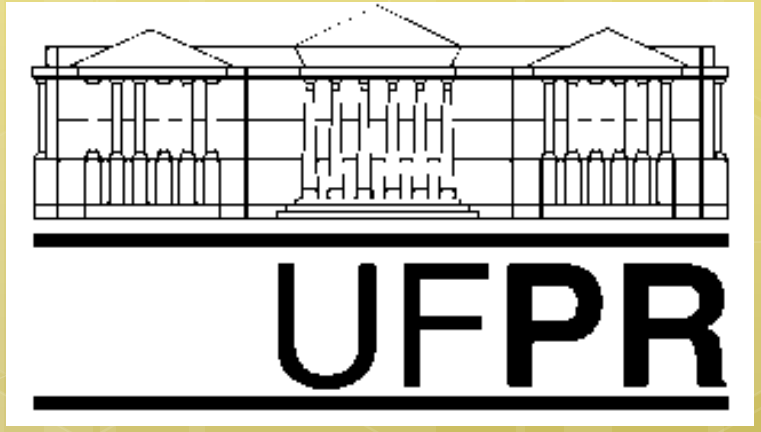


TE231
Capitulo 1 – Erros e Aritmética Computacional

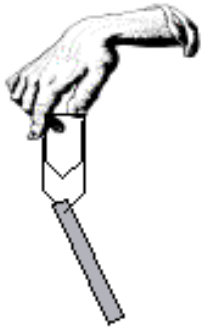
Prof. Mateus Duarte Teixeira

Sumário

1. Origem dos erros
2. Erros absolutos e relativos
3. Erros na mudança de base
4. Representação numérica de ponto flutuante
5. Convergência nos processos numéricos

1. Origem dos erros

ENTENDIMENTO DE UM PROBLEMA



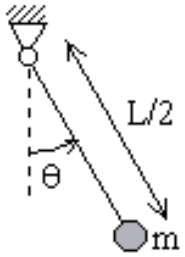
PROBLEMA REAL

Modelagem:

•Observação do fenômeno.

•Levantamento dos efeitos dominantes (idealizações).

•Referência aos conhecimentos prévios físicos e matemáticos.

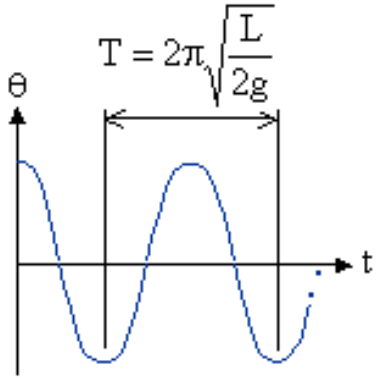


$$\frac{d^2\theta}{dt^2} + \frac{2g}{L}\,\mathrm{sen}\,\theta = 0$$

$$\theta << 1 \Rightarrow \frac{d^2\theta}{dt^2} + \frac{2g}{L}\theta = 0$$

MODELO MATEMÁTICO

•Permite analisar situações reais ou fazer previsões de comportamentos.



SOLUÇÃO

Resolução:

•Procedimentos analíticos ou numéricos.

•Exige uma confrontação com situações experimentais para validação do modelo matemático.

1. Origem dos erros

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.

Erro Inerente

- Erro sempre 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.

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.

Natureza dos Erros

- Geralmente os erros provém de três fontes:
  - Precisão dos dados;
  - Modelagem matemática do problema;
  - Erros de processamento numérico (na fase de resolução).

Precisão dos dados

- Relativos à imprecisão no processo de aquisição/entrada, externos ao processo numérico.
- Proveniência → Processo de aquisição/entrada (medidas experimentais)
  - Sujeitos às limitação/aferição dos instrumentos usados no processo de mensuração
  - Erros inerentes são inevitáveis!

Modelagem matemática do problema;

- 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

Modelagem matemática do problema

- 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 de processamento numérico

- Erros nos dados (experimentos): erros inerentes aos próprios instrumentos de medidas. Depende do tipo de aparelho utilizado.
  - Ex: Estufa de secagem de 40L com respeito ao controlador:
  - controlador on/off (variação de 15ºC)
  - controlador PID (variação de 1°C)

- Erros Inerentes: fora de controle (teórico), implementação numérica com dados errados do modelo matemático, cuidado ao usar hipóteses simplificativas;
- Erros por truncamento
- Erros por arredondamento

Erros computacionais (EC)

- Computadores nunca foram perfeitos e totalmente inteligentes. Nem mesmo quando se trata do desenvolvimento de suas aplicações.
- Muitas vezes, erros gerados por um programador em uma mera linha de código podem gerar grande caos aos sistemas computacionais.
- Os erros mais comuns são encontradas entre o teclado e a cadeira e até mesmo os pequenos erros podem custar bilhões de dólares para empresas ou nações.

