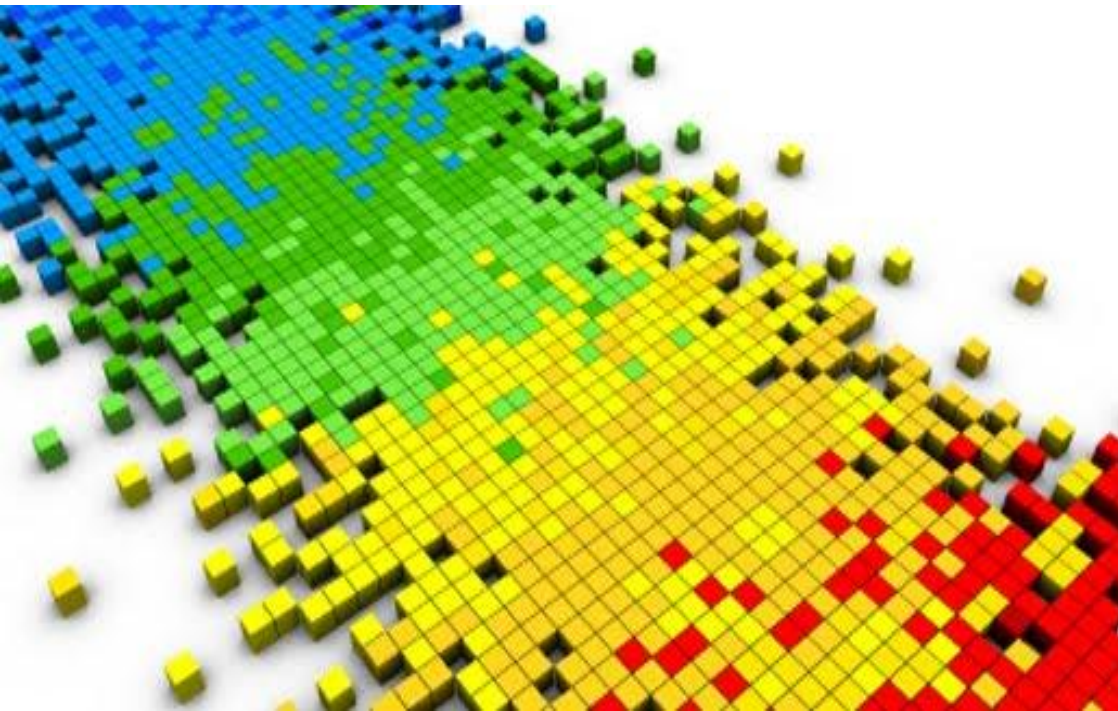




Cálculo Numérico

Prof. Guilherme Amorim

24/10/2013

Aula 2 – Erros e Aritmética de Ponto Flutuante

The background of the slide is a dense, overlapping field of 3D-rendered numbers (0-9) in a light gray color. The numbers are scattered across the entire frame, creating a textured, mathematical background. A semi-transparent blue rectangular box is centered horizontally and vertically, containing the title text in orange.

Noções de Aritmética de Máquina

Representação de Números...

$\Pi =$

3.141592653589793238462643383279502884197169399375105820974944
592307816406286208998628034825342117067982148086513282306647
093844609550582231725359408128481117450284102701938521105559
644622948954930381964428810975665933446128475648233786783165
271201909145648566923460348610454326648213393607260249141273
724587006606315588174881520920962829254091715364367892590360
011330530548820466521384146951941511609433057270365759591953
092186117381932611793105118548074462379962749567351885752724
891227938183011949129833673362440656643086021394946395224737
190702179860943702770539217176293176752384674818467669405132
000568127145263560827785771342757789609173637178721468440901
224953430146549585371050792279689258923542019956112129021960
864034418159813629774771309960518707211349999998372978049951
059731732816096318595024459455346908302642522308253344685035
261931188171010003137838752886587533208381420617177669147303
598253490428755468731159562863882353787593751957781857780532
171226806613001927876611195909216420198938095257201065485863
278865936153381827968230301952035301852968995773622

Quantas decimais terá o número π ?

