

# Chapter 1

## Sobre Erros

### 1.1 Introdução

Podemos distinguir vários tipos de erros que podem limitar a precisão dos resultados de cálculos,

1. erros inerentes;
2. erros de arredondamento;
3. erros de aproximação.

Os erros inerentes estão além do controle do cálculo e surgem, em geral, na aquisição de dados, que podem ser obtidos com instrumentos que possuem uma precisão limitada.

Os erros de arredondamento surgem quando realizamos os cálculos com números cuja representação é restrita a um número finito de dígitos, como é o caso dos computadores digitais.

Vejam os terceiro tipo de erro. Muitos métodos não fornecem a solução exata de um problema  $P$ , mesmo que os cálculos sejam realizados sem erros de arredondamento, eles fornecem, na verdade, a solução de um outro problema mais simples  $\tilde{P}$  que aproxima o problema  $P$ . Por exemplo, seja  $P$  o problema de somar os termos de uma série infinita como abaixo,

$$e = \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \dots$$

este problema pode ser substituído por um problema  $\hat{P}$  mais simples de somar somente um número finito de termos da série. Isto é

$$e \approx 1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!}$$

Muitas aproximações de problemas são obtidas por discretização do problema original: por exemplo, integrais definidas são aproximadas por somas finitas, diferenciais por diferenças, etc.. Em tais casos os erros de aproximação são também chamados erros de discretização. Estes tipos de erros serão discutidos individualmente, sempre que possível, para cada método estudado.

### 1.2 Representação de Números

Por razões técnicas, a maioria dos computadores digitais representam os números internamente na forma binária, ao invés da decimal.

Os computadores fazem uso de um número finito e fixo de posições quando internamente representam um número, o comprimento da palavra. Esse número  $n$  é determinado pela construção da máquina, embora algumas máquinas possuam extensões para múltiplos de  $n$ ,  $2n$  (precisão dupla),  $3n$  (precisão tripla), para fornecer uma precisão maior, se for necessário.

### 1.2.1 Conversão de Números: Decimal $\leftrightarrow$ Binário

Se o número é inteiro:

Ex:

$$\begin{aligned}(347)_{10} &= 3 \times 10^2 + 4 \times 10^1 + 7 \times 10^0 \\ (10111)_2 &= 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0\end{aligned}$$

#### Binário $\rightarrow$ Decimal

$$N = (a_n a_{n-1} \dots a_2 a_1 a_0)_\beta = a_n \beta^n + a_{n-1} \beta^{n-1} + \dots + a_2 \beta^2 + a_1 \beta^1 + a_0 \beta^0$$

logo

$$(10111)_2 = 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 16 + 0 + 4 + 2 + 1 = (23)_{10}$$

#### Decimal $\rightarrow$ Binário

$$\frac{347}{2} = 178 \times 2 + \underline{1}_9$$

$$\frac{178}{2} = 86 \times 2 + \underline{1}_8$$

$$\frac{86}{2} = 43 \times 2 + \underline{0}_7$$

$$\frac{43}{2} = 21 \times 2 + \underline{1}_6$$

$$\frac{21}{2} = 10 \times 2 + \underline{1}_5$$

$$\frac{10}{2} = 5 \times 2 + \underline{0}_4$$

$$\frac{5}{2} = 2 \times 2 + \underline{1}_3$$

$$\frac{2}{2} = \underline{1}_1 \times 2 + \underline{0}_2$$

ou

$$(347)_{10} = (101011011)_2$$

Se o número é fracionário

$$r = r_I + r_F = \text{Parte Inteiro} + \text{Parte Fracionária}$$

Se a base é decimal (10)

$$r_F = b_1 \times 10^{-1} + b_2 \times 10^{-2} + b_3 \times 10^{-3} + \dots + b_k \times 10^{-k} + \dots$$

ex:

$$r_F = 0.125 = 1 \times 10^{-1} + 2 \times 10^{-2} + 5 \times 10^{-3} \quad (\text{representação finita})$$

$$r_F = 0.11111\dots = 1 \times 10^{-1} + 1 \times 10^{-2} + 1 \times 10^{-3} + \dots \quad (\text{sem representação finita})$$

Se a base é binária (2)

$$r_F = b_1 \times 2^{-1} + b_2 \times 2^{-2} + b_3 \times 2^{-3} + \dots + b_k \times 2^{-k} + \dots$$

### Binário $\rightarrow$ Decimal

Ex:

$$(0.001)_2 = 0 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} = 0 + 0 + 0.125 = (0.125)_{10}$$

### Decimal $\rightarrow$ Binário

$r_F \rightarrow$  parte decimal

$$\begin{aligned} \rightarrow r_F &= \sum_{k=1}^{\infty} b_k 2^{-k} \\ \times 2 \rightarrow 2r_F &= \sum_{k=1}^{\infty} b_k 2^{-k+1} = \underbrace{b_1}_{(*)} + \underbrace{\sum_{k=1}^{\infty} b_{k+1} 2^{-k}}_{(**)} \end{aligned}$$

onde  $(*)$  é a parte inteira binária de  $2r_F$  e  $(**)$  é a parte fracionária binária.

Novamente

$$\begin{aligned} \rightarrow (2r_F)_F &= \sum_{k=1}^{\infty} b_{k+1} 2^{-k} \\ \rightarrow 2(2r_F)_F &= \sum_{k=1}^{\infty} b_{k+1} 2^{-k+1} = \underbrace{b_2}_{(*)} + \underbrace{\sum_{k=1}^{\infty} b_{k+2} 2^{-k}}_{(**)} \end{aligned}$$

onde  $(*)$  é a parte inteira binária de  $2(2r_F)_F$  e  $(**)$  é a parte fracionária binária.

Ex:

$$1. \quad r_F = (0.5)_{10}$$

$$2r_F = 2 \left\{ \sum_{k=1}^{\infty} b_k 2^{-k} \right\} = \sum_{k=1}^{\infty} b_k 2^{-k+1} = b_1 + \sum_{k=1}^{\infty} b_{k+1} 2^{-k}$$

$$2r_F = 2(0.5) = 1 = b_1 + \sum_{k=1}^{\infty} b_{k+1}2^{-k}$$

$$\implies b_1 = 1 \text{ e } b_k = 0, \quad k = 2, 3, \dots$$

assim

$$r_F = (0.5)_{10} = (0.1)_2$$

2.  $r_F = (0.1)_{10}$

$$\begin{array}{rcl} 2r_F & = & 0.2 \longrightarrow b_1 = 0 \\ 2(2r_F)_F & = & 0.4 \longrightarrow b_2 = 0 \\ 2(2(2r_F)_F)_F & = & 0.8 \longrightarrow b_3 = 0 \\ 2(2(2(2r_F)_F)_F)_F & = & 1.6 \longrightarrow b_4 = 1 \\ 2(0.6) & = & 1.2 \longrightarrow b_5 = 1 \\ 2(0.2) & = & 0.4 \longrightarrow b_6 = 0 \\ 2(0.4) & = & 0.8 \longrightarrow b_7 = 0 \\ 2(0.8) & = & 1.6 \longrightarrow b_8 = 1 \\ 2(0.6) & = & 1.2 \longrightarrow b_9 = 1 \\ 2(0.2) & = & 0.4 \longrightarrow b_{10} = 0 \\ 2(0.4) & = & 0.8 \longrightarrow b_{11} = 0 \\ 2(0.8) & = & 1.6 \longrightarrow b_{12} = 1 \\ 2(0.6) & = & 1.2 \longrightarrow b_{13} = 1 \\ 2(0.2) & = & 0.4 \longrightarrow b_{14} = 0 \\ \vdots & = & \vdots \longrightarrow \vdots \end{array}$$

logo  $(0.1)_{10} = (0.000110011001100110011\dots)_2$ .

Note que  $(0.1)_{10}$  tem representação finita, porém, na base binária não  $(0.0001100\dots)_2$ .

Ex:

$$\sum_{k=1}^{100} 0.1 \neq 10 \text{ num computador}$$

na máquina será armazenada uma aproximação para o número que depende do comprimento da palavra do computador.

Vimos que uma fonte de erros está no processo de conversão do sistema decimal para o sistema binário e vice-versa. Outra fonte de erros está no fato de que os computadores possuem um número finito de dígitos para representar os números reais.

## 1.2.2 Representação em Ponto Flutuante

Neste tipo de representação o ponto decimal (binário) não é fixo, sua posição com respeito ao primeiro dígito é indicado separadamente, através de um expoente. Isto é,

### Definição:

Um número real  $x$  é dito um número de ponto flutuante normalizado se:

- (1)  $x = mb^e$
- (2)  $m = \pm 0.d_1d_2d_3 \dots d_n$ ,  $n \in \mathbb{N}$
- (3)  $1 \leq d_1 \leq b-1$ ,  $0 \leq d_i \leq b-1$ ;  $i = 2, 3, 4, \dots, n$
- (4)  $e_1 \leq e \leq e_2$ ,  $e_1 \leq 0$ ,  $e_2 \geq 1$ ,  $e_1, e_2 \in \mathbb{Z}$

onde

- $b$  - base,  $b \geq 2$
- $e$  - expoente
- $m$  - mantissa
- $n$  - número de dígitos usados para representar a mantissa

Todos os números de ponto flutuante e o número 0 que é representado como

$$0 = 0.\underbrace{000 \dots 00}_n \times b^{e_1}$$

formam um sistema de ponto flutuante denotado por  $F(b, n, e_1, e_2)$ .