Erros computacionais (EC)

- Em 1989, um EC resultou no envio de mais de 40.000 cartas aos cidadãos de Paris acusando-os de crimes como homicídio, extorsão, prostituição e crime organizado.
- O aeroporto de Denver com "sistema de bagagens automatizada" tinha 41,83 Km de pista subterrânea, tal sistema tinha vários erros computacionais em meados da década de 90, adiando a abertura do aeroporto a um custo de US$ 234 milhões
- Em 1994, uma linha de código fez um Banco, em NY, deduzir o dobro do montante que os clientes retiraram das máquinas ATM. Em torno de US$ 5 milhões foram retirados por engano nas contas dos clientes do banco.
- Em 1992 a Pepsi nas Filipinas ofereceu um prêmio de um milhão de pesos (cerca de R$ 85.000) para a sorte dos clientes que encontrassem o número 349 impresso na tampinha da garrafa. Devido a um "erro de software", 800000 tampinhas foram impressas com o número do vencedor em vez de apenas uma.
- A Scientific America em 1998 relatou um caso em que um membro da tripulação do *cruiser USS Yorktown* que faz o controle de mísseis guiados cometeu o trivial erro da "divisão por zero" gerando um brutal erro no software. Os erros em cascata aconteceram em todo cruiser, causando o desligamento do sistema de propulsão e de saída do Yorktown por várias horas.
- Em 1999, a Mars Climate Orbiter não foi para o espaço por uma falha computacional gerando um custo de 125 milhões dólares porque não havia um padrão para medição de dispositivos. Quando um módulo, no dispositivo, passava informações para o outro, o sistema não foi capaz de processá-las.
- O Northeast Blackout no USA em 2003 resultou na perda de eletricidade a mais de 50 milhões de lares. As perdas foram estimadas em 6 bilhões de dólares.
- Em 2006, um erro no sistema da Verizon levou à sobrecarga de cerca de 11.000 clientes na região mid-Atlantic. O erro de programação gerou um custo de cerca de US$ 200 milhões de dólares.

Erros por Truncamento

- Erros de truncamento advém de métodos numéricos originados a partir de considerações de um número finito de termos de uma série ou números irracionais.

- Cometido na interrupção de processos infinitos.
  - Está associado ao método de aproximação empregado, como vimos quando fazemos aproximações usando polinômios de Taylor, como e^x, sen(x), cos(x), ln(x), log(x) ...
  - Discretizações – processo de refinamento deveria ser infinito mas só usamos pontos nodais.

Erros por Truncamento

- Determinação do valor de e.
- Lembrar que $e = \sum_{n=0}^{\infty} \frac{1}{n!}$. Logo:

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2{,}718281828459\overline{05}$$

- Um truncamento no sexto termo gera:

$$e = \sum_{n=0}^{5} \frac{1}{n!} = 2{,}7166666666\overline{6667}$$

Erros por Arredodamento

- Os erros de arredondamento são consequências de se trabalhar com uma aritmética de precisão finita (máquina).
- Eles ocorrem quando números com uma quantidade limitada de algarismos significativos são usados para representar números exatos.
  - 1/3 – armazenamento de racionais ilimitados
  - sqrt(2) , pi – armazenamento de irracionais
  - Mudança de base $(0,1)_{10}$ – $(???)_2$ mudança de base gerando racional ilimitado
  - Regiões de overflow e underflow – abrangência limitada da notação em ponto flutuante

Consequências dos erros de arredondamento

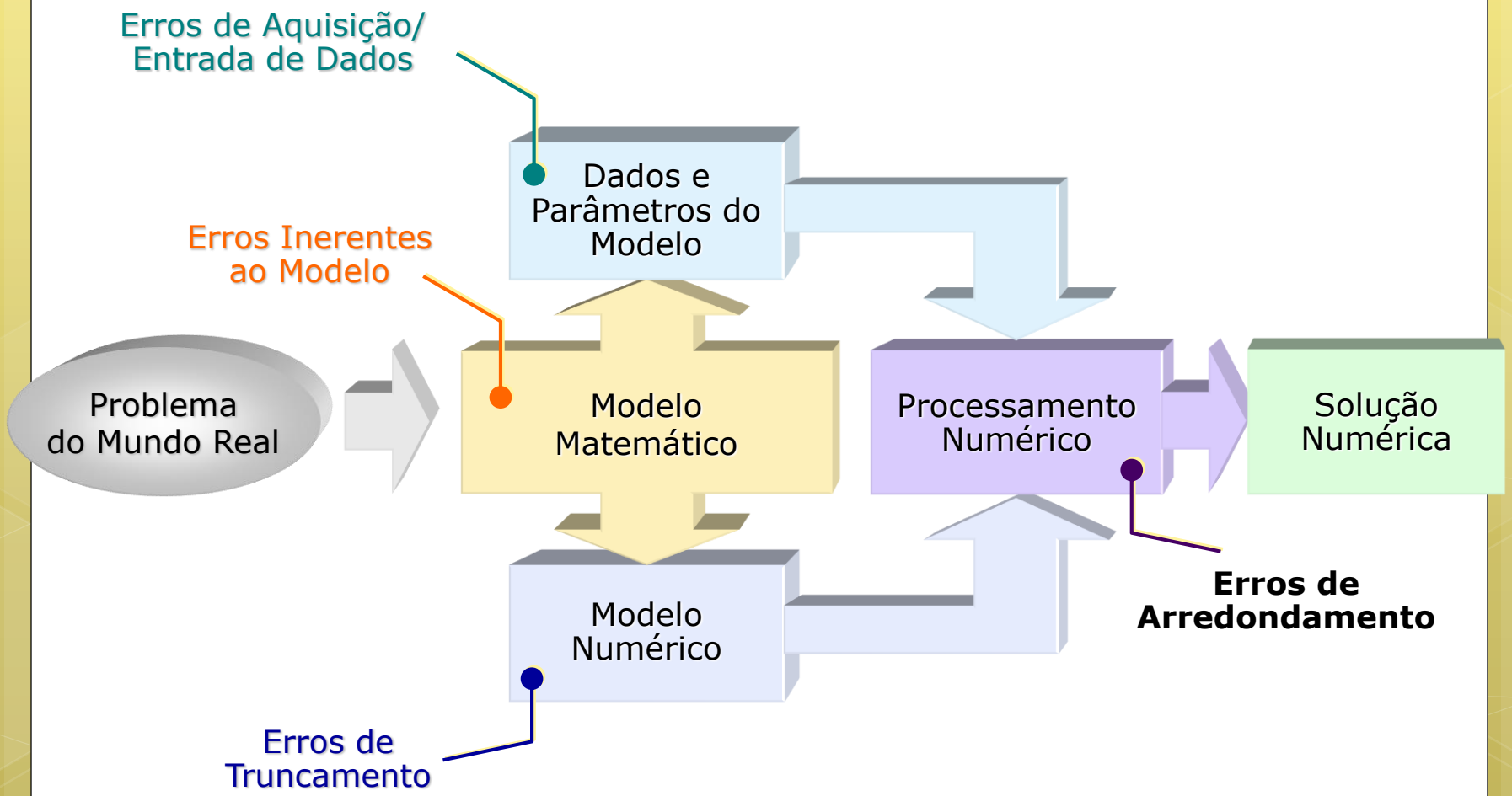
- Perda de significação – ocorre na subtração ou divisão. Na subtração de números próximos e na divisão onde o divisor (denominador) é pequeno em relação ao dividendo (numerador);
  - Obtenha $S = 1590/(9-x^2)$ para $x = 2,999$ ou $x = 2,9990005$
- Instabilidade numérica – deve-se ao uso recursivo (iterativo) de valores com erros, podendo deturpar o resultado final;