- Infinitas!
- O que poderíamos fazer para representar π num computador?
- Um computador é algo intrinsecamente finito em recursos, certo?

Representação de números..

- Um computador pode representar um conjunto finito dos números racionais.
- São chamados números de PONTO FLUTUANTE (Floating Point Numbers - *Floats*)
- Logo, cada operação realizada a partir de pontos flutuantes leva a resultados aproximados..

E se falamos de aproximações...

- ❑ Precisamos pensar em **ERROS**..
- ❑ E erros provocados por computador podem causar sérios danos..
- ❑ Exemplo: Ariane 5



Erros



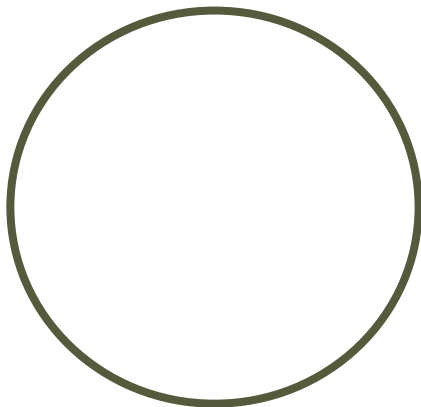
- O que são?
- Como são calculados?
- Para que servem?

Tipos de Erros

- ❑ Inerentes
- ❑ Truncamentos
- ❑ Arredondamentos

Erros Inerentes

- “São erros que o usuário não tem condições de evitá-los. Eles surgem de modelos matemáticos, medidas, etc.”
- Exemplo:
 - ▣ Calcular o comprimento de uma circunferência.



$$C = 2\pi r$$

Como π é irracional, C não pode ser calculado exatamente em um computador.

Erros de Truncamento

- “São erros que surgem quando substituímos um processo matemático infinito por uma parte finita dele, pois, na implementação de algoritmos numéricos em um computador, podemos realizar apenas um número de operações aritméticas.”
- Trocamos uma série infinita por uma finita.
- Exemplo:

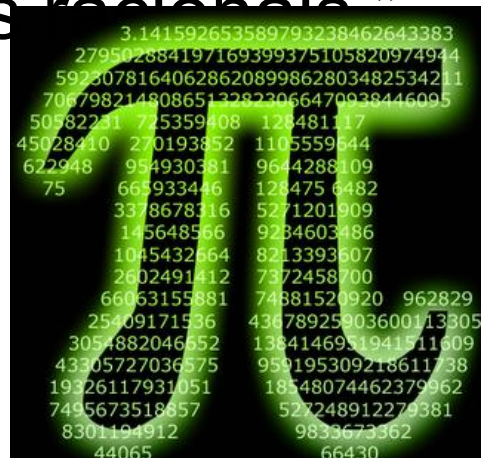
$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad \longrightarrow \quad e^x \cong \sum_{k=0}^{n-1} \frac{x^k}{k!}$$

Para um número natural n de parcelas “conveniente”

Erros de Arredondamento

- “São cometidos pelos computadores ao realizarem operações aritméticas. Isto se deve ao fato de que um computador possui uma palavra (local onde armazena dados) de tamanho finito e conseqüentemente só consegue representar um subconjunto finito de números reais.”

- Exemplo:



```
3.141592653589793238462643383
279502884197169399375105820974944
59230781640628620899862803482534211
70679821480865132823066470938446095
50582231 725359408 128481117
45028410 270193852 1105559644
622948 954930381 9644288109
75 665933446 128475 6482
3378678316 5271201909
145648566 9284603486
1045432664 8213393607
2602491412 7372458700
66063155881 74881520920 962829
25409171586 43678925903600113805
3054882046652 1384146931941511609
43305727036575 959195309218611738
19326117931051 18548074462379962
7495673518857 527248912279381
8301194912 9833673362
44065 66430
```

Representação e cálculo de Erros

- Precisamos de medidas para representar erros.
- As mais comuns são:
 - ▣ Erro absoluto
 - ▣ Erro relativo
 - ▣ Erro Percentual

Erros

- **Erro absoluto**, denotado por $\Delta \bar{x}$ é a diferença entre o valor exato x e o valor aproximado $\bar{x}$, isto é, $\Delta \bar{x} = x - \bar{x}$.

