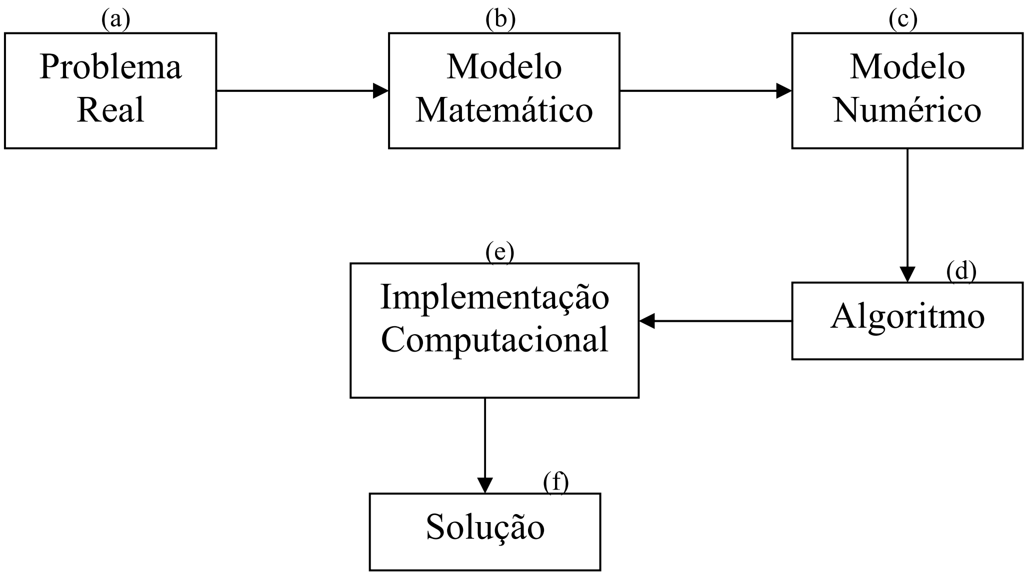
REPRESENTAÇÃO NUMÉRICA E ERROS

Introdução e Definições

Fases do processo de solução de um problema real:



- Na passagem de cada fase para a fase seguinte é inevitável que erros sejam incorporados.
    - Por exemplo, a transformação do problema real em um modelo matemático introduz erros devido à desconsideração de fenômenos com grau de incerteza elevado (resistência do ar, velocidade do vento, etc.).
- Já nas transformações entre as etapas designadas por (b), (c), (d) e (e) existe um outro tipo de erro associado:
    - O erro numérico. Esse erro depende fundamentalmente do tipo de representação numérica, bem como do volume de cálculos efetuado.

O erro numérico pode ser formalmente definido como:

Ev = Valor verdadeiro – Valor aproximado

podendo ser classificado como:

- Erro de truncamento
- Erro de arredondamento

O *erro de truncamento* é decorrente da representação de um processo infinito através de um processo finito.

Por exemplo:

A avaliação de funções "implícitas" em computadores tais como exponencial, funções trigonométricas, etc., é realizada através do seu desenvolvimento em série de Taylor:

$$f(x+h) = f(x) + f'(x)h + f''(x)\frac{h^2}{2!} + f'''(x)\frac{h^3}{3!} + \cdots + f^{(n)}(x)\frac{h^n}{n!}$$

O *erro de arredondamento* é proveniente da representação finita de um número em um computador.

O arredondamento pode ser efetuado de duas formas:

- Descarte, ou;
- Assumindo o número significativo mais próximo.

A representação científica de um número é feita na forma:

$$x \to m \times b^e$$

onde:

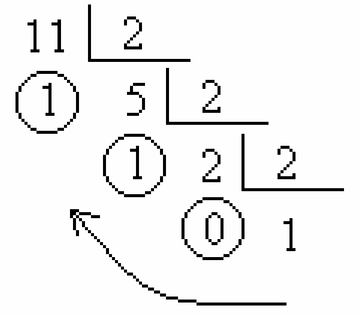
m → mantissa

b → base

e → expoente

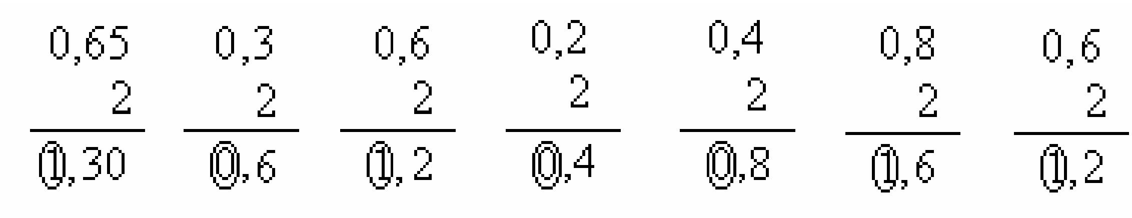
Mudança de Base

a) $11_{10} = ?_2$



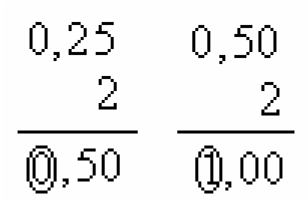
$\Rightarrow 11_{10} = 1011_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$

b) $0{,}65_{10} = ?_2$



$\Rightarrow 0{,}65_{10} = 0{,}101001 \cdots_2$

c) $11{,}25_{10} = 11 + 0{,}25$



$\Rightarrow 0{,}01_2$

$11{,}25_{10} = 1011_2 + 0{,}01_2 = 1011{,}01_2$

$$1011{,}01_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} =$$
$$= 8 + 0 + 2 + 1 + 0 + 0{,}25 = 11{,}25_{10}$$

Representação Numérica

Representação Geral

$$x = \pm\left[\frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \cdots + \frac{d_t}{\beta^t}\right] \cdot \beta^e$$

onde:

$d_i \longrightarrow$ inteiros / $0 \le d_i \le \beta - 1$; $i = 1,2,\cdots,n$

$t \longrightarrow$ nº de dígitos significativos

$e \longrightarrow$ expoente ; $e_i \le e \le e_s$

Ex1: Para $\beta = 10$

$$0{,}415_{10} = +\left[\frac{4}{10} + \frac{1}{10^2} + \frac{5}{10^3}\right] \cdot 10^0$$

$$41{,}5_{10} = 0{,}415_{10} \times 10^2 = \left[\frac{4}{10} + \frac{1}{10^2} + \frac{5}{10^3}\right] \cdot 10^2$$

Ex2: A representação binária do decimal 5 é:

$$101_2 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 5_{10}$$

De acordo com a representação geral, tem-se:

$$101_2 = 0{,}101_2 \times 2^3 = \left[\frac{1}{2} + \frac{0}{2^2} + \frac{1}{2^3} \right] \cdot 2^3$$

O expoente 3 pode ainda ser escrito na forma binária:

$$3_{10} = 11_2$$

Assim, o número 5 pode ser caracterizado, na forma binária, pelo vetor:

$$\underbrace{\begin{array}{|c|c|c|c|}\hline \pm & 1 & 0 & 1 \\ \hline \end{array}}_{\text{mantissa}}\underbrace{\begin{array}{|c|c|c|}\hline \pm & 1 & 1 \\ \hline \end{array}}_{\text{exp oente}}$$

O sinal + ou – pode ser caracterizado por mais um dígito (bit); adota-se 0 para + e 1 para -. Assim:

$$\underbrace{\begin{array}{|c|c|c|c|}\hline 0 & 1 & 0 & 1 \\ \hline \end{array}}_{mantissa}\underbrace{\begin{array}{|c|c|c|}\hline 0 & 1 & 1 \\ \hline \end{array}}_{\exp oente}$$

Ex3: Uma calculadora possui um sistema de representação numérica de base $\beta = 2$, com t = 10 algarismos significativos da mantissa e expoentes inferior e superior e_i = -15 e e_s = 15, respectivamente. Verificar o nº de bits necessários à representação de um número, p. ex., 25.

$$25_{10} = 11001_2 = 0{,}11001_2 \times 2^5 = 0{,}\overbrace{11001}^{mant}{}_2 \times 2^{\overbrace{101}^{exp}}$$

O máximo valor absoluto de expoente a ser representado é $15_{10} = 1111_2$. Portanto, para o expoente, devem ser reservados 04 bits e mais 01 para o seu sinal. Contando com o bit do sinal da mantissa e com os 10 algarismos significativos, tem-se um total de 16 bits. A representação do número 25 ficaria:

0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1
sin *al*											sin *al*				

$$0{,}1100100000 \times 2^{0101}$$

O maior número com representação possível nessa máquina é:

$$0{,}1111111111 \cdot 2^{1111} = 32736_{10}$$

Arredondamento

É importante observar que, nem todos os números reais podem ser representados através do sistema acima.