Exemplo:

$$\begin{aligned} 0.1236 &= 0.1236 \times 10^0 \\ &= 1.236 \times 10^{-1} \\ &= 12.36 \times 10^{-2} \\ &= 0.01236 \times 10^1 \\ &= 0.001236 \times 10^2 \end{aligned}$$

Dado um sistema  $F$  de ponto flutuante

$$F = F(b, n, e_1, e_2)$$

Temos que:

$0.1 \times b^{e_1} \longrightarrow$  menor número positivo não nulo

$0.[b-1][b-1][b-1] \dots [b-1] \times b^{e_2} \longrightarrow$  maior número não nulo

Além disso

$$\#F = \underbrace{2}_{\pm} \times \underbrace{(b-1)b^{n-1}}_{\text{mantissas } \neq} \times \underbrace{(e_2 - e_1 + 1)}_{\text{expoentes } \neq} + \underbrace{1}_0$$

Exemplo:

$$F = F(b, n, e_1, e_2) = F(2, 3, -1, 2)$$

$$\#F = \underbrace{2}_{b} \times \underbrace{1 \times 2^2}_{2^n} \times \underbrace{(2 - (-1) + 1)}_{e_1 + 1} + \underbrace{1}_{e_2} = 33$$

$$b = 2$$

$n = 3 \rightarrow$  três dígitos para a mantissa, logo todas as mantissas possíveis são:

0.100

0.101

0.110

0.111

$e_1 = -1$ ;  $e_2 \rightarrow$  todos os expoentes possíveis são:

-1

0

1

2

Temos então os seguintes números reais positivos neste sistema

$$(0.100 \times 2^{-1})_2 = (0.01)_2 = (0 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2})_{10} = \frac{1}{4}$$

$$(0.100 \times 2^0)_2 = (0.1)_2 = (0 \times 2^0 + 1 \times 2^{-1})_{10} = \frac{1}{2}$$

$$(0.100 \times 2^1)_2 = (1.0)_2 = (1 \times 2^0)_{10} = 1$$

$$(0.100 \times 2^2)_2 = (10.0)_2 = (1 \times 2^1 + 0 \times 2^0)_{10} = 2$$

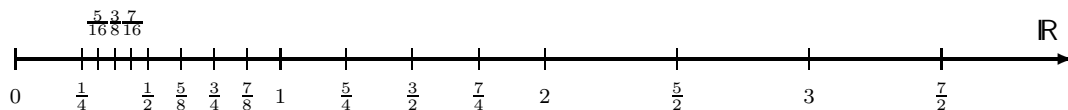
$$(0.101 \times 2^{-1})_2 = (0.0101)_2 = (0 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4})_{10} = \frac{5}{16}$$

$$(0.101 \times 2^0)_2 = (0.101)_2 = (0 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3})_{10} = \frac{5}{8}$$

e assim sucessivamente. Resumindo

$e \backslash m$	0.100	0.101	0.110	0.111
-1	1/4	5/16	3/8	7/16
0	1/2	5/8	3/4	7/8
1	1	5/4	3/2	7/4
2	2	5/2	3	7/2

Representando estes números na reta dos reais



Vemos claramente que os números de ponto flutuante de  $F$  não estão uniformemente distribuídos no intervalo  $[-7/2, 7/2]$ , na verdade, existem regiões onde eles estão uniformemente distribuídos.

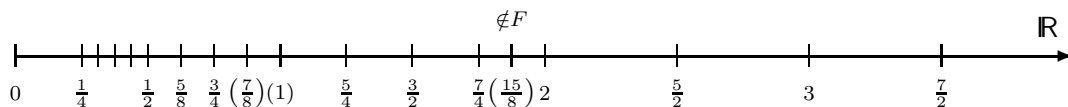
Podemos destacar duas regiões notórias num sistema de ponto flutuante.

Região de *UNDERFLOW*: entre o maior número de ponto flutuante negativo e zero e entre zero e o menor número de ponto flutuante positivo. Para diminuir esta região devemos aumentar o valor de  $n$ .

Região de *OVERFLOW*: a região aquém e além do menor e maior número de ponto flutuante, respectivamente. Para melhorar devemos aumentar e diminuir  $e_2$  e  $e_1$ , respectivamente.

Além disso podemos observar também que se efetuarmos uma operação aritmética com pontos de um sistema de ponto flutuante, seu resultado pode não pertencer ao sistema. Por exemplo, no sistema mostrado acima temos

$$1 + \frac{7}{8} = \frac{15}{8} \notin F \quad \left\{ \begin{array}{l} 1 + \frac{7}{8} = \frac{7}{4} \in F \\ 1 + \frac{7}{8} = 2 \in F \end{array} \right.$$



logo, podemos ter

$$x \oplus y \neq x + y$$

$$x \ominus y \neq x - y$$

$$x \otimes y \neq x \times y$$

$$x \oslash y \neq x / y$$

onde  $\odot$  denota uma operação aritmética elementar efetuada em um sistema de ponto flutuante.

Devido a estes erros as operações aritméticas podem fornecer como resultados números incorretos, mesmo que os operandos estejam certos. Ademais, outras propriedades das operações aritméticas deixam de serem válidas, e o resultado pode depender do algoritmo usado.

Exemplo:

Seja  $F = F(10, 8, -99, 99)$ , e os seguintes números reais com representação no sistema de ponto flutuante

$$a := 0,23371258 \times 10^{-4}$$

$$b := 0,33678429 \times 10^2$$

$$c := -0,33677811 \times 10^2$$

temos, então,

$$\begin{aligned} a + (b + c) &= 0,23371258 \times 10^{-4} + (0,33678429 \times 10^2 - 0,33677811 \times 10^2) \\ &= 0,23371258 \times 10^{-4} + 0,00000618 \times 10^2 \\ &= 0,23371258 \times 10^{-4} + 0,61800000 \times 10^{-3} \\ &= 0,02337126 \times 10^{-3} + 0,61800000 \times 10^{-3} \\ &= 0,64137126 \times 10^{-3} \end{aligned}$$

$$\begin{aligned} (a + b) + c &= (0,23371258 \times 10^{-4} + 0,33678429 \times 10^2) - 0,33677811 \times 10^2 \\ &= (0,00000023 \times 10^2 + 0,33678429 \times 10^2) - 0,33677811 \times 10^2 \\ &= 0,33678452 \times 10^2 - 0,33677811 \times 10^2 \\ &= 0,64100000 \times 10^{-3} \end{aligned}$$

que são diferentes do resultado exato

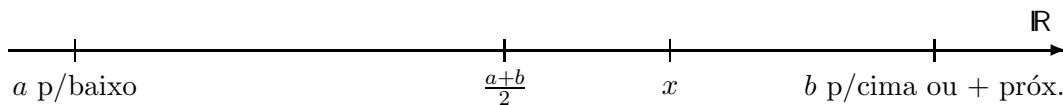
$$a + b + c = 0,641371258 \times 10^{-3}$$

Portanto as operações aritméticas em ponto flutuante não são necessariamente associativas ou distributivas.

Arredondamento

A aproximação de um número real por um número de ponto flutuante pode ser feito de maneiras diferentes. As mais conhecidas são:

- Arredondamento para cima;
- Arredondamento para baixo;
- Arredondamento para o número de máquina mais próximo;



Exemplo:  $F = F(2, 3, -1, 2)$

$$\frac{9}{8} \notin F$$

$$\frac{9}{8} = (1.125)_{10} = (0.1001 \times 2^1)_2$$

$$\text{para baixo } \frac{9}{8} = (0.100 \times 2^1)_2 = 1$$

$$\text{para cima } \frac{9}{8} = (0.101 \times 2^1)_2 = \frac{5}{4}$$



### 1.2.3 Erros Absolutos e Relativos

Erro Absoluto:

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

ou

$$EA_x = |x - \tilde{x}| \quad \longrightarrow \text{mais usada}$$

onde  $x$  é o valor exato e  $\tilde{x}$  é o valor aproximado.

Erro Relativo:

$$ER_x = \frac{x - \tilde{x}}{x} \quad \text{ou} \quad ER_x = \frac{x - \tilde{x}}{\tilde{x}}$$

ou

$$ER_x = \frac{|x - \tilde{x}|}{|x|} \quad \text{ou} \quad ER_x = \frac{|x - \tilde{x}|}{|\tilde{x}|} \quad \longrightarrow \text{mais usadas}$$

Os erros relativos são mais utilizados que os absolutos pois são mais representativos.

Exemplo:

$$x = 0.00005$$

$$\tilde{x} = 0.00006$$

logo,

$$EA_x = |x - \tilde{x}| = |0.00005 - 0.00006| = 0.00001 \quad \longrightarrow \text{parece pequeno}$$

$$ER_x = \frac{|x - \tilde{x}|}{|x|} = \frac{|0.00005 - 0.00006|}{|0.00005|} = \frac{0.00001}{0.00005} = 0.2 = 20\%$$

## Chapter 2

# Resolução de Sistemas Lineares

### 2.1 Introdução

Seja uma sistema linear geral de  $n$  equações e  $n$  incógnitas  $x_1, x_2, x_3, \dots, x_n$

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + \dots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + \dots + a_{3n}x_n = b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + \dots + a_{4n}x_n = b_4 \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + a_{n4}x_4 + \dots + a_{nn}x_n = b_n \end{array} \right. \quad (2.1)$$

Que reescrito na forma matricial torna-se

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \dots & a_{3n} \\ a_{41} & a_{42} & a_{43} & a_{44} & \dots & a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \dots & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ \vdots \\ b_n \end{bmatrix} \quad (2.2)$$

ou na forma compacta

$$\mathbf{Ax} = \mathbf{b}$$

onde

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots & a_{3n} \\ a_{41} & a_{42} & a_{43} & a_{44} & \cdots & a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nn} \end{bmatrix}$$

é chamada matriz de coeficientes do sistema

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \end{bmatrix}$$

é o vetor de incógnitas e

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ \vdots \\ b_n \end{bmatrix}$$

é o vetor segundo membro.

Teoricamente se  $\mathbf{A}$  é inversível ( $\det(\mathbf{A}) \neq 0$ ) então  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$  é a solução do sistema, entretanto, calcular inversas de matrizes pode ser uma tarefa trabalhosa.

## 2.2 Definições Preliminares

$\mathbf{A} \in \mathbb{R}^{n \times n}$ , são chamadas matrizes quadradas

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots & a_{3n} \\ a_{41} & a_{42} & a_{43} & a_{44} & \cdots & a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nn} \end{bmatrix}$$

### Operações Básicas com Matrizes

$$\begin{aligned} \mathbf{C} &= \mathbf{A} + \mathbf{B} & ; & \quad c_{ij} = a_{ij} + b_{ij} \\ \mathbf{C} &= \alpha \mathbf{A} & ; & \quad c_{ij} = \alpha a_{ij} \\ \mathbf{C} &= \mathbf{AB} & ; & \quad c_{ij} = \sum_{k=1}^n a_{ik} \cdot b_{kj} \\ \mathbf{C} &= \mathbf{A}^T & ; & \quad c_{ij} = a_{ji} \end{aligned}$$

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Chama-se inversa de uma matriz  $\mathbf{A}$  a matriz  $\mathbf{A}^{-1}$  tal que  $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ .

Se  $\mathbf{A}^{-1}$  existe,  $\mathbf{A}$  é dita não-singular, caso contrário  $\mathbf{A}$  é dita singular.

### Determinantes:

$$\text{Se } \mathbf{A} = a \in \mathbb{R}^{1 \times 1} \longrightarrow \det(\mathbf{A}) = a$$

$$\text{Se } \mathbf{A} \in \mathbb{R}^{n \times n} \longrightarrow$$

$$\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{j+1} a_{ij} \det(\mathbf{A}_{1j})$$

onde a matriz  $\mathbf{A}_{1j}$  é obtida retirando-se a primeira linha e a  $j$ -ésima coluna de  $\mathbf{A}$ .

Propriedades dos determinantes:

1.  $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ ;
2.  $\det(\mathbf{A}^T) = \det(\mathbf{A})$ ;
3.  $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$ ;
4.  $\det(\mathbf{A}) \neq 0 \iff \mathbf{A}$  é não-singular;

## Matrizes Especiais:

- Simétrica se  $\mathbf{A}^T = \mathbf{A}$
- Anti-simétrica se  $\mathbf{A}^T = -\mathbf{A}$
- Positiva definida se  $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0, \vec{0} \neq \mathbf{x} \in \mathbb{R}^n$
- Não positiva definida se  $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0, \mathbf{x} \in \mathbb{R}^n$
- Ortogonal se  $\mathbf{A}^T \mathbf{A} = \mathbf{I}$
- Nilpotente se  $\mathbf{A}^k = \mathbf{0}$  para algum  $k$
- Idempotente se  $\mathbf{A}^2 = \mathbf{A}$
- Positiva se  $a_{ij} > 0, \forall i, j$
- Não negativa se  $a_{ij} \geq 0, \forall i, j$
- Diagonal dominante se  $|a_{ii}| > \sum_{j \neq i}^n |a_{ij}|, \forall i$

Classificação quanto a forma da matriz:

- Diagonal se  $a_{ij} = 0$  para  $i \neq j$ ;
- Tridiagonal se  $a_{ij} = 0$  para  $|i - j| > 1$ ;
- Triangular superior se  $a_{ij} = 0$  para  $i > j$ ;
- Estritamente triangular superior se  $a_{ij} = 0$  para  $i \geq j$ ;
- Triangular inferior se  $a_{ij} = 0$  para  $i < j$ ;
  
- Matriz esparsa se tem a maioria de seus elementos nulos;
  
- Matriz densa se a maioria de seus elementos são diferentes de zero;
  
- Matriz em Banda Uma matriz é de banda superior  $s$  e de banda inferior  $r$  se  $a_{ij} = 0$  para  $i > j + r$  e  $j > i + s$ , se  $s = r$  a matriz é chamada simplesmente de banda  $r$ .

**Definição:** Uma norma em  $\mathbb{R}^n$  é uma função real  $\|\cdot\|$  satisfazendo as seguintes propriedades.

- $\forall \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| \geq 0$  e  $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \vec{0}$
- $\forall \mathbf{x} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}, \|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
- $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

**Exemplos:**

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2} \quad \text{Norma Euclidiana}$$

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i| \quad \text{Norma da Soma}$$

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i| \quad \text{Norma do Mximo}$$

Analogamente podemos definir normas para matrizes como:

**Definio:** Uma norma em  $\mathbb{R}^{m \times n}$   uma funo real  $\|\cdot\|$  satisfazendo as seguintes propriedades.

- $\forall \mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\|\mathbf{A}\| \geq 0$  e  $\|\mathbf{A}\| = 0 \iff \mathbf{A} = \hat{\mathbf{0}}$
- $\forall \mathbf{A} \in \mathbb{R}^{m \times n}, \forall \alpha \in \mathbb{R}$ ,  $\|\alpha \mathbf{A}\| = |\alpha| \|\mathbf{A}\|$
- $\forall \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$ ,  $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$

**Exemplos:**

$$\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \left( \sum_{i=1}^n |a_{ij}| \right) \quad \text{Norma do Mximo das Colunas}$$

$$\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq m} \left( \sum_{j=1}^n |a_{ij}| \right) \quad \text{Norma do Mximo das Linhas}$$

$$\|\mathbf{A}\|_E = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2} \quad \text{Norma Euclidiana}$$

**Observao:** Uma norma matricial  consistente com uma norma vetorial se:

$$\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \|\mathbf{x}\| \quad \forall \mathbf{A} \in \mathbb{R}^{n \times n}, \forall \mathbf{x} \in \mathbb{R}^n$$

Existem duas grandes classes de mtodos numricos para a resoluo de sistemas lineares. Os Mtodos Diretos e os Mtodos Iterativos. Nos mtodos diretos a soluo do sistema  obtida aps um nmero finito de passos. Diferentemente, os mtodos iterativos partem de uma aproximao inicial para a soluo do sistema e geram uma seqncia de aproximaoes que esperamos convirja para a soluo do sistema.

## 2.3 Mtodos Diretos

Nesta classe de mtodos a soluo do sistema  obtida aps um nmero finito de passos e so em geral baseados em mtodos de eliminao onde se transforma o sistema a resolver em outro sistema de resoluo mais fcil que o sistema original. Estes mtodos so normalmente empregados na soluo de sistemas de pequeno a mdio porte em que a matriz de coeficientes  densa.

### 2.3.1 Eliminação de Gauss / Fatoração LU

O objetivo do processo de Eliminação de Gauss é transformar um sistema linear  $\mathbf{Ax} = \mathbf{b}$  em um outro sistema, que possui a mesma solução do primeiro,  $\mathbf{Ux} = \mathbf{c}$  tal que a matriz  $\mathbf{U}$  seja triangular superior. Esta transformação é feita através de combinações lineares das linhas do sistema. As operações sobre as linhas não alteram a solução do sistema original e se resumem a:

- Multiplicação de uma linha por um escalar;
- Soma de duas linhas;
- Troca de duas linhas.

Os sistemas lineares onde a matriz do sistema é triangular superior são de fácil solução. Seja o sistema triangular superior abaixo

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1,n-1}x_{n-1} + a_{1n}x_n = b_1 \\ a_{22}x_2 + a_{23}x_3 + \cdots + a_{2,n-1}x_{n-1} + a_{2n}x_n = b_2 \\ a_{33}x_3 + \cdots + a_{3,n-1}x_{n-1} + a_{3n}x_n = b_3 \\ \vdots \\ a_{n-1,n-1}x_{n-1} + a_{n-1,n}x_n = b_n \\ a_{nn}x_n = b_n \end{array} \right. \quad (2.3)$$

Vemos que a última equação do sistema acima só possui dependência da última incógnita  $x_n$ , logo podemos calcular esta incógnita,

$$x_n = \frac{b_n}{a_{nn}}$$

da penúltima equação como já conhecemos  $x_n$ , podemos determinar  $x_{n-1}$ , logo

$$x_{n-1} = \frac{1}{a_{n-1,n-1}} \{b_n - a_{n-1,n}x_n\}$$

Conhecendo  $x_{n-1}$  e  $x_n$  podemos da antipenúltima equação do sistema determinar  $x_{n-2}$ . Com  $x_n$ ,  $x_{n-1}$  e  $x_{n-2}$  calculamos  $x_{n-3}$  e repetindo este procedimento podemos determinar todas as incógnitas do sistema. Este processo é chamado *retro-substituição*.

Para  $i = n, n-1, \dots, 1$

$$x_i = b_i$$

Para  $j = i+1, \dots, n$

$$x_i = x_i - a_{ij}x_j$$

$$x_i = x_i / a_{ii}$$

**Exemplo:**

$$\left\{ \begin{array}{l} 3x_1 + 4x_2 - 4x_3 = 3 \\ 2x_2 + x_3 = 3 \\ 4x_3 = 4 \end{array} \right.$$





$$\text{Da terceira equação} \longrightarrow x_3 = \frac{1}{-4} \{3 - 3x_1 - 4x_2\} = \frac{1}{-4} \{3 - 3 - 4\} = \frac{-4}{-4} = 1$$

*Resp:* (1; 1; 1)

Já vimos que muito fácil resolver qualquer sistema triangular superior. Veremos agora como transformar um sistema qualquer em um sistema triangular superior usando combinações lineares das linhas do sistema. Vejamos o procedimento no seguinte exemplo:

**Exemplo:** Seja o sistema linear

$$\begin{cases} 3x_1 + x_2 + 6x_3 = 2 \\ 2x_1 + x_2 + 3x_3 = 7 \\ x_1 + x_2 + x_3 = 4 \end{cases}$$

Substituiremos a (2ª linha) por  $\left\{ (2^\text{ª linha}) - \frac{2}{3}(1^\text{ª linha}) \right\}$ , isto é

$$\begin{array}{rcl} (2^\text{ª linha}) & 2x_1 + x_2 + 3x_3 = & 7 \\ -\frac{2}{3}(1^\text{ª linha}) & -\frac{2}{3}3x_1 - \frac{2}{3}x_2 - \frac{2}{3}6x_3 = & -\frac{2}{3}2 \\ \hline & 0x_1 + \frac{1}{3}x_2 - x_3 = & \frac{17}{3} \end{array}$$

o sistema linear se transforma

$$\begin{cases} 3x_1 + x_2 + 6x_3 = 2 \\ 0x_1 + \frac{1}{3}x_2 - x_3 = \frac{17}{3} \\ x_1 + x_2 + x_3 = 4 \end{cases}$$