- **Erro relativo**, denotado por $\delta \bar{x}$, é o quociente entre o erro absoluto $\Delta \bar{x}$ e o valor exato x ($x \neq 0$), isto é,

$$\delta \bar{x} = \Delta \bar{x} / x = (x - \bar{x}) / x.$$

- **Percentual**, denotado por $p \bar{x}$, é apenas o produto do erro relativo por 100, dando-lhe a representação percentual

$$p \bar{x} = 100 \times \delta \bar{x} \%.$$

Erros – Exemplo 1

- Em 10.000 itens a serem contados, foram encontrados 9.999.
- Valor exato: $x = 10000$
- Valor aproximado: $\bar{x} = 9999$
- Erro absoluto $\Delta \bar{x} = 10000 - 9999 = 1$;
- Erro relativo: $\delta \bar{x} = 1 / 10000 = 0,0001$
- Erro percentual $p \bar{x} = 100 \times 0,0001\% = 0,01 \%$.

Erros – Exemplo 2

- Em 10 itens a serem contados, foram encontrados 9.

$$x = 10$$

- Valor exato: $\bar{x} = 9$

- Valor aproximado: $\Delta \bar{x} = 10 - 9 = 1$

- Erro absoluto: $\delta \bar{x} = 1 / 10 = 0,1$

- Erro relativo: $p \bar{x} = 100 \times 0,1 \% = 10 \%$

- Erro percentual:

Notar que...

- Em ambos os casos, os erros absolutos foram iguais
- Os erros relativos e percentuais, entretanto, foram bastante distintos:

Exemplo 1

$$\delta \bar{x} = 1 / 10000 = 0,0001$$

$$p \bar{x} = 100 \times 0,0001\% = 0,01 \%$$

Exemplo 2

$$\delta \bar{x} = 1 / 10 = 0,1$$

$$p \bar{x} = 100 \times 0,1 \% = 10 \%$$

- Portanto, o erro relativo (e percentual) são mais adequados

Revisão – Mudança de base

- **Bases:** um número pode ser representado em diferentes bases...
- As mais tradicionais são binária, decimal, hexadecimal..
- Exemplos:

$$11_{10} = 1011_2$$

$$\begin{aligned} &\rightarrow 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = \\ &= 8 + 0 + 2 + 1 = \\ &= 11 \end{aligned}$$

$$25_{10} = ?_2$$

$$\begin{array}{r} 25 \overline{) 12} \\ 1 \\ \underline{12} \\ 0 \\ 0 \\ \underline{0} \\ 0 \\ \underline{0} \\ 1 \end{array} \Rightarrow 11001_2$$

Revisão – Mudança de base

- E no caso de números fracionários (base 2 para a base 10)...

Multiplicar cada algarismo do numero na base 2, após o ponto, por potências decrescentes de 2, da esquerda para a direita e somar as parcelas.

0,110 na base 2 para a base 10

$$0,110 = 1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3}$$

$$\frac{1}{2} + \frac{1}{4} + 0 = 0,75$$

Revisão – Mudança de base

- E no caso de números fracionários (base 10 para a base 2)...

Multiplique a parcela decimal por 2. Continue multiplicando a parte decimal do resultado obtido por 2. O número na base 2 será então obtido tomando-se a parte inteira do resultado de cada multiplicação.

0,75 na base 10 para a base 2

$$0,75 \times 2 = 1,50$$

$$0,50 \times 2 = 1,00$$

$$0,00 \times 2 = 0,00$$

$$\text{Logo, } 0,75_{10} = 0,110_2$$

Revisão – Mudança de base

- Mais um exemplo...

3,75 na base 10 para a base 2

Separamos a parte inteira da parte decimal.

