times 10^{-t+1}$$

Observações

- Erro:
  - Erro de arredondamento = E_A
  - Erro de truncamento = E_T
    - **$E_A < E_T$**
- Tempo:
  - Tempo de execução do arredondamento = T_A
  - Tempo de execução do truncamento = T_T
    - **$T_A > T_T$**
    - **truncamento é mais utilizado**

Análise de Erros nas Operações

- Dada sequência de operações, por exemplo:
  - $u = x + y - z$
- É preciso ter uma noção de como o erro se propaga ao longo das operações realizadas
- O erro total em uma operação é composto pelo erro nas parcelas ou fatores da operação e pelo erro no resultado da operação
- Nos exemplos a seguir, será utilizado um sistema de ponto flutuante de 4 dígitos, na base 10, com acumulador de precisão dupla

Cálculo da Adição

- Requer o alinhamento dos pontos decimais dos dois números dados
  - Desloca-se a mantissa de menor expoente para a direita para realizar esse alinhamento
    - O deslocamento de casas decimais é igual à diferença entre os dois expoentes dos números considerados
  - Exemplo de alinhamento em dois números:
    - $x = 0.937 \times 10^4$; $y = 0.1272 \times 10^2$
    - Alinhando-se os pontos decimais:
    - $x = 0.937 \times 10^4$ e $y = 0.001272 \times 10^4$

Cálculo da Adição

- Exemplo:
  - x = 0.937 x 10^4 ; y = 0.1272 x 10^2, calcular x+y:
  - Alinhando-se os pontos decimais tem-se que:
    - **x = 0.937 x 10^4 e y = 0.001272 x 10^4 (números alinhados)**
    - **x+y = (0.937 + 0.001272) x 10^4 = 0.938272 x 10^4 (exato)**
  - Como t = 4, o resultado deve ser truncado ou arredondado dependendo do que se deseja:
    - **arredondamento: $\overline{x+y}$ = 0.9383 x 10^4**
    - **truncamento: $\overline{x+y}$ = 0.9382 x 10^4**

Cálculo da Multiplicação

- Não requer alinhamento dos pontos decimais dos dois números dados
  - Basta realizar a multiplicação dos números
  - Depois ajusta-se o resultado da multiplicação
  - O resultado é ajustado pela base e mantissa
  - Assim como na adição pode-se ter 2 opções:
    - Truncamento
    - Arredondamento

Cálculo da Multiplicação

- Exemplo:
  - $x = 0.937 \times 10^4$ e $y = 0.1272 \times 10^2$, calcular xy:
    - **$xy = (0.937 \times 10^4) \times (0.1272 \times 10^2)$**
      $= (0.937 \times 0.1272) \times 10^6$
      $= 0.1191864 \times 10^6$
  - Como t = 4, o resultado deve ser truncado ou arredondado dependendo do que se deseja:
    - **arredondamento: $\overline{xy} = 0.1192 \times 10^6$**
    - **truncamento: $\overline{xy} = 0.1191 \times 10^6$**

Erro relativo de uma operação

- Mesmo que as parcelas ou fatores de uma operação estejam representados exatamente no sistema, não se pode esperar que o resultado armazenado seja exato
- Normalmente, o resultado exato da operação (OP) é normalizado e depois arredondado ou truncado para t dígitos, obtendo-se então o resultado aproximado $\overline{OP}$
- Baseando-se no cálculo de erro relativo anterior e supondo que as parcelas ou fatores não contêm erro, o erro relativo de qualquer operação será dado pelas expressões abaixo:

No truncamento: $|ER_{OP}| < 10^{-t+1}$

No arredondamento: $|ER_{OP}| < \frac{1}{2} \times 10^{-t+1}$