Agora substituiremos a (3ª linha) por  $\left\{ (3^\text{ª linha}) - \frac{1}{3}(1^\text{ª linha}) \right\}$ , isto é

$$\begin{array}{rcl} (3^\text{ª linha}) & x_1 + x_2 + x_3 = & 4 \\ -\frac{1}{3}(1^\text{ª linha}) & -\frac{1}{3}3x_1 - \frac{1}{3}x_2 - \frac{1}{3}6x_3 = & -\frac{1}{3}2 \\ \hline & 0x_1 + \frac{2}{3}x_2 - x_3 = & \frac{10}{3} \end{array}$$

e o sistema linear se torna

$$\begin{cases} 3x_1 + x_2 + 6x_3 = 2 \\ 0x_1 + \frac{1}{3}x_2 - x_3 = \frac{17}{3} \\ 0x_1 + \frac{2}{3}x_2 - x_3 = \frac{10}{3} \end{cases}$$

Enfim, substituiremos a ( $3^{\text{a}} \text{ linha}$ ) por  $\{(3^{\text{a}} \text{ linha}) - 2(2^{\text{a}} \text{ linha})\}$ , isto é

$$\begin{array}{rcl} (3^{\text{a}} \text{ linha}) & 0x_1 + \frac{2}{3}x_2 - x_3 = & \frac{10}{3} \\ \\ -2(2^{\text{a}} \text{ linha}) & -2 \cdot 0x_1 - 2\frac{1}{3}x_2 - 2x_3 = & -2\frac{17}{3} \\ \hline & 0x_1 + 0x_2 + x_3 = & -\frac{24}{3} \end{array}$$

resultando no seguinte sistema linear

$$\begin{cases} 3x_1 + x_2 + 6x_3 = 2 \\ 0x_1 + \frac{1}{3}x_2 - x_3 = \frac{17}{3} \\ 0x_1 + 0x_2 + x_3 = -\frac{24}{3} \end{cases}$$

ou

$$\begin{cases} 3x_1 + x_2 + 6x_3 = 2 \\ \frac{1}{3}x_2 - x_3 = \frac{17}{3} \\ x_3 = -\frac{24}{3} \end{cases}$$

que é triangular superior, logo de solução imediata.

$$x_3 = -8, \quad x_2 = -7, \quad x_1 = 19$$

## Caso Geral

Num primeiro passo do algoritmo um múltiplo adequado da primeira linha é subtraída de todas as outras equações do sistema de forma que os coeficientes de  $x_1$  se anulem, e dessa forma restando  $x_1$  somente na primeira equação. Isso é possível se  $a_{11} \neq 0$ , condição que pode ser obtida rearranjando as equações do sistema até que pelo menos um  $a_{i1} \neq 0$  (se for possível), como veremos mais adiante. Ao invés de trabalhar com as equações propriamente ditas, poderíamos efetuar as operações sobre a matriz estendida com o segundo membro que corresponde a um sistema linear completo.

$$(\mathbf{A}, \mathbf{b}) = \left[ \begin{array}{cccccc|c} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots & a_{3n} & b_3 \\ a_{41} & a_{42} & a_{43} & a_{44} & \cdots & a_{4n} & b_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nn} & b_n \end{array} \right] \begin{array}{l} \\ \text{(ii)} \leftarrow \text{(ii)} - \frac{a_{21}}{a_{11}} \text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{a_{31}}{a_{11}} \text{(i)} \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{a_{41}}{a_{11}} \text{(i)} \\ \\ \text{(n)} \leftarrow \text{(n)} - \frac{a_{n1}}{a_{11}} \text{(i)} \end{array}$$

Vejam as operações da primeira linha: Substituiremos a ( $2^{\text{a}}$  linha) por  $\left\{ (2^{\text{a}} \text{ linha}) - \frac{a_{21}}{a_{11}} (1^{\text{a}} \text{ linha}) \right\}$ , isto é

$$\begin{array}{l|l} \text{(ii)} & a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \cdots + a_{2n}x_n = b_2 \\ -\frac{a_{21}}{a_{11}} \text{(i)} & -\frac{a_{21}}{a_{11}}a_{11}x_1 - \frac{a_{21}}{a_{11}}a_{12}x_2 - \frac{a_{21}}{a_{11}}a_{13}x_3 \cdots - \frac{a_{21}}{a_{11}}a_{1n}x_n = -\frac{a_{21}}{a_{11}}b_1 \end{array}$$

---


$$\underbrace{\left\{ a_{21} - \frac{a_{21}}{a_{11}}a_{11} \right\}}_{=0} x_1 + \left\{ a_{22} - \frac{a_{21}}{a_{11}}a_{12} \right\} x_2 + \left\{ a_{23} - \frac{a_{21}}{a_{11}}a_{13} \right\} x_3 \cdots + \left\{ a_{2n} - \frac{a_{21}}{a_{11}}a_{1n} \right\} x_n = \left\{ b_2 - \frac{a_{21}}{a_{11}}b_1 \right\}$$

Observe que os múltiplos da primeira linha são calculados da seguinte maneira, é a razão entre o coeficiente que se deseja anular e o elemento  $a_{11}$  da diagonal da primeira linha.

Logo o primeiro passo do processo de eliminação de Gauss transforma o sistema acima  $(\mathbf{A}, \mathbf{b})$  num sistema da forma

$$(\mathbf{A}', \mathbf{b}') = \begin{bmatrix} a'_{11} & a'_{12} & a'_{13} & a'_{14} & \cdots & a'_{1n} & b'_1 \\ 0 & a'_{22} & a'_{23} & a'_{24} & \cdots & a'_{2n} & b'_2 \\ 0 & a'_{32} & a'_{33} & a'_{34} & \cdots & a'_{3n} & b'_3 \\ 0 & a'_{42} & a'_{43} & a'_{44} & \cdots & a'_{4n} & b'_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a'_{n2} & a'_{n3} & a'_{n4} & \cdots & a'_{nn} & b'_n \end{bmatrix}$$

esta primeira etapa pode ser descrita como:

**Etapa 1** Para  $k = 2, 3, \dots, n$ , subtraia da linha  $(k)$  o múltiplo

$$l_{k1} = \frac{a_{k1}}{a_{11}}$$

da linha (1) da matriz  $(\mathbf{A}, \mathbf{b})$ . O resultado será a matriz desejada  $(\mathbf{A}', \mathbf{b}')$ .

A transição  $(\mathbf{A}, \mathbf{b}) \rightarrow (\mathbf{A}', \mathbf{b}')$  pode ser descrita usando multiplicação de matrizes

$$(\mathbf{A}', \mathbf{b}') = \mathbf{G}_1(\mathbf{A}, \mathbf{b})$$

onde  $\mathbf{G}_1$  é uma matriz triangular inferior da seguinte forma

$$\mathbf{G}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -l_{21} & 1 & 0 & 0 & \cdots & 0 \\ -l_{31} & 0 & 1 & 0 & \cdots & 0 \\ -l_{41} & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -l_{n1} & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

onde os coeficientes  $l_{21}, l_{31}, l_{41}, \dots, l_{n1}$  são dados respectivamente por  $\frac{a_{21}}{a_{11}}, \frac{a_{31}}{a_{11}}, \frac{a_{41}}{a_{11}}, \dots, \frac{a_{n1}}{a_{11}}$ .

Matrizes tais como a acima  $(\mathbf{G}_1)$ , as quais diferem da matriz identidade em uma linha somente são chamadas matrizes de *Frobenius*.  $\mathbf{G}_1$  é não-singular. Observe que:

$$\mathbf{G}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ l_{21} & 1 & 0 & 0 & \cdots & 0 \\ l_{31} & 0 & 1 & 0 & \cdots & 0 \\ l_{41} & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Por esta razão os sistemas  $\mathbf{Ax} = \mathbf{b}$  e  $\mathbf{A}'\mathbf{x} = \mathbf{b}'$  têm a mesma solução:  $\mathbf{Ax} = \mathbf{b}$  implica que  $\mathbf{G}_1\mathbf{Ax} = \mathbf{A}'\mathbf{x} = \mathbf{b}' = \mathbf{G}_1\mathbf{b}$ , e  $\mathbf{A}'\mathbf{x} = \mathbf{b}'$  implica que  $\mathbf{G}_1^{-1}\mathbf{A}'\mathbf{x} = \mathbf{Ax} = \mathbf{b} = \mathbf{G}_1^{-1}\mathbf{b}'$ .

O elemento  $a_{11}$  na diagonal da primeira linha na Etapa 1 é chamado *elemento pivô* ou simplesmente *pivô*, a linha que contém o pivô é chamada *linha pivotal*.

Observe que o sistema resultante dessa primeira etapa não possui a variável  $x_1$  nas linhas abaixo da linha pivotal, isto é, as linhas 2, 3, 4, ..., n, tornam-se agora um sistema de ordem  $n - 1$ , independente da variável  $x_1$ . Logo, podemos repetir o procedimento anterior usando agora o elemento  $a'_{22}$  ( $a'_{22} \neq 0$ ) como pivô e a segunda linha como linha pivotal. Enfim, num segundo passo do algoritmo um múltiplo adequado da segunda linha é subtraído de todas as linhas abaixo da segunda linha, de forma que todos os elementos de  $x_2$  abaixo de  $a_{22}$  se anulem. Isto é,

$$(\mathbf{A}', \mathbf{b}') = \begin{bmatrix} a'_{11} & a'_{12} & a'_{13} & a'_{14} & \cdots & a'_{1n} & b'_1 \\ 0 & a'_{22} & a'_{23} & a'_{24} & \cdots & a'_{2n} & b'_2 \\ 0 & a'_{32} & a'_{33} & a'_{34} & \cdots & a'_{3n} & b'_3 \\ 0 & a'_{42} & a'_{43} & a'_{44} & \cdots & a'_{4n} & b'_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a'_{n2} & a'_{n3} & a'_{n4} & \cdots & a'_{nn} & b'_n \end{bmatrix} \begin{array}{l} \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{a'_{32}}{a'_{22}} \text{(ii)} \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{a'_{42}}{a'_{22}} \text{(ii)} \\ \\ \text{(n)} \leftarrow \text{(n)} - \frac{a'_{n2}}{a'_{22}} \text{(ii)} \end{array}$$

Logo o segundo passo do processo de eliminação de Gauss transforma o sistema acima  $(\mathbf{A}', \mathbf{b}')$  em um sistema da forma

$$(\mathbf{A}'', \mathbf{b}'') = \begin{bmatrix} a''_{11} & a''_{12} & a''_{13} & a''_{14} & \cdots & a''_{1n} & b''_1 \\ 0 & a''_{22} & a''_{23} & a''_{24} & \cdots & a''_{2n} & b''_2 \\ 0 & 0 & a''_{33} & a''_{34} & \cdots & a''_{3n} & b''_3 \\ 0 & 0 & a''_{43} & a''_{44} & \cdots & a''_{4n} & b''_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & a''_{n3} & a''_{n4} & \cdots & a''_{nn} & b''_n \end{bmatrix}$$

agora a segunda etapa pode ser descrita como:

**Etapa 2** Para  $k = 3, \dots, n$ , subtraia da linha ( $k$ ) o múltiplo

$$l_{k2} = \frac{a'_{k2}}{a'_{22}}$$

da linha (2) da matriz  $(\mathbf{A}', \mathbf{b}')$ . O resultado será a matriz desejada  $(\mathbf{A}'', \mathbf{b}'')$ .