3 na base 10 para a base 2

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

$$\text{Logo, } 3,75_{10} = 11,110_2$$

0,75 na base 10 para a base 2

$$0,75 \times 2 = 1,50$$

$$0,50 \times 2 = 1,00$$

$$0,00 \times 2 = 0,00$$

$$\text{Logo, } 0,75_{10} = 0,110_2$$

Representação dos Números

- Quantas formas diferentes temos de representar o número 0,3?

$$\frac{3}{10}$$

$$\frac{6}{20}$$

$$0,3 \times 10^0$$

$$3 \times 10^{-1}$$

$$30 \times 10^{-2}$$

Representação dos Números

- E qual a forma mais utilizada de apresentar o número 0,3 no computador?

$$3 \times 10^{-1}$$

Aritmética de Ponto Flutuante

- Definição: Um número real x ($x \in \mathbb{R}$) é dito um número de máquina (de ponto flutuante normalizado) se são válidos os seguintes itens:

1) $x = m \times b^e$;

2) $m = \pm d_0, d_1 d_2 \dots d_{t-1}$, $t \in \mathbf{N}$ (Naturais);

3) $d_0 \neq 0$, $d_i \in \{0, 1, 2, \dots, (b-1)\}$, $i = 1, 2, \dots, (t-1)$;

4) $e_1 \leq e \leq e_2$, sendo $e_1 \leq 0$ e $e_2 \geq 1$, $e_1, e_2 \in \mathbf{Z}$ (Inteiros),

onde:

Aritmética de Ponto Flutuante

b – é a base, $b \geq 2$, $b \in \mathbf{N}$;

e – é o expoente, $e \in \mathbf{Z}$;

e_1 – é o menor expoente;

e_2 – é o maior expoente ;

m – é o significando normalizado, isto é, possui $d_0 \neq 0$,
portanto $1 \leq |m| < b$;

t – é o número de dígitos significativos usados na representação do significando normalizado;

d_i – são dígitos da base b que formam o significando.

Logo..

$$x = m b^e$$

- A representação em forma de ponto flutuante de um número real x usa uma base $b \geq 2$, t dígitos significativos (ou precisão) e um expoente e .

Exemplos

- Represente os seguintes números em ponto flutuante $b=10$ e $t=5$:

- π $\pi = 3,1415926 \dots = 3,141\textbf{6} \times 10^0$

- $2/7$ $\frac{2}{7} = 0,28571428 \dots = 2,8571 \times 10^{-1}$

- $0,01523$ $0,01523 = 1,5230 \times 10^{-2}$

- $\sqrt{2}$ $\sqrt{2} = 1,41421356 \dots = 1,4142 \times 10^0$

Observações

- A representação de ponto flutuante na forma normalizada ($d_0 \neq 0$) de um número real x é única.
- 0 (zero) é um número de máquina especial representado por $\pm 0,000 \dots 0 \times b^{e_1}$;
- Se x é um número de máquina, então $-x$ também o é;
- Os dígitos $d_0, d_1, d_2, \dots, d_{t-1}$, de m são ditos os t primeiros dígitos significativos de x .

Observações

- Se $d_r = d_{r+1} = d_{r+2} = d_{t-1} = 0$, $r \geq 1$, ainda assim estes dígitos são considerados significativos;
- O menor número de máquina positivo é:
 - ▣ $x_{min} = 1,000 \dots 0 \times b^{e_1}$
- O maior número de máquina positivo é:
 - ▣ $x_{max} = (b - 1), (b - 1)(b - 1) \times b^{e_2}$
- Logo, a região dos números reais que pode ser representada em um computador é dada por:
 - ▣ $\mathbb{R} = [-x_{max}; -x_{min}] \cup \{0\} \cup [x_{min}, x_{max}]$

Exemplo

b, t, e₁, e₂	x_{min}	x_{max}
10, 5, -5, 5	$1,0000 \times 10^{-5}$	$9,9999 \times 10^5$
2, 4, -10, 10	$1,000 \times 2^{-10}$	$1,1111 \times 2^{10}$

Sistema de Ponto Flutuante

O conjunto de todos os números de máquina é chamado de sistema de ponto flutuante, geralmente representado por $F(b, t, e_1 \text{ e } e_2)$.

