Sistemas de Numeração e Erros

Os números representáveis em qualquer máquina são finitos, ou melhor, não é possível representar em um computador todos os números de um dado intervalo [a, b]. O resultado de um simples cálculo de uma função, realizado com esses números, podem conter erros.

Esses erros causados podem diminuir e, algumas vezes, destruir a precisão dos resultados.

Representação de um Número Inteiro

Assim dado um número inteiro $n \neq 0$, ele possui uma única representação:

$$n = \pm(n_k\ n_{k-1}\ n_{k-2} \ldots n_1\ n_0) = \pm(n_0\beta^0 + n_1\beta^1 + \ldots + n_{k-2}\beta^{k-2} + n_{k-1}\beta^{k-1} + n_k\beta^k)$$

onde

β é um número inteiro ≥ 2, chamada base.

n_i são inteiros tal que $0 \leq n_i < \beta$ e $n_k \neq 0$

Exemplos:

o número 1967 é representado na base 10 como:

$$(1967)_{10} = 7x10^0 + 6x10^1 + 9x10^2 + 1x10^3 = (1967)_{10}$$

já o número 1101 na base 2 é:

$$(1101)_2 = 1x2^0 + 0x2^1 + 1x2^2 + 1x2^3 = 1 + 0 + 4 + 8 = (13)_{10}$$

Representação de um Número Real

No computador, existem duas representações de um número real: Ponto Fixo e Ponto Flutuante.

a) Ponto Fixo

Um número $x \neq 0$ será representado em ponto fixo assim:

$$x = \pm\sum_{i=k}^{n} x_i\beta^{-i}$$

onde k e n são inteiros sendo $k<n$ e usualmente $k \leq 0$ e $n>0$ e os x_i são inteiros tal que $0 \leq x_i < \beta$

Exemplo:

$$1967{,}25 = \sum_{i=-3}^{2} x_i\beta^{-i} = 1x10^3 + 9x10^2 + 6x10^1 + 7x10^0 + 2x10^{-1} + 5x10^{-2}$$

Esse sistema foi usado por muitos computadores no passado e, hoje, não é mais utilizado.

b) Ponto Flutuante

Um número $x \neq 0$ será representado em ponto flutuante assim:

$$x = \pm d \times \beta^e$$

onde $\boldsymbol{\beta}$ é a base, **d** é a mantissa e **e** é o expoente. A mantissa, **d**, é um número em ponto fixo:

$$d = \pm\sum_{i=k}^{t} d_i\beta^{-i}$$

onde **k** é geralmente igual a 1, tal que $x \neq 0$, então $d_i \neq 0$ (forma normalizada). A quantidade de dígitos é igual a **t**, com $\beta^{-1} \leq d < 1$.

Exemplos:

a) $0{,}35 = (3x10^{-1} + 5x10^{-2})x10^0 = 0{,}35x10^0$

b) $-5{,}172 = (5x10^{-1} + 1x10^{-2} + 7x10^{-3} + 2x10^{-4})x10^1 = -0{,}5172x10^1$

c) $0{,}0123 = (1x10^{-1} + 2x10^{-2} + 3x10^{-3})x10^{-1} = 0{,}123x10^{-1}$

d) $0{,}0003 = (3x10^{-1})x10^{-3} = 0{,}3x10^{-3}$

e) $5391{,}3 = (5x10^{-1} + 3x10^{-2} + 9x10^{-3} + 1x10^{-4} + 3x10^{-5})x10^4 = 0{,}53913x10^4$

Para simplificar a representação de um sistema de números em ponto flutuante normalizado, na base β, com t dígitos significativos e com limites do expoente m e M , usaremos a notação: F(β, t,m,M).
Assim um número em F(β, t, m,M) será representado por:

$$\pm 0.d_1 d_2 d_3 \ldots d_t \times \beta^e$$

onde $d_1 \neq 0$ e $-m \leq e \leq M$.

Exemplo - Considere o sistema F(10, 3, 2, 2). Represente nesse sistema os números:

a) $0{,}35 = 0{,}350x10^0$

b) $-5{,}172 = -0{,}517x10^1$

c) $0{,}0123 = 0{,}123x10^{-1}$

d) $0{,}0003 = \left(3x10^{-1}\right)x10^{-3} = \nexists$ no F(10,3,2,2)

e) $5391{,}3 = \left(5x10^{-1} + 3x10^{-2} + 9x10^{-3} + 1x10^{-4} + 3x10^{-5}\right)x10^4 = \nexists$ no F(10,3,2,2)

Observe que o número 0,0003 não pode ser representado no sistema, pois o expoente é menor que -2 causando **underflow**. Já o número 5391,3 não pode ser representado no sistema, pois o expoente é maior que 2, causando **overflow.**

Mudança de Base

Um número pode ser representado em mais de uma base. Através de uma mudança de base, é possível determinar a representação em uma nova base.