A transição  $(\mathbf{A}', \mathbf{b}') \rightarrow (\mathbf{A}'', \mathbf{b}'')$  também pode ser descrita usando multiplicação de matrizes

$$(\mathbf{A}'', \mathbf{b}'') = \mathbf{G}_2(\mathbf{A}', \mathbf{b}')$$

onde  $\mathbf{G}_2$  é uma matriz triangular inferior da seguinte forma

$$\mathbf{G}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -l_{32} & 1 & 0 & \cdots & 0 \\ 0 & -l_{42} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -l_{n2} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

onde os coeficientes  $l_{32}, l_{42}, l_{52}, \dots, l_{n2}$  são dados respectivamente por  $\frac{a'_{32}}{a'_{22}}, \frac{a'_{42}}{a'_{22}}, \frac{a'_{52}}{a'_{22}}, \dots, \frac{a'_{n2}}{a'_{22}}$ .

Assim como a matriz  $\mathbf{G}_1$ ,  $\mathbf{G}_2$  é uma matriz de Frobenius e possui inversa da forma

$$\mathbf{G}_2^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & l_{32} & 1 & 0 & \cdots & 0 \\ 0 & l_{42} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & l_{n2} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Assim como na etapa anterior, podemos mostrar que  $\mathbf{A}'\mathbf{x} = \mathbf{b}'$  e  $\mathbf{A}''\mathbf{x} = \mathbf{b}''$  têm a mesma solução.

O elemento  $a'_{22}$  na diagonal da segunda linha na Etapa 2 agora é chamado *pivô* e a segunda linha é agora a *linha pivotal*.

Observando o sistema resultante vemos que a partir da terceira linha as equações não possuem nenhuma dependência das incógnitas  $x_1$  e  $x_2$  e portanto podemos visualizar um sistema de ordem  $n - 2$  com incógnitas  $x_3, x_4, \dots, x_n$  e então repetir o procedimento anterior.

$$(\mathbf{A}'', \mathbf{b}'') = \begin{bmatrix} a''_{11} & a''_{12} & a''_{13} & a''_{14} & \cdots & a''_{1n} & b''_1 \\ 0 & a''_{22} & a''_{23} & a''_{24} & \cdots & a''_{2n} & b''_2 \\ 0 & 0 & a''_{33} & a''_{34} & \cdots & a''_{3n} & b''_3 \\ 0 & 0 & a''_{43} & a''_{44} & \cdots & a''_{4n} & b''_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & a''_{n3} & a''_{n4} & \cdots & a''_{nn} & b''_n \end{bmatrix} \quad \begin{array}{l} \text{(iv)} \leftarrow \text{(iv)} - \frac{a''_{43}}{a''_{33}} \text{(iii)} \\ \text{(n)} \leftarrow \text{(n)} - \frac{a''_{n3}}{a''_{33}} \text{(iii)} \end{array}$$

Como nas etapas anteriores a eliminação de Gauss transforma o sistema acima  $(\mathbf{A}'', \mathbf{b}'')$  em um sistema da forma

$$(\mathbf{A}''', \mathbf{b}''') = \begin{bmatrix} a'''_{11} & a'''_{12} & a'''_{13} & a'''_{14} & \cdots & a'''_{1n} & b'''_1 \\ 0 & a'''_{22} & a'''_{23} & a'''_{24} & \cdots & a'''_{2n} & b'''_2 \\ 0 & 0 & a'''_{33} & a'''_{34} & \cdots & a'''_{3n} & b'''_3 \\ 0 & 0 & 0 & a'''_{44} & \cdots & a'''_{4n} & b'''_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & a'''_{n4} & \cdots & a'''_{nn} & b'''_n \end{bmatrix}$$

e a terceira etapa será descrita como:

**Etapa 3** Para  $k = 4, 5, \dots, n$ , subtraia da linha  $(k)$  o múltiplo

$$l_{k3} = \frac{a''_{k3}}{a''_{33}}$$

da linha (3) da matriz  $(\mathbf{A}'', \mathbf{b}'')$ . O resultado será o sistema desejado  $(\mathbf{A}''', \mathbf{b}''')$ .

A transição  $(\mathbf{A}'', \mathbf{b}'') \rightarrow (\mathbf{A}''', \mathbf{b}''')$  será descrita por

$$(\mathbf{A}''', \mathbf{b}''') = \mathbf{G}_3(\mathbf{A}'', \mathbf{b}'')$$

onde  $\mathbf{G}_3$  é da forma

$$\mathbf{G}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -l_{43} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & -l_{n3} & 0 & \cdots & 1 \end{bmatrix}$$

onde os coeficientes  $l_{43}, l_{53}, \dots, l_{n3}$  são dados por  $\frac{a''_{43}}{a''_{33}}, \frac{a''_{53}}{a''_{33}}, \dots, \frac{a''_{n3}}{a''_{33}}$  respectivamente.

Como  $\mathbf{G}_1$  e  $\mathbf{G}_2$ ,  $\mathbf{G}_3$  possui sua inversa da forma

$$\mathbf{G}_3^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & l_{43} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & l_{n3} & 0 & \cdots & 1 \end{bmatrix}$$

Mais uma vez podemos mostrar que  $\mathbf{A}''\mathbf{x} = \mathbf{b}''$  e  $\mathbf{A}'''\mathbf{x} = \mathbf{b}'''$  têm a mesma solução.

O pivô nesta etapa é o elemento  $a''_{33}$  na diagonal da terceira linha, e esta é chamada linha pivotal.

O sistema resultante a partir da quarta linha não possui dependência das incógnitas  $x_1, x_2$  e  $x_3$  sendo assim um sistema de ordem  $n - 3$  com incógnitas  $x_4, x_5, \dots, x_n$  e mais uma vez podemos repetir o procedimento no novo sistema reduzido.

Procedendo sucessivamente de maneira análoga, reduziremos o sistema original a um sistema triangular superior, que pode ser facilmente resolvido.

Usando a representação matricial da eliminação de Gauss temos

$$\begin{array}{rcl}
 \mathbf{Ax} = \mathbf{b} & \iff & \mathbf{Ax} = \mathbf{b} \\
 \mathbf{A}'\mathbf{x} = \mathbf{b}' & \iff & \mathbf{G}_1\mathbf{Ax} = \mathbf{G}_1\mathbf{b} \\
 \mathbf{A}''\mathbf{x} = \mathbf{b}'' & \iff & \mathbf{G}_2\mathbf{G}_1\mathbf{Ax} = \mathbf{G}_2\mathbf{G}_1\mathbf{b} \\
 \mathbf{A}'''\mathbf{x} = \mathbf{b}''' & \iff & \mathbf{G}_3\mathbf{G}_2\mathbf{G}_1\mathbf{Ax} = \mathbf{G}_3\mathbf{G}_2\mathbf{G}_1\mathbf{b} \\
 \vdots & & \vdots \\
 \mathbf{A}^{(n-1)}\mathbf{x} = \mathbf{b}^{(n-1)} & \iff & \mathbf{G}_{n-1}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{Ax} = \mathbf{G}_{n-1}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{b} \\
 \underbrace{\mathbf{U}} & & \underbrace{\mathbf{U}} \\
 \mathbf{Ux} = \mathbf{c} & \iff & \mathbf{Ux} = \mathbf{G}_{n-1}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{b}
 \end{array}$$

Premultiplicando o sistema acima pela inversa de  $\mathbf{G}_{n-1}$  que sabemos que existe

$$\mathbf{G}_{n-1}^{-1}\mathbf{Ux} = \underbrace{\mathbf{G}_{n-1}^{-1}\mathbf{G}_{n-1}}_{\mathbf{I}}\mathbf{G}_{n-2}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{b}$$

ou

$$\mathbf{G}_{n-1}^{-1}\mathbf{Ux} = \mathbf{G}_{n-2}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{b}$$

Analogamente premultiplicamos o sistema por  $\mathbf{G}_{n-2}^{-1}$

$$\mathbf{G}_{n-2}^{-1}\mathbf{G}_{n-1}^{-1}\mathbf{Ux} = \underbrace{\mathbf{G}_{n-2}^{-1}\mathbf{G}_{n-2}}_{\mathbf{I}}\mathbf{G}_{n-3}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{b}$$

logo

$$\mathbf{G}_{n-2}^{-1}\mathbf{G}_{n-1}^{-1}\mathbf{Ux} = \mathbf{G}_{n-3}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{b}$$

Da mesma forma podemos premultiplicar o sistema por  $\mathbf{G}_{n-3}^{-1}$

$$\mathbf{G}_{n-3}^{-1}\mathbf{G}_{n-2}^{-1}\mathbf{G}_{n-1}^{-1}\mathbf{Ux} = \underbrace{\mathbf{G}_{n-3}^{-1}\mathbf{G}_{n-3}}_{\mathbf{I}}\mathbf{G}_{n-4}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{b}$$

logo

$$\mathbf{G}_{n-3}^{-1}\mathbf{G}_{n-2}^{-1}\mathbf{G}_{n-1}^{-1}\mathbf{Ux} = \mathbf{G}_{n-4}\dots\mathbf{G}_2\mathbf{G}_1\mathbf{b}$$