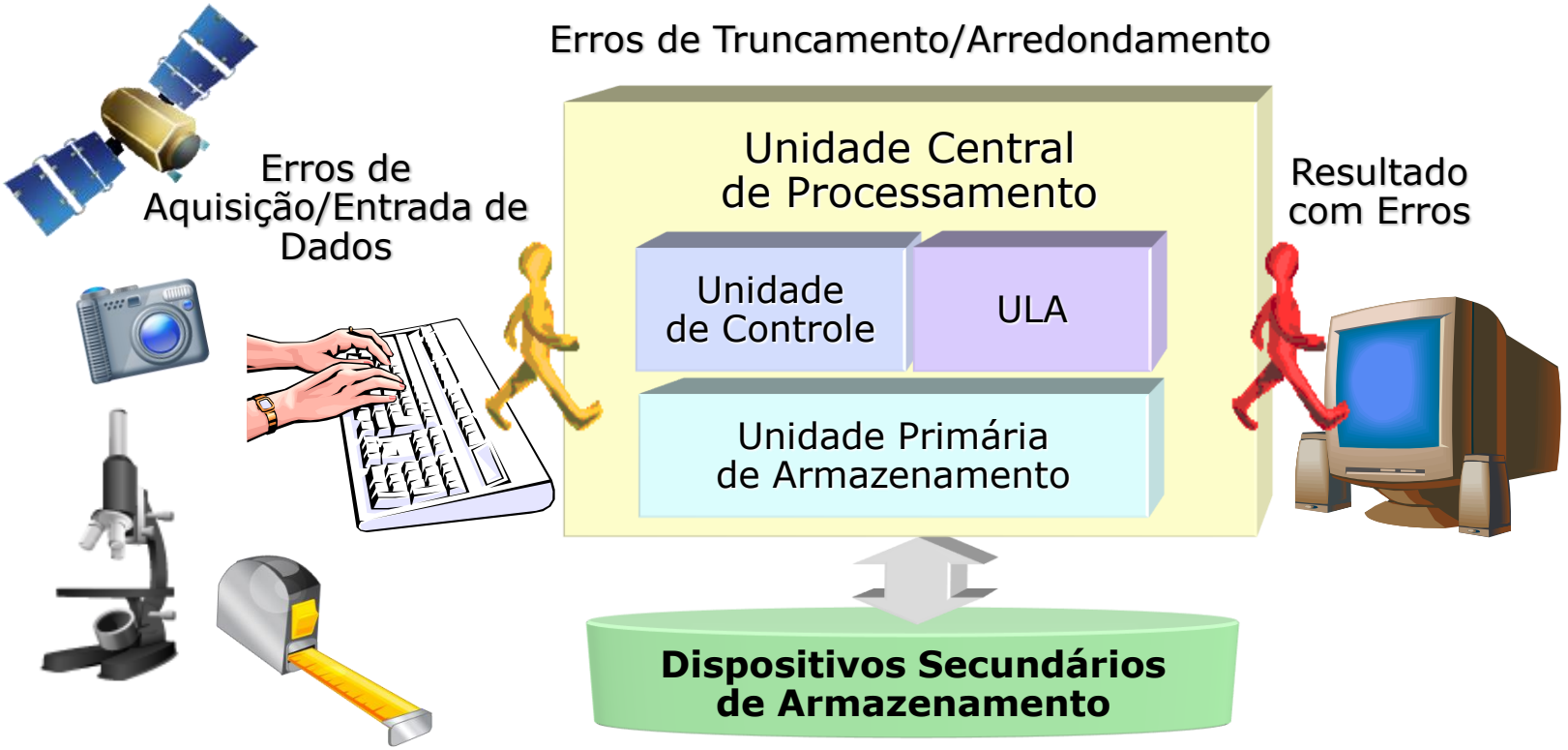
Consequências dos erros de arredondamento