Exemplo

- Seja uma máquina que trabalha com o sistema de ponto flutuante $F(10, 4, -9, 9)$.
- O número real 34,21 é um número desta máquina?
 - ▣ Sim. $3,421 \times 10^1$
- O número real 0,42162 é um número desta máquina?
 - ▣ Não. $4,216 \times 10^{-1}$

Proposição

□ **F é um subconjunto finito dos números racionais.**

□ Demonstração:

▣ Parte 1: $F \subset \mathbb{Q}$ (Conjunto dos Números Racionais)

a) Se $x = 0$, então $x \in \mathbb{Q}$

$$\begin{aligned} \text{b) Se } x &= \pm m \times b^e = \pm d_0, d_1 d_2 \dots d_{t-1} \times b^e \\ &= \pm \left(d_0 + d_1 b^{-1} + d_2 b^{-2} \dots + d_{t-1} b^{-(t-1)} \right) b^e \\ &= \pm \underbrace{\left(d_0 b^{t-1} + d_1 b^{t-2} + \dots + d_{t-1} \right)}_{b^{t-1}} b^e \end{aligned}$$

Quociente de inteiros $\in \mathbb{Q}$ $x \in \mathbb{Q}$

Proposição

- Demonstração (Parte 2): F é finito.
 - ▣ Podemos calcular a quantidade de números que podem ser representados, ou seja, o número de elementos de F .
 - ▣ Calculando a quantidade, garantimos que F é finito

$$N(F) = 2 \cdot (b-1) b^{t-1} \cdot (e_2 - e_1 + 1) + 1$$

Proposição

□ Demonstração (Parte 2): F é finito.

⊛ $d_0 : 1, \dots, b-1$

⊛ $d_1 d_2 \dots d_{t+1}$
 $b \geq b \geq \dots \geq b$

⊛ $e : e_1, \dots, e_2$

⊛ Sinal : $\oplus$ ou $\ominus$

⊛ Zero:

$$N(F) = 2 \cdot (b-1) b^{t-1} \cdot (e_2 - e_1 + 1) + 1$$

Exemplo

Exemplo 1.6 - Quantos elementos distintos existem no sistema de ponto flutuante $F(10, 4, -9, 9)$?

SOLUÇÃO: Temos que $b = 10$, $t = 4$, $e_1 = -9$ e $e_2 = 9$. Facilmente obtemos,

$$n(\mathbf{F}) = 2 \times (10 - 1) \times 10^{(4-1)} \times (9 - (-9) + 1) + 1 = 342001.$$

Definição

Dois números de máquina x_1 e x_2 , $x_1 < x_2$, são ditos consecutivos se e somente se entre x_1 e x_2 não existe outro elemento de máquina.

Proposição

- Fixado o expoente e , dois números consecutivos de uma máquina qualquer $x_1 = m_1 \times b^e$ e $x_2 = m_2 \times b^e$, $x_1 < x_2$, se diferenciam por: $b^{(e-t+1)}$

- Demonstração

$$\begin{aligned}x_2 - x_1 &= m_2 \cdot b^e - m_1 \cdot b^e \\&= (m_2 - m_1) \cdot b^e \\&= 0,000 \dots 01 \cdot b^e \\&= (1 \cdot b^{-(t-1)}) \cdot b^e \\&= b^{e-t+1}\end{aligned}$$


Exemplo

Exemplo 1.7 - Considere o sistema de ponto flutuante $F(10, 4, -9, 9)$. Determine a distância entre dois pares de números consecutivos:

a) $x_1 = 3,147 \times 10^{-9}$ e $x_2 = 3,148 \times 10^{-9}$;

b) $y_1 = 3,147 \times 10^9$ e $y_2 = 3,148 \times 10^9$,

SOLUÇÃO: No primeiro caso, temos que

$$x_2 - x_1 = (3,148 - 3,147) \times 10^{-9} = 0,001 \times 10^{-9} = 1 \times 10^{(-9-4+1)} = 10^{-12},$$

enquanto que no segundo os valores são: $y_2 - y_1 = 0,001 \times 10^9 = 1 \times 10^{(9-4+1)} = 10^6 = 1000000$.

Observação

- “Notar que em a e b são representados valores totalmente diferentes. No primeiro caso, temos uma distância de 10^{-12} , no segundo chegamos a 10^6 . Isto mostra que em uma máquina não existe uma distribuição uniforme de seus números, embora para um fixado expoente ela seja uniforme.”