Procedendo da mesma forma, premultiplicamos sucessivamente o sistema por  $\mathbf{G}_{n-4}^{-1}$ ,  $\mathbf{G}_{n-5}^{-1}$ ,  $\mathbf{G}_{n-6}^{-1}$ ,  $\dots$ ,  $\mathbf{G}_2^{-1}$ ,  $\mathbf{G}_1^{-1}$ , obtendo enfim

$$\mathbf{G}_1^{-1} \mathbf{G}_2^{-1} \mathbf{G}_3^{-1} \dots \mathbf{G}_{n-3}^{-1} \mathbf{G}_{n-2}^{-1} \mathbf{G}_{n-1}^{-1} \mathbf{U} \mathbf{x} = \mathbf{b}$$

Vimos que a inversa de cada matriz de Frobenius é fácil de calcular. Temos agora que calcular o produto dessas matrizes. Porém, cada matriz  $\mathbf{G}_i$  é triangular inferior, logo o produto delas é também triangular inferior, além disso devido a estrutura de cada uma delas este produto possui a forma

$$\begin{aligned} & \mathbf{G}_1^{-1} \mathbf{G}_2^{-1} \mathbf{G}_3^{-1} \dots \mathbf{G}_{n-3}^{-1} \mathbf{G}_{n-2}^{-1} \mathbf{G}_{n-1}^{-1} = \\ & = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & 0 & \dots & 0 \\ l_{31} & 0 & 1 & 0 & \dots & 0 \\ l_{41} & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & 0 & 0 & 0 & \dots & 1 \end{bmatrix}}_{\mathbf{G}_1^{-1}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & l_{32} & 1 & 0 & \dots & 0 \\ 0 & l_{42} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & l_{n2} & 0 & 0 & \dots & 1 \end{bmatrix}}_{\mathbf{G}_2^{-1}} \dots \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & l_{n,n-1} & 1 \end{bmatrix}}_{\mathbf{G}_{n-1}^{-1}} = \\ & = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ l_{21} & 1 & 0 & \dots & 0 & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 & 0 \\ l_{41} & l_{42} & l_{43} & \dots & \vdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{n,n-1} & 1 \end{bmatrix}}_{\mathbf{L}} = \mathbf{L} \end{aligned}$$

Observando a matriz  $\mathbf{L}$ , vemos que seus elementos são os valores dos múltiplos utilizados no processo de eliminação do método de eliminação de Gauss.

Decomposições triangulares tem grande importância na solução de sistemas de equações lineares. Se esta decomposição é conhecida para uma matriz  $\mathbf{A}$ , então o sistema

$$\mathbf{A} \mathbf{x} = \mathbf{b}$$

pode ser imediatamente resolvido para qualquer segundo membro  $\mathbf{b}$ . Isto é,

$$\mathbf{A} \mathbf{x} = \mathbf{b} \implies \mathbf{L} \mathbf{U} \mathbf{x} = \mathbf{b} \implies \begin{cases} \mathbf{L} \mathbf{y} = \mathbf{b} & \longrightarrow \mathbf{y} \\ \mathbf{U} \mathbf{x} = \mathbf{y} & \longrightarrow \mathbf{x} \end{cases}$$

**Observação:** Nem sempre toda matriz inversível possui uma decomposição da forma acima, às vezes, é necessária uma conveniente troca de linhas da matriz para a obtenção da fatoração.

Resumindo temos os seguintes algoritmos:

**Eliminação de Gauss:**

Para  $k = 1, 2, \dots, n - 1$

Para  $i = k + 1, \dots, n$

$$l_{ik} = a_{ik}/a_{kk}$$

Para  $j = k + 1, \dots, n$

$$a_{ij} = a_{ij} - l_{ik}a_{kj}$$

$$b_i = b_i - l_{ik}b_k$$

**Exemplo:**

$$\begin{cases} 3x_1 + x_2 + 6x_3 = 2 \\ 1x_1 + x_2 + x_3 = 4 \\ 2x_1 + x_2 + 3x_3 = 7 \end{cases}$$

**Solução:**

$$\left( \begin{array}{ccc|c} 3 & 1 & 6 & 2 \\ 1 & 1 & 1 & 4 \\ 2 & 1 & 3 & 7 \end{array} \right) \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} - 1/3\text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} - 2/3\text{(i)} \end{array}$$

$$\left( \begin{array}{ccc|c} 3 & 1 & 6 & 2 \\ 0 & 2/3 & -1 & 10/3 \\ 0 & 1/3 & -1 & 17/3 \end{array} \right) \text{(iii)} \leftarrow \text{(iii)} - 1/2\text{(ii)}$$

$$\left( \begin{array}{ccc|c} 3 & 1 & 6 & 2 \\ 0 & 2/3 & -1 & 10/3 \\ 0 & 0 & -1/2 & 4 \end{array} \right)$$

*Resp:* (19; -7; -8)

então

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 2/3 & 1/2 & 1 \end{pmatrix} \quad \text{e} \quad \mathbf{U} = \begin{pmatrix} 3 & 1 & 6 \\ 0 & 2/3 & -1 \\ 0 & 0 & -1/2 \end{pmatrix}$$

Se possuímos a fatoração  $\mathbf{LU}$  da matriz  $\mathbf{A}$  o sistema acima seria resolvido da seguinte maneira.

**Fatoração  $\mathbf{LU}$  com pivô nulo.**

$$\begin{cases} -x_1 + 2x_2 + 3x_3 + x_4 = 1 \\ 2x_1 - 4x_2 - 5x_3 - x_4 = 0 \\ -3x_1 + 8x_2 + 8x_3 + x_4 = 2 \\ x_1 + 2x_2 - 6x_3 + 4x_4 = -1 \end{cases}$$

$$\left[ \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 1 \\ 2 & -4 & -5 & -1 & 0 \\ -3 & 8 & 8 & 1 & 2 \\ 1 & 2 & -6 & 4 & -1 \end{array} \right] \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} - \frac{2}{-1}(\text{i}) \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{-3}{-1}(\text{i}) \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{1}{-1}(\text{i}) \end{array} \implies \mathbf{L} = \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 3 & ? & 1 & 0 \\ -1 & ? & ? & 1 \end{array} \right]$$

$$\left[ \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 2 & -1 & -2 & -1 \\ 0 & 4 & -3 & 5 & 0 \end{array} \right]$$

trocando a segunda e terceira linha contornamos o pivô nulo

$$\implies \mathbf{L} = \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -2 & ? & 1 & 0 \\ -1 & ? & ? & 1 \end{array} \right]$$

Observe que os elementos de  $\mathbf{L}$  abaixo da diagonal são também permutados.

$$\left[ \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 1 \\ 0 & 2 & -1 & -2 & -1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 4 & -3 & 5 & 0 \end{array} \right] \begin{array}{l} \text{(iii)} \leftarrow \text{(iii)} - \frac{0}{2}(\text{ii}) \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{4}{2}(\text{ii}) \end{array} \implies \mathbf{L} = \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ -1 & 2 & ? & 1 \end{array} \right]$$

$$\left[ \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 1 \\ 0 & 2 & -1 & -2 & -1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 9 & 2 \end{array} \right] \text{ (iv) } \leftarrow \text{ (iv) } - \frac{-1}{1} \text{ (iii)} \quad \Rightarrow \mathbf{L} = \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ -1 & 2 & -1 & 1 \end{array} \right]$$

$$\left[ \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 1 \\ 0 & 2 & -1 & -2 & -1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 10 & 4 \end{array} \right]$$

*Resp:*  $(\frac{28}{5}; \frac{7}{10}; \frac{8}{5}; \frac{2}{5})$

Resumindo temos a fatoração  $\mathbf{PA} = \mathbf{LU}$ .

$$\mathbf{L} = \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ -1 & 2 & -1 & 1 \end{array} \right] \quad \mathbf{U} = \left[ \begin{array}{cccc} -1 & 2 & 3 & 1 \\ 0 & 2 & -1 & -2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 10 \end{array} \right]$$

Calculando o produto  $\mathbf{LU}$

$$\left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ -1 & 2 & -1 & 1 \end{array} \right] \cdot \left[ \begin{array}{cccc} -1 & 2 & 3 & 1 \\ 0 & 2 & -1 & -2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 10 \end{array} \right] = \left[ \begin{array}{cccc} -1 & 2 & 3 & 1 \\ -3 & 8 & 8 & 1 \\ 2 & -4 & -5 & -1 \\ 1 & 2 & -6 & 4 \end{array} \right] = \mathbf{PA}$$

Veremos a seguir, que por razões numéricas trocamos linhas do sistema para utilizar um pivô melhor, até mesmo quando temos um pivô diferente de zero.

### 2.3.2 Sobre erros de arredondamento e Sistemas Mal Condicionados

Teoricamente a solução de um sistema linear não singular está perfeitamente estabelecida. Na eliminação de Gauss pode ser necessário algumas trocas de linhas, mas sempre será possível resolver corretamente (teoricamente) o sistema. A prática, porém é diferente.

Lembre-se que para um sistema linear de tamanho pequeno, digamos  $100 \times 100$ , a Eliminação de Gauss envolve aproximadamente 300.000 operações aritméticas ( $n^3/3$ ). Para cada operação podemos esperar um erro de arredondamento.

A questão é, *como estes erros contribuem para o erro final na solução?*

Os exemplos abaixo podem ilustrar alguns pontos importantes sobre os erros de arredondamento. Sejam duas matrizes,

$$\mathbf{A} = \begin{bmatrix} 1. & 1. \\ 1. & 1.0001 \end{bmatrix} \quad \text{e} \quad \mathbf{A}' = \begin{bmatrix} 0.0001 & 1. \\ 1. & 1. \end{bmatrix}$$

O primeiro ponto é: **Algumas matrizes são extremamente sensíveis à pequenas mudanças, e outras não.**

Qualitativamente  $\mathbf{A}$  é *quase* singular enquanto  $\mathbf{A}'$  não é. Se trocarmos o último elemento de  $\mathbf{A}$  para  $a_{22} = 1$ , a matriz se torna singular. Considere agora dois segundos membros *próximos* para o sistema  $\mathbf{Ax} = \mathbf{b}$

$$\begin{cases} u + v = 2 \\ u + 1.0001v = 2 \end{cases} \quad \text{e} \quad \begin{cases} u + v = 2 \\ u + 1.0001v = 2.0001 \end{cases}$$

A solução do primeiro sistema é  $u = 2, v = 0$ ; a solução do segundo sistema é  $u = v = 1$ . *Uma mudança na quinta casa decimal de  $b$  foi amplificada para uma mudança na primeira casa decimal da solução. Nenhum método numérico pode evitar tal sensibilidade a pequenas perturbações.* O mal-condicionamento pode se modificado de um lugar para outro, mas não pode ser removido. A solução real do sistema é muito sensível, e a solução calculada computacionalmente não pode ser menos sensível.