- Falha no lançamento de mísseis
- 25/02/1991 – Guerra no golfo – míssil Patriot, erro de 0,34 s no cálculo do tempo de lançamento.
- Computador (24 bits)

Consequências dos erros de arredondamento

- Falha no lançamento de foguetes
- 04/06/1996 – Guiana Francesa – foguete Ariane 5
- Erro de trajetória 36,7 s após o lançamento
- Prejuízo: U$7,5 bilhões
- Limitação na representação numérica
  - erro numérico (overflow)


Erros de Aquisição/
Entrada de Dados
Dados e
Parâmetros do
Modelo
Erros Inerentes
ao Modelo
Problema
do Mundo Real
Modelo
Matemático
Processamento
Numérico
Solução
Numérica
Modelo
Numérico
Erros de
Arredondamento
Erros de
Truncamento


Erros de Truncamento/Arredondamento
Erros de Aquisição/Entrada de Dados
Unidade Central de Processamento
Unidade de Controle
ULA
Unidade Primária de Armazenamento
Resultado com Erros
Dispositivos Secundários de Armazenamento

Exatidão (Acurácia)

- Acurácia refere-se ao quão próximo um número simulado pelo computador está do valor exato. A exatidão de uma quantidade pode ser medida através do erro absoluto ou relativo.

Caso 1 ⇨ $x = 1{,}234 \qquad \hat{x} = 1{,}233 \qquad |x - \hat{x}| = 10^{-3}$

Caso 2 ⇨ $x = 0{,}002 \qquad \hat{x} = 0{,}001 \qquad |x - \hat{x}| = 10^{-3}$

O erro absoluto neste caso não é uma boa medida para quantificar a exatidão do número, use o erro relativo.

Precisão

- Precisão é governada pelo número de dígitos empregados na representação (precisão dupla no lugar de simples). A precisão está relacionada à máquina utilizada para realizar os cálculos:
- O número π = 3,141592654 representado por:

$$\pi 1 = 3{,}1416304958$$

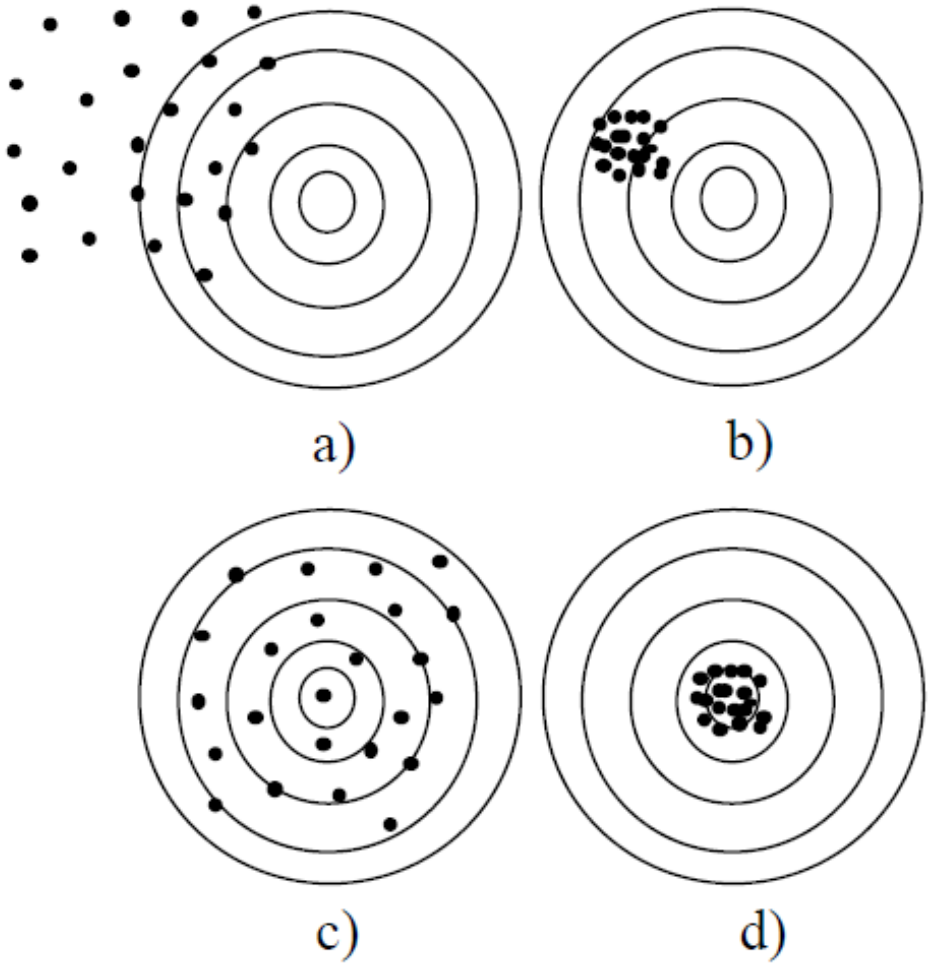
$$\pi 2 = 3{,}1415809485$$

- O número $\pi 2$ possui maior acurácia (erro absoluto menor) que $\pi 1$ quando comparados a π, embora ambos possuam a mesma precisão.