Exemplo: Mudar a representação dos números:

a) 1101 da base 2 para a base 10

Solução: $(1101)_2 = 1x2^0 + 0x2^1 + 1x2^2 + 1x2^3 = 1 + 0 + 4 + 8 = (13)_{10}$

b) 0,110 da base 2 para a base 10

Solução: $(0{,}110)_2 = 1x2^{-1} + 1x2^{-2} + 0x2^{-3} = \frac{1}{2} + \frac{1}{4} + 0 = (0{,}75)_{10}$

c) 13 da base 10 para a base 2

$$\begin{array}{lllll}
13 & \underline{|\,2} & & & \\
\mathbf{1} & 6 & \underline{|\,2} & & \\
 & \mathbf{0} & 3 & \underline{|\,2} & \\
 & & \mathbf{1} & 1 & \underline{|\,2} \\
 & & & \mathbf{1} & 0
\end{array}$$

Solução: O número na base 2 será obtido tomando-se todos os restos das divisões: $(13)_{10} = (1101)_2$

d) 0,75 da base 10 para a base 2

$$\begin{array}{lll}
0{,}75 & \times 2 = & \mathbf{1}{,}50 \\
0{,}50 & \times 2 = & \mathbf{1}{,}00 \\
0{,}00 & \times 2 = & \mathbf{0}{,}00
\end{array}$$

Solução: O número na base 2 será obtido tomando-se a parte inteira do resultado de cada multiplicação. Assim: $(0{,}75)_{10} = (0{,}110)_2$

e) 3,8 da base 10 para a base 2 $\quad (3{,}8)_{10} = (11{,}11001100\ldots)_2$

f) 12,20 da base 4 para a base 3 $\quad (12{,}20)_4 = (20{,}111\ldots)_3$

Exercícios

1 – Mudar a representação do número:

i)	413 da base 10,	para a base 2
ii)	724128 da base 10,	para a base 16
iii)	F0CA da base 16,	para a base 10
iv)	1120,3 da base 4,	para a base 8

Um número **F(β, t, m, M)** será representado por $\pm 0.d_1 d_2 d_3 ... d_t \times \beta^e$ onde $d_1 \neq 0$ e $-m \leq e \leq M$

2 – Considere o sistema F(10, 3, 5, 5). Efetue as operações indicadas:

i) $(1{,}386 - 0{,}987) + 7{,}6485$ e $1{,}386 - (0{,}987 - 7{,}6485)$

ii) $\dfrac{1{,}338 - 2{,}038}{4{,}577}$ e $\dfrac{1{,}338}{4{,}577} - \dfrac{2{,}038}{4{,}577}$

3 – Seja

$$x = \frac{17{,}678}{3{,}471} + \frac{(9{,}617)^2}{3{,}716 \times 1{,}85}$$

a) Calcule x com todos os algarismos da sua calculadora, sem efetuar arredondamento.

b) Calcule x considerando o sistema F(10, 3, 4, 3). Faça arredondamento a cada operação efetuada.

4 – Seja $P(x) = 2{,}3x^3 - 0{,}6x^2 + 1{,}8x - 2{,}2$. Deseja-se obter o valor de P(x) para x = 1,61.

a) Calcule P(1,61) com todos os algarismos da sua calculadora, sem efetuar arredondamento.

b) Calcule P(1,61) considerando o sistema F(10, 3, 4, 3). Faça arredondamento a cada operação efetuada.

5 – Resolver a equação $x^2 - 1634x + 2 = 0$

Efeitos Numéricos

1) Cancelamento – ocorre na subtração de dois números quase iguais.

calcule $\sqrt{9876} - \sqrt{9875}$

2) Propagação de Erro – ocorre numa soma quando uma soma parcial é muito grande quando comparada com o resultado final.

calcule $e^{-5,25}$ sabendo que $e^{-x} = \sum_{k=0}^{\infty} (-1)^k \dfrac{x^k}{k!}$

3) Instabilidade Numérica – ocorre quando um resultado intermediário é contaminado por um erro de arredondamento, influenciando todos os resultados subseqüentes.

calcule $I_n = e^{-1} \int_0^1 x^n e^x dx$ sabendo que $I_n = 1 - n\, I_{n-1}$

4) Mal Condicionamento – ocorre quando pequenas perturbações nos dados alteram de modo significativo o resultado final.

calcule $\begin{cases} x + \quad y = 2 \\ x + 1{,}01y = 2{,}01 \end{cases}$