Exemplo gráfico:  $\implies$  retas quase paralelas.

O segundo ponto é: **Mesmo uma matriz bem-condicionada pode ser afetada por um algoritmo.**

Infelizmente para a matriz  $\mathbf{A}'$ , o processo de Eliminação de Gauss Clássica não é um bom algoritmo. Suponha que 0.0001 seja utilizado como o primeiro pivô, e 10000 vezes a primeira linha seja subtraída da segunda. O elemento na posição (22) da matriz tornar-se-á  $-9999$ , que arredondado se transforma em  $-10000$ . Os vestígios do valor 1 que estava originalmente naquela posição desapareceram.

Considere o exemplo:

$$\begin{cases} 0.0001u + v = 1 \\ u + v = 2. \end{cases} \quad (ii) \leftarrow (ii) - \frac{1}{0,0001}(i) \quad (ii) \leftarrow (ii) - 10000(i)$$

$$\begin{array}{r}
 u + v = 2 \\
 - \\
 u + 10000v = 10000 \\
 \hline
 - 9999v = -9998
 \end{array}$$

Após a eliminação a segunda equação poderia tornar-se

$$-9999v = -9998, \quad \text{ou} \quad v = 0.99990$$

Um arredondamento resultaria em  $-10000v = -10000$ , ou  $v = 1$ . Independentemente, a alteração da segunda equação não resultou em uma solução "errada" para  $v$ . Entretanto, quando é feita a retro-substituição a primeira equação com o valor correto de  $v$  torna-se

$$0.0001u + 0.9999 = 1 \quad \text{ou} \quad u = 1.$$

Se ao invés disso usarmos o valor  $v = 1$  que está incorreto somente na quarta casa decimal, teremos:

$$0.0001u + 1 = 1 \quad \text{ou} \quad u = 0.$$

O valor calculado computacionalmente é completamente errôneo. Mesmo a matriz  $\mathbf{A}'$  sendo bem-condicionada o processo de Eliminação de Gauss Clássico é extremamente instável.

O pivô de pequeno valor ( $=0.0001$ ) trouxe instabilidade, e a cura é a troca de linhas. Esse é o terceiro ponto:

**Teoricamente apenas os pivôs nulos forçam a troca de linhas na Eliminação de Gauss, e um pivô pequeno força uma troca prática.**

A menos que se tenha garantias especiais, um computador precisa comparar cada pivô com todos os possíveis pivô na mesma coluna. Escolhendo entre esses candidatos o com maior valor absoluto, e trocando as respectivas linhas tal que o maior valor seja o novo pivô, teremos a chamada *Eliminação de Gauss com pivoteamento Parcial*.

### Outro exemplo de Sistema Mal-Condicionado

Seja o sistema

$$\left\{ \begin{array}{l}
 10x_1 + 7x_2 + 8x_3 + 7x_4 = 32 \\
 7x_1 + 5x_2 + 6x_3 + 5x_4 = 23 \\
 8x_1 + 6x_2 + 10x_3 + 9x_4 = 33 \\
 7x_1 + 5x_2 + 9x_3 + 10x_4 = 31
 \end{array} \right.$$

se substituirmos  $\mathbf{x} = (9.2, -12.6, 4.5, -1, 1)$  no sistema acima obtemos, no lado esquerdo:

$$\begin{cases} b_1 = 32.1 \\ b_2 = 22.9 \\ b_3 = 33.1 \\ b_4 = 30.9 \end{cases}$$

o que nos leva a crer que  $\mathbf{x}$  é uma boa aproximação da solução do sistema.

Mas, se fizermos  $\mathbf{x} = (1.82, -0.36, 1.35, 0.79)$  obtemos

$$\begin{cases} b_1 = 32.01 \\ b_2 = 22.99 \\ b_3 = 33.01 \\ b_4 = 30.99 \end{cases}$$

que é outra aproximação da solução, porém, ainda longe do valor correto que é  $\mathbf{x} = (1, 1, 1, 1)$ .

## Estratégias de Pivoteamento

Vimos que quando encontramos um pivô nulo temos que procurar um outro elemento para substituí-lo. Entretanto, como também já vimos o método da eliminação de Gauss pode ser instável se utilizamos um pivô com valor muito pequeno, isto sugere que sempre seja feita uma seleção para a escolha de um novo pivô. Isto é feito geralmente de duas maneiras chamadas, *pivoteamento parcial* e *pivoteamento total*.

### Pivoteamento parcial

Nesta estratégia a pesquisa por um novo pivô é feita na mesma coluna do pivô natural (aquele que é usado na eliminação de Gauss clássica), os elementos candidatos ao cargo de pivô são os situados abaixo do pivô natural. O eleito será o de maior valor absoluto, isto é,

$$|a_{r1}| = \max_i |a_{i1}| \quad i = 1, 2, 3, \dots, n$$

então trocamos as linhas  $r$  e  $1$  tornando  $a_{r1}$  o novo pivô e continuamos a eliminação de Gauss.

Na Etapa 2 a pesquisa pelo novo pivô é feita agora na segunda coluna abaixo do pivô  $a'_{22}$ , isto é,

$$|a_{r2}| = \max_i |a_{i2}| \quad i = 2, 3, \dots, n$$

Na prática o método da eliminação de Gauss (ou Fatoração  $LU$ ) sempre é usado com alguma seleção de pivô, principalmente com a estratégia de pivoteamento parcial.

#### Observações:

- a) Quando calculamos a fatoração  $LU$  de uma matriz  $A$  usando pivoteamento parcial, devemos de alguma forma guardar as trocas de linhas realizadas na eliminação para a posterior trocas das linhas do vetor segundo membro.

### Pivoteamento total

No pivoteamento parcial a pesquisa por um novo pivô é restrita a uma coluna (aquela que contém o pivô natural). No pivoteamento total a pesquisa é feita em toda a matriz de coeficientes e novo pivô será aquele que tiver o maior valor absoluto. Neste caso após a escolha devemos trocar convenientemente as linhas e colunas do sistema para continuar o processo de eliminação.

#### Observações:

- a) Como nesta estratégia de pivoteamento trocamos colunas do sistema deve-se atentar para a troca na ordem das incógnitas visto que estas estão associadas a cada coluna do sistema.
- b) O pivoteamento total é muito pouco utilizado devido ao alto custo da pesquisa por um novo pivô, na  $k$ -ésima etapa da fatoração são necessários  $(n - k + 1)^2$  testes para se determinar quem será o novo pivô, enquanto que no pivoteamento parcial o número de testes necessários é igual a  $(n - k + 1)$ , visto que a procura é restrita a uma única coluna.



### 2.3.3 Custo da Eliminação Gaussiana

*Quantas operações aritméticas são necessárias para se resolver um sistema de  $n$  equações e  $n$  incógnitas?*

Suporemos que não sejam necessárias trocas de linhas, isto é, não existam pivôs nulos nem os erros de arredondamento.

Vejam os em primeiro lugar somente as operações na matriz (ignoraremos inicialmente as operações no segundo membro).

Essas operações são de dois tipos.

Uma operação é a divisão pelo pivô, para encontrar o múltiplo da linha pivotal que deverá ser subtraído. E a seguir, efetuamos realmente essa subtração.

Consideremos cada divisão e cada multiplicação–subtração uma única operação.

No início, quando a primeira equação tem  $n$  elementos na matriz, são necessárias  $n$  operações para cada zero(0) obtido na primeira coluna de cada linha ( 1 operação para determinarmos o múltiplo (divisão) e  $n - 1$  para os outros elementos da linha fatorada (multiplicações e subtrações)). Existem  $n-1$  linhas abaixo da primeira linha, assim a primeira etapa da eliminação necessita de  $n(n-1) = n^2 - n$  operações (outra abordagem é a seguinte: todos os  $n^2$  elementos precisam ser modificados, menos os  $n$  elementos da primeira linha  $\rightarrow n^2 - n$ ). Agora, observe que as etapas seguintes são mais "rápidas", porque as equações se tornam progressivamente menores. Quando a eliminação é feita em  $k$  equações ( $k < n$ ), somente  $k^2 - k$  operações são necessárias para zerar a coluna abaixo do pivô (pela mesma razão usada na primeira linha).

Juntas, o número de operações na matriz é a soma de  $k^2 - k$  operações para  $k$  variando de 1 até  $n$ . Isto é,

$$\begin{aligned} \sum_{k=1}^n k^2 - k &= \sum_{k=1}^n k^2 - \sum_{k=1}^n k = \\ &= (1^2 + 2^2 + \dots + n^2) - (1 + 2 + \dots + n) = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2} = \frac{n^3 - n}{3} \approx \frac{n^3}{3} \end{aligned}$$

#### Observação:

Se o  $n$  é muito grande, uma estimativa do número de operações é  $\frac{1}{3}n^3$ .

No segundo membro na primeira etapa temos  $n - 1$  operações (uma multiplicação–subtração para cada elemento do segundo membro). Na segunda etapa como temos um sistema  $(n - 1) \times (n - 1)$ , serão necessárias  $n - 2$  operações, e assim por diante. Logo, somando todas as operações, teremos

$$\begin{aligned} (n - 1) + (n - 2) + \dots + 2 + 1 &= 1 + 2 + \dots + (n - 1) = \frac{(n - 1)((n - 1) + 1)}{2} = \frac{n(n - 1)}{2} = \\ &= \frac{n^2 - n}{2} \approx \frac{n^2}{2} \end{aligned}$$

#### Observação:

Se o  $n$  é muito grande, uma estimativa do número de operações no segundo membro é  $\frac{1}{2}n^2$ .

A retro-substituição é bem mais "rápida".