Acurácia x Precisão

- Os erros em quaisquer métodos são caracterizados pela acuracidade (acurácia) e precisão.
- A acurácia representa o quanto estamos próximos do valor real (o valor procurado, o alvo).
- A precisão está relacionada com o conceito de repetibilidade do resultado (número de dígitos de uma grandeza).
- Ex: Prática do tiro ao alvo.
  - A Fig.a apresenta baixa acurácia e precisão.
  - A Fig.b apresenta uma alta precisão (repetibilidade), porém uma baixa acurácia.
  - A Fig.c tem boa acurácia porém uma baixa precisão (não há repetibilidade na posição dos tiros).
  - A Fig.d apresenta uma excelente acurácia e precisão.

Acurácia x Precisão



2. Erro Absoluto e Relativo

- Define-se como erro absoluto a diferença entre o valor exato de um número e seu valor aproximado dado por

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

- Erro relativo é definido como o erro absoluto dividido pelo valor aproximado, amplamente utilizado.

$$e_R = \frac{x - \bar{x}}{\bar{x}} \qquad e_R = \frac{|x - \bar{x}|}{|\bar{x}|}$$

- Medidas do comprimento de uma ponte e de um prego:
  - Ponte: p = 10000 cm; p* = 9999 cm
    - Ea = 1 cm; Er = 0,000=0,01%
  - Prego: p = 10 cm; p* = 9 cm
    - Ea = 1 cm; Er = 0,1=10%

- Note que o erro relativo representou de uma forma mais adequada o erro na aferição das medidas.
- Na prática, costuma-se trabalhar com um limitante superior para o erro, ao invés do próprio erro ($|E| < \varepsilon$, sendo ε é o limitante)

Exemplo:

- Cálculo do erro relativo na representação dos números **a = 2112,9** e **e = 5,3**, sendo $|EA| < 0,1$
  - $|ERa| = |a - \bar{a}| / |a| = 0,1/2112,9 \cong 4,7 \times 10^{-5}$
  - $|ERe| = |e - \bar{e}| / |e| = 0,1/5,3 \cong 0,02$
- Conclusão: **a** é representado com maior precisão do que **e.**

Exercício:

- Considere $f(x)=x^2-a=0$, onde $a=2$ e $x_0=1$. Utilize o processo iterativo abaixo para uma tolerância $\varepsilon=0.0001$:

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right), \qquad n = 0,1,2 \ldots$$

3. Erros na mudança de base

Representação dos números

- No cotidiano usamos números na base 10, contudo microcomputadores e estações de trabalho utilizam a base 2, ou computadores de grande porte da IBM utilizam a base 16.

Exemplos:

$$(347)_{10} = 3\times10^{2} + 4\times10^{1} + 7\times10^{0}$$

$$(1101)_{2} = 1\times2^{3} + 1\times2^{2} + 0\times2^{1} + 1\times2^{0} \longrightarrow ???$$

Representação dos números

Parte Inteira — Parte Fracionária

$$N = a_m \beta^m + a_{m-1} \beta^{m-1} + \ \dots \ + a_1 \beta^1 + a_0 \ + \ a_{-1} \beta^{-1} + a_{-2} \beta^{-2} + \ \dots \ + a_{-n} \beta_{-n}$$

Onde: $0 \le a_k \le \beta - 1$

$$(0,2345)_{10} = (23,45 \times 10^{-2})_{10}$$

$$(0,1101)_2 = (11,01 \times 2^{-2})_2$$

Conversão de Bases

a) Base β ⟶ Base Decimal

Levar os coeficientes á expressão polinomial geral e calcular o valor. Observe que a solução numérica da expressão polinomial sempre resulta num número N na base decimal.

Exemplo:

$$(0{,}111)_2 = 1\times 2^{-1} + 1\times 2^{-2} + 1\times 2^{-3} = (0{,}875)_{10}$$

Conversão de Bases

b) Decimal $\longrightarrow$ Base β

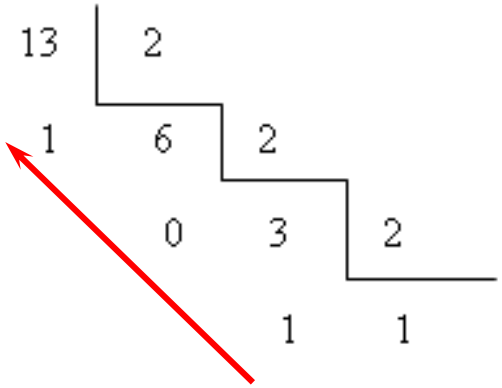
A conversão da base decimal para uma base β qualquer se dá em duas etapas, numa primeira etapa se converte a parte inteira do número e numa segunda etapa a parte fracionária do número.

b.1) Parte Inteira

Dividir o número sucessivamente pela base β até que o último quociente seja maior que zero e menor que β. O número é representado pelo último quociente e os restos na ordem inversa.