Por exemplo, o número 17 pode ser representado na seguinte forma: $17_{10} = 10001 = 0{,}10001 \times 2^{101}$

0	1	0	0	0	1						0	0	1	0	1

O número, imediatamente superior a 17, que pode ser representado nesse sistema é: $0{,}10001\,00001 \times 2^{101}$

0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1

que corresponde a $17 + \dfrac{2^5}{2^{10}} = 17{,}0313$.

Portanto, o número 17,03 não pode ser representado de forma exata. Nesse caso a representação é feita assumindo-se o número significativo mais próximo: $17 + 1/2^5$

Precisão

Define o número de casas decimais exatas da mantissa. É determinada pelo último bit da mantissa. No caso acima, tem-se:

$$\text{precisão} \leq \frac{1}{\beta^t} = \frac{1}{2^{10}} \approx 10^{-3}$$

EXERCÍCIO

1) Uma máquina de 16 bits possui β = 4 e *t* = 10. Determinar:

a) O maior número que pode ser representado na máquina;

b) A sua precisão.

c) Representar o número 15,25 nessa máquina.

Solução:

$$\left.\begin{array}{l}16\text{ bits}\\ t=10\end{array}\right\}\Rightarrow$$

Logo, como $\beta = 4$ ($\beta - 1 = 3$), os valores máximo e mínimo para o expoente serão, respectivamente:

											0	3	3	3	3
											1	3	3	3	3

e, o maior número que pode ser representado nesta máquina é:

0	3	3	3	3	3	3	3	3	3	3	0	3	3	3	3

que equivale:

$$x_{máx} = \left[\frac{3}{4} + \frac{3}{4^2} + \frac{3}{4^3} + \frac{3}{4^4} + \frac{3}{4^5} + \frac{3}{4^6} + \frac{3}{4^7} + \frac{3}{4^8} + \frac{3}{4^9} + \frac{3}{4^{10}}\right] \cdot 4^e \Rightarrow$$

$$x_{máx} = \left[3 \cdot 4^9 + 3 \cdot 4^8 + 3 \cdot 4^7 + 3 \cdot 4^6 + 3 \cdot 4^5 + 3 \cdot 4^4 + 3 \cdot 4^3 + \right.$$
$$\left. + 3 \cdot 4^2 + 3 \cdot 4^1 + 3 \cdot 4^0\right] \cdot 4^{-10} \cdot 4^e \Rightarrow$$

$$x_{máx} = [1.048.575] \cdot 4^{-10} \cdot 4^e \approx 1 \cdot 4^e \Rightarrow$$

$$\boxed{x_{máx} \approx 4^e}$$

onde:

$$e = \left[\frac{3}{4} + \frac{3}{4^2} + \frac{3}{4^3} + \frac{3}{4^4}\right] \cdot 4^4 = \left[3 \cdot 4^3 + 3 \cdot 4^2 + 3 \cdot 4^1 + 3 \cdot 4^0\right] \cdot 4^{-4} \cdot 4^4 \Rightarrow$$

$$\boxed{e = 255 \cdot 4^{-4} \cdot 4^4 = 255}$$

logo:

$$\boxed{x_{máx} \approx 4^{255} = 3{,}3520 \times 10^{153}}$$

A precisão da máquina em questão é dada por:

$$\boxed{\text{precisão} \leq \frac{1}{\beta^t} = \frac{1}{4^{10}} = 9{,}5367 \times 10^{-7} \approx 10^{-6}}$$

O número 15,25 é representado nessa máquina, por:

$$\boxed{\begin{array}{l|l} 15 & 4 \\ \hline 3 & 3 \end{array} \Rightarrow 15_{10} = 33_4} \text{; e } \boxed{\begin{array}{l} 0{,}25 \\ \times\ \ 4 \\ \hline 1{,}00 \end{array} \Rightarrow 0{,}25_{10} = 0{,}1_4}$$

logo:

$$\boxed{15{,}25_{10} = 33{,}1_4 = \left(0{,}331 \times 4^2\right)_4}$$

0	3	3	1	0	0	0	0	0	0	0	0	0	0	0	2