Exemplo:

$(13)_{10} \rightarrow (1101)_2$



Conversão de Bases

b.2) Parte Fracionária

Multiplicar sucessivamente a parte fracionária por β até que a mesma seja zero se a representação for exata. No caso da representação não ser exata, haverá uma sequência infinita na parte fracionária. O número é representado pelas partes inteiras resultantes.

Exemplo:

$(0{,}875)_{10} \rightarrow (0{,}111)_2$

0,875	0,750	0,500
$\times 2$	$\times 2$	$\times 2$
1,750	1,500	1,000

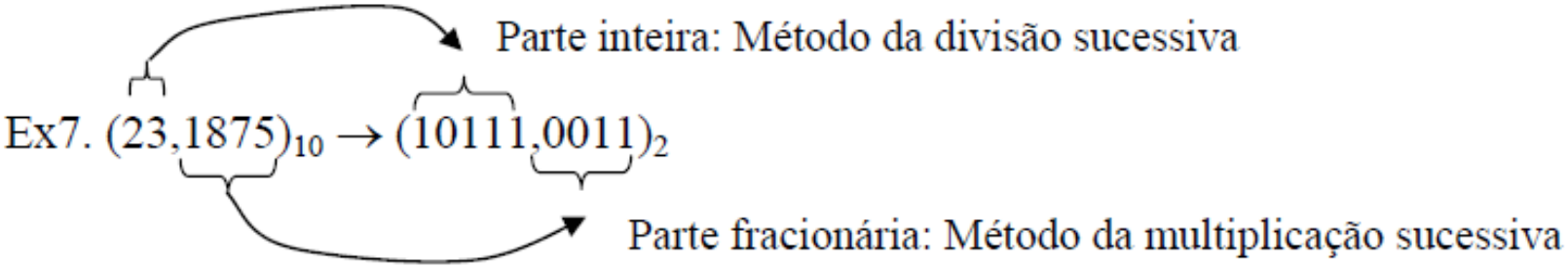
A representação de $(0{,}1)_{10}$ não possui representação exata na base 2.

$(0{,}10)_{10} \rightarrow (0{,}00011001100\ldots)_2$

Em virtude da representação não ser exata, a operação seguinte poderá não ter resultado exato na utilização de microcomputadores dependendo da linguagem de programação.

$\sum_{i=1}^{100} 0{,}1 = 9{,}9999999\ldots\ldots$ → Teste em Matlab e C

Conversão de Bases



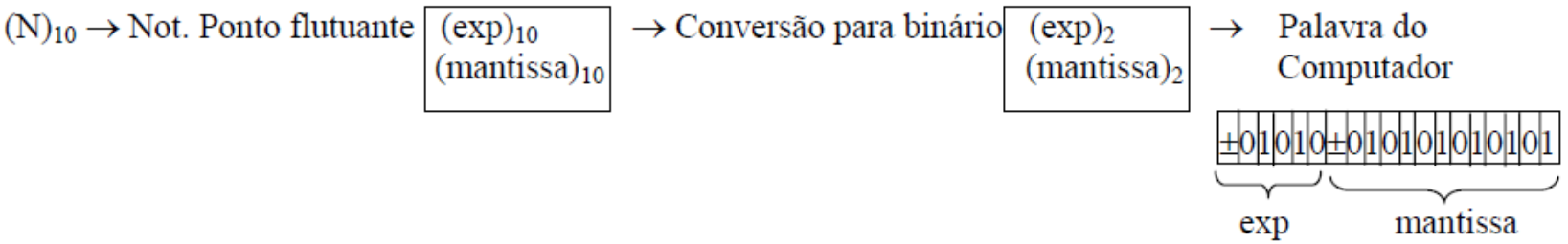
Exercício:

- Converta os seguintes números:
  - $(0{,}1875)_{10}$ para binário;
  - $(13{,}25)_{10}$ para binário;
  - $(0,\ 2)_{10}$ para binário;
  - $(111.011)_2$ para decimal;

4. Representação numérica de ponto flutuante

- Nas maquinas digitais, um digito binário é denominado BIT (do inglês, binary digit). Um grupo de oito bits corresponde a 1 byte. Dessa forma, percebemos que a representação dos números binários num computador é feita com um número finito de bits. A esse tamanho finito de bits é dado o nome palavra de computador.
- O tamanho da palavra do computador depende de características internas à arquitetura do mesmo. Em geral, os microcomputadores padrão PC tem tamanho de palavra de 16 e 32 bits. Computadores modernos tem palavras de 64 bits ou mais.
- Quanto maior o tamanho da palavra do computador mais veloz e mais preciso será o computador.

- Uma máquina digital (que opera em base 2) armazena um número internamente da seguinte forma esquematizada abaixo:



- Em princípio, representação de um número inteiro no computador não apresenta qualquer dificuldade. Qualquer computador trabalha internamente com uma base fixa β , onde β é um inteiro ≥ 2.

- Como seria a representação do número 1100 numa base β = 2?
  - $(1100)_2 = 1x2^3 + 1x2^2 + 0x2^1 + 0x2^0$

- Como seria a representação do número 1997 em uma base β = 10 ?
  - $(1997)_{10} = 1x10^3 + 9x10^2 + 9x10^1 + 7x10^0$

Um sistema de ponto flutuante $F \subset \Re$ é um subconjunto dos números reais cujos elementos tem a forma:

$$y = \pm \left(\frac{d_1}{\beta^1} + \frac{d_2}{\beta^2} + \frac{d_3}{\beta^2} + \ldots + \frac{d_t}{\beta^t} \right) \beta^e = \pm (.\, d_1 d_2 d_3 \ldots d_t)\, \beta^e$$

Onde $0 \le d_i < \beta, \quad i = 1, \ldots, t$

A aritmética de ponto flutuante F é caracterizada por quatro números inteiros:

- base β (binária, decimal, hexadecimal, etc.);
- precisão t (número de algarismos da mantissa);
- limites do expoente e ($e_{min} \le e \le e_{max}$);

$F(\beta, t, e_{min}, e_{max})$. A mantissa é fracionária nesta representação (<1).

O zero é representado de uma forma especial todos os dígitos da mantissa e expoente são nulos

- Considere uma máquina que opera no sistema F[10,3,-5,5], ou seja, um sistema onde β=10, t=3 e $e \in [-5,5]$. Nesse sistema os números são representados na seguinte forma de sistema:

$$\mathbf{(0.d_1d_2d_3)10^{e}, 0 \leq d_j \leq 9, d_1 \neq 0 \text{ e } e \in [-5,5]}$$

- O menor número representado nessa máquina é:
  - $m = 0.100 \cdot 10^{-5} = 10^{-6}$

- O maior:
  - $M = 0.999 \cdot 10^{5} = 99900$

Exemplo, se: $\beta = 2, t = 3, e_{\min} = -1 \quad e \quad e_{\max} = 3$

Mantissa fracionária

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

Expoentes

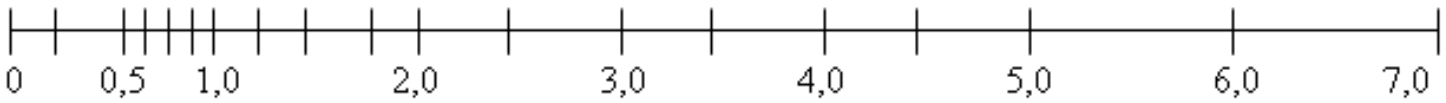
2^e $\quad e = -1, 0, 2 \; e \; 3$

Observe que os números em aritmética de ponto flutuante não são igualmente espaçados. O número total de elementos de uma máquina de aritmética em ponto flutuante é dado por:

$$ne = 2(\beta - 1)\beta^{t-1}(e_{\max} - e_{\min} + 1) + 1$$

Considerando apenas a parte positiva, tem-se os seguintes números:

0; 0,25; 0,3125; 0,4375; 0,5; 0,625; 0,750; 0,875; 1,0; 1,25; 1,5; 1,75; 2,0; 2,5; 3,0; 3,5; 4,0; 5,0; 6,0; 7,0, que podem ser representados na reta numerada:



Exemplo

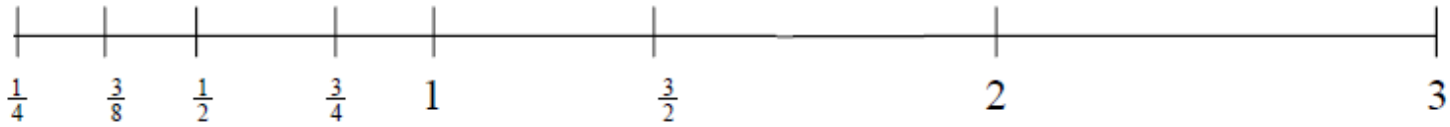
Considere $F(2,2,-1,2)$, com número normalizado, isto é, $d_1 \neq 0$. Os números serão:
$\pm .10 \times 2^e$ ou $\pm .11 \times 2^e$, sendo $-1 \leq e \leq 2$.

Convertendo para decimal, temos:
$.10 = ½$ e $.11 = ¾$

Com isso, os únicos números positivos representáveis nesse computador são:

Mantissa		Expoentes
$1/2 \times 2^e$		
$3/4 \times 2^e$	para	e= -1, 0, 1 e 2

Ou seja, ¼, ½, 1, 2, 3/8, 3/4, 3/2 e 3, que podem ser representados na reta numerada:



Alem desses números, os seus respectivos números negativos e o numero zero também serão representados.

- O número x = 235.89 representado em F[10,3,-5,5]:

- Veja que x = 235.89 pode ser representado por $x = 0.23589 .10^3$. Porém a mantissa em nosso sistema possui somente 3 dígitos. Assim, as únicas representações possíveis seriam: **$0.235.10^3$** usando truncamento ou **$0.236.10^3$** usando arredondamento.

- A consequência mais importante dessa representação é que, ao contrário do conjunto $\mathbb{R}$ dos números reais, o conjunto $\mathbb{F}$ efetivamente representáveis pela máquina é intrinsicamente finito, discreto e limitado. Isto é, nem todo número real pode ser representado na máquina.
- IEEE Standard for Floating-Point Arithmetic (IEEE 754) fornece o padrão técnico para as calculadoras - estabelecido pelo IEEE.
- Se uma operação aritmética resultar em um número que seja maior em módulo que o maior número representável naquela máquina ocorrerá um **overflow**, caso contrário **underflow**.

- IEEE Standard for Floating-Point Arithmetic (IEEE 754) fornece o padrão técnico para as calculadoras - estabelecido pelo IEEE.
- Se uma operação aritmética resultar em um número que seja maior em módulo que o maior número representável naquela máquina ocorrerá um **overflow**, caso contrário **underflow**.
- O zero é representado de uma forma especial, todos os dígitos da mantissa e do expoente são nulos.
- Os números em ponto flutuante são discretos e não contínuos como os números reais.
- O conceito de sempre existir um número real entre dois números reais quaisquer não é válido para os números em ponto flutuante. A falha deste conceito pode ter consequências desastrosas.

- Sejam x =875 e y=3172 . Calcular x × y. Primeiro, deve-se arredondar os números e armazena-los no formato indicado. A operação de multiplicação é efetuada usando 2t dígitos.
  - $x = 0{,}88.10^{3}$ e $y = 0{,}32.10^{4}$, $x.y = 0.2816 \times 10^{7}$.

- Sejam x =0,0064 e y=7312 Calcular x ÷ y. Primeiro, deve-se arredondar os números e armazená-los no formato indicado. A operação de divisão é efetuada usando 2t dígitos.
  - $x = 0{,}64.10^{-2}$ e $y = 0{,}73.10^{4}$, $x \div y = 0.8767.10^{-6}$

Exemplo:

- Escrever os números reais $x_1 = 0.35$, $x_2 = -5.172$, $x_3 = -0.0123$, $x_4 = 0.0003$ e $x_5 = 5391.3$, onde estão todos na base $\beta = 10$ em notação de um sistema de aritmética de ponto flutuante.

Solução:

- $0.35 = (3\times10^{-1} + 5\times10^{-2})\times10^{0} = 0.35\times100$
- $-5.172 = -(5\times10^{-1} + 1\times10^{-2} + 7\times10^{-3} + 2\times10^{-4})\times10^{1} = -0.5172\times10^{1}$
- $0.0123 = (1\times10^{-1} + 2\times10^{-2} + 3\times10^{-3})\times10^{-1} = 0.123\times10^{-1}$
- $5391.3 = (5\times10^{-1} + 3\times10^{-2} + 9\times10^{-3} + 1\times10^{-4} + 3\times10^{-5})\times10^{4} = 0.53913\times10^{4}$
- $0.0003 = (3\times10^{-1})\times10^{-3} = 0.3\times10^{-3}$

- Considerando agora que estamos diante de uma máquina que utilize apenas três dígitos significativos e que tenha como limite inferior e superior para o expoente, respectivamente, -2 e 2, como seriam representados nesta máquina os números do exemplo anterior?
- Solução:
  - Solução: Temos então para esta máquina $t = 3$, $I = -2$ e $S = -2$. Desta forma $-2 \leq e \leq 2$. Sendo assim temos:

- $0.35 = 0.350 \times 10^{0}$
- $-5.172 = -0.517 \times 10^{1}$
- $0.0123 = 0.123 \times 10^{-1}$
- $5391.3 = 0.53913 \times 10^{4}$ Não pode ser representado por esta máquina. **Erro de overflow**.
- $0.0003 == 0.3 \times 10^{-3}$ Não pode ser representado por esta máquina. **Erro de underflow**.

5. Propagação de erros

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.

- Será mostrado um exemplo que ilustra como os erros descritos anteriormente podem influenciar no desenvolvimento de um cálculo.
- Suponhamos que as operações indicadas nos itens a) e b) sejam processadas numa máquina com **4 dígitos significativos (t)**.

a) $(x_2 + x_1) - x_1$

b) $x_2 + (x_1 - x_1)$

- Fazendo $x_1 = 0.3491 \times 10^4$ e $x_2 = 0.2345 \times 10^0$ temos:

a) $(x_2 + x_1) - x_1 = (0.2345 \times 10^0 + 0.3491 \times 10^4) - 0.3491 \times 10^4$

$= 0.3491 \times 10^4 - 0.3491 \times 10^4$

$= 0.0000$

b) $x_2 + (x_1 - x_1) = 0.2345 \times 10^0 + (0.3491 \times 10^4 - 0.3491 \times 10^4)$

$= 0.2345 \times 10^0 + 0.0000$

$= 0.2345$

- A causa da diferença nas operações anteriores foi um arredondamento que foi feito na adição (x2 + x1) do item a), cujo resultado tem oito dígitos. Como a máquina só armazena 4 dígitos, os menos significativos foram desprezados.
- Ao se utilizar uma máquina de calcular deve-se está atento a essas particularidades causadas pelo erro de arredondamento, não só na adição, mas também nas demais operações.

Ex. 15: Dados $a = 50 \pm 3$ e $b = 21 \pm 1$, calcular $a + b$.

- Variação de a $\rightarrow$ 47 a 53
- Variação de b $\rightarrow$ 20 a 22
- Menor valor da soma $\rightarrow 47 + 20 = 67$
- Maior valor da soma $\rightarrow 53 + 22 = 75$
- $a + b = (50 + 21) \pm 4 = 71 \pm 4 \rightarrow$ 67 a 75