A última incógnita ( $x_n$ ) é determinada com somente uma operação (uma divisão). A penúltima incógnita requer 2 operações (uma subtração–multiplicação e uma divisão) e assim por diante. Logo o número total de operações para a retrosubstituição é

$$1 + 2 + \dots + n = \frac{n(n+1)}{2} = \frac{n^2 + n}{2} \approx \frac{n^2}{2}$$

Somando todas as operações efetuadas, temos:

$$\frac{n^3 - n}{3} + \frac{n^2 - n}{2} + \frac{n^2 + n}{2} = \frac{n^3 - n}{3} + n^2$$

**Observação:**

Novamente, podemos dizer, se  $n$  for muito grande que o número de operações necessários para se resolver um sistema por Eliminação de Gauss é  $\frac{1}{3}n^3$ .

## Exemplos

### Eliminação de Gauss

$$\left\{ \begin{array}{l} 2x_1 + 2x_2 + x_3 + x_4 = 7 \\ x_1 - x_2 + 2x_3 - x_4 = 1 \\ 3x_1 + 2x_2 - 3x_3 - 2x_4 = 4 \\ 4x_1 + 3x_2 + 2x_3 + x_4 = 12 \end{array} \right.$$

$$\left[ \begin{array}{cccc|c} 2 & 2 & 1 & 1 & 7 \\ 1 & -1 & 2 & -1 & 1 \\ 3 & 2 & -3 & -2 & 4 \\ 4 & 3 & 2 & 1 & 12 \end{array} \right] \begin{array}{l} \\ \text{(ii)} \leftarrow \text{(ii)} - \frac{1}{2}\text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{3}{2}\text{(i)} \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{4}{2}\text{(i)} \end{array}$$

$$\left[ \begin{array}{cccc|c} 2 & 2 & 1 & 1 & 7 \\ 0 & -2 & \frac{3}{2} & -\frac{3}{2} & -\frac{5}{2} \\ 0 & -1 & -\frac{9}{2} & -\frac{7}{2} & -\frac{13}{2} \\ 0 & -1 & 0 & -1 & -2 \end{array} \right] \begin{array}{l} \\ \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{-1}{-2}\text{(ii)} \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{-1}{-2}\text{(ii)} \end{array}$$

$$\left[ \begin{array}{cccc|c} 2 & 2 & 1 & 1 & 7 \\ 0 & -2 & \frac{3}{2} & -\frac{3}{2} & -\frac{5}{2} \\ 0 & 0 & -\frac{21}{4} & -\frac{11}{4} & -\frac{21}{4} \\ 0 & 0 & -\frac{3}{4} & -\frac{1}{4} & -\frac{3}{4} \end{array} \right] \begin{array}{l} \\ \\ \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{-3/4}{-21/4}\text{(iii)} \end{array}$$

$$\left[ \begin{array}{cccc|c} 2 & 2 & 1 & 1 & 7 \\ 0 & -2 & \frac{3}{2} & -\frac{3}{2} & -\frac{5}{2} \\ 0 & 0 & -\frac{21}{4} & -\frac{11}{4} & -\frac{21}{4} \\ 0 & 0 & 0 & \frac{1}{7} & 0 \end{array} \right]$$

$$\left\{ \begin{array}{l} 2x_1 + 2x_2 + x_3 + x_4 = 7 \\ -2x_2 + \frac{3}{2}x_3 - \frac{3}{2}x_4 = -\frac{5}{2} \\ -\frac{21}{4}x_3 - \frac{11}{4}x_4 = -\frac{21}{4} \\ \frac{1}{7}x_4 = 0 \end{array} \right.$$

Da quarta equação  $\rightarrow x_4 = 0$

$$\text{Da terceira equação } \rightarrow x_3 = \frac{1}{-\frac{21}{4}} \left\{ -\frac{21}{4} + \frac{11}{4}x_4 \right\} = \frac{1}{-\frac{21}{4}} \left\{ -\frac{21}{4} + 0 \right\} = 1$$

$$\text{Da segunda equação } \rightarrow x_2 = -\frac{1}{2} \left\{ -\frac{5}{2} - \frac{3}{2}x_3 + \frac{3}{2}x_4 \right\} = -\frac{1}{2} \left\{ -\frac{5}{2} - \frac{3}{2} \cdot 1 + \frac{3}{2} \cdot 0 \right\} = \frac{8}{4} = 2$$

$$\text{Da primeira equação } \rightarrow x_1 = \frac{1}{2} \{7 - 2x_2 - x_3 - x_4\} = \frac{1}{2} \{7 - 2 \cdot 2 - 1 - 0\} = 1$$

*Resp:* (1; 2; 1; 0)

**Fatoração  $LU$  c/ e s/ pivoteamento**

1) Calcular a fatoração  $LU$  da matriz abaixo e calcular seu determinante

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 4 & 0 \\ 4 & -1 & 5 & 1 \\ -2 & 2 & -2 & 3 \\ 0 & 3 & -9 & 4 \end{bmatrix}$$

usaremos as posições da matriz abaixo da diagonal principal que forem sendo *zeradas* para armazenar os elementos da matriz  $\mathbf{L}$ .

$$\begin{bmatrix} 2 & -1 & 4 & 0 \\ 4 & -1 & 5 & 1 \\ -2 & 2 & -2 & 3 \\ 0 & 3 & -9 & 4 \end{bmatrix} \begin{array}{l} \\ \text{(ii)} \leftarrow \text{(ii)} - \frac{4}{2}(\text{i}) \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{-2}{2}(\text{i}) \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{0}{2}(\text{i}) \end{array}$$

$$\begin{bmatrix} 2 & -1 & 4 & 0 \\ \hline 2 & | & 1 & -3 & 1 \\ -1 & | & 1 & 2 & 3 \\ 0 & | & 3 & -9 & 4 \end{bmatrix} \begin{array}{l} \\ \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{1}{1}(\text{ii}) \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{3}{1}(\text{ii}) \end{array}$$

$$\begin{bmatrix} 2 & -1 & 4 & 0 \\ \hline 2 & | & 1 & -3 & 1 \\ -1 & 1 & | & 5 & 2 \\ 0 & 3 & | & 0 & 1 \end{bmatrix} \begin{array}{l} \\ \\ \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{0}{5}(\text{iii}) \end{array}$$

$$\left[ \begin{array}{cccc|cccc} 2 & -1 & 4 & 0 & & & & & \\ \hline & 2 & & 1 & -3 & 1 & & & \\ & & & 1 & & & & & \\ -1 & & & & 5 & 2 & & & \\ & & & & & & & & \\ 0 & 3 & 0 & & & & 1 & & \end{array} \right]$$

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{U} = \begin{bmatrix} 2 & -1 & 4 & 0 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

logo

$$\det \mathbf{A} = \det \mathbf{LU} = \det \mathbf{L} \cdot \det \mathbf{U} = 1 \cdot \det \mathbf{U} = \det \mathbf{R} = 2 \times 1 \times 5 \times 1 = 10$$

como o determinante da matriz é diferente de zero então o sistema linear abaixo possui uma única solução.

$$\begin{bmatrix} 2 & -1 & 4 & 0 \\ 4 & -1 & 5 & 1 \\ -2 & 2 & -2 & 3 \\ 0 & 3 & -9 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \\ 1 \\ -2 \end{bmatrix} \quad \begin{bmatrix} 2 & -1 & 4 & 0 \\ 4 & -1 & 5 & 1 \\ -2 & 2 & -2 & 3 \\ 0 & 3 & -9 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 10 \\ 50 \\ 1 \\ 2 \end{bmatrix}$$

ache a solução deste sistema usando a decomposição  $\mathbf{LU}$  de  $\mathbf{A}$ .

$$\mathbf{Ax} = \mathbf{b} \implies \mathbf{LUx} = \mathbf{b} \implies \begin{cases} \mathbf{Ly} = \mathbf{b} & \longrightarrow \mathbf{y} \\ \mathbf{Ux} = \mathbf{y} & \longrightarrow \mathbf{x} \end{cases}$$

$$\mathbf{Ly} = \mathbf{b} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \\ 1 \\ -2 \end{bmatrix} \longrightarrow \begin{aligned} y_1 &= 5 \\ y_2 &= -1 \\ y_3 &= 7 \\ y_4 &= 1 \end{aligned}$$

$$Ux = y \longrightarrow \begin{bmatrix} 2 & -1 & 4 & 0 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 7 \\ 1 \end{bmatrix} \begin{array}{l} \longrightarrow x_1 = 1 \\ \longrightarrow x_2 = 1 \\ \longrightarrow x_3 = 1 \\ \longrightarrow x_4 = 1 \end{array}$$

Agora repetiremos o problema usando pivoteamento parcial.

$$\begin{bmatrix} 2 & -1 & 4 & 0 \\ 4 & -1 & 5 & 1 \\ -2 & 2 & -2 & 3 \\ 0 & 3 & -9 & 4 \end{bmatrix}$$

Com pivoteamento parcial precisamos trocar a primeira com a segunda linha para que novo pivô seja o 4.

$$\begin{bmatrix} 4 & -1 & 5 & 1 \\ 2 & -1 & 4 & 0 \\ -2 & 2 & -2 & 3 \\ 0 & 3 & -9 & 4 \end{bmatrix} \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} - \frac{2}{4} \text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{-2}{4} \text{(i)} \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{0}{4} \text{(i)} \end{array}$$

$$\begin{bmatrix} 4 & -1 & 5 & 1 \\ \hline 1/2 & | & -1/2 & 3/2 & -1/2 \\ -1/2 & | & 3/2 & 1/2 & 7/2 \\ 0 & | & 3 & -9 & 4 \end{bmatrix}$$

Precisamos agora de nova troca de linhas, isto é, a segunda linha será trocada com a quarta linha

$$\left[ \begin{array}{cccc|cccc} 4 & -1 & 5 & 1 & & & & \\ \hline & 0 & 3 & -9 & 4 & & & \\ & & & & & & & \\ -1/2 & 3/2 & 1/2 & 7/2 & & & & \\ & & & & & & & \\ 1/2 & -1/2 & 3/2 & -1/2 & & & & \end{array} \right] \begin{array}{l} \\ \\ \\ \text{(iii)} \leftarrow \text{(iii)} - \frac{3/2}{3}(\text{ii}) \\ \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{-1/2}{3}(\text{ii}) \end{array}$$

$$\left[ \begin{array}{cccc|cccc} 4 & -1 & 5 & 1 & & & & \\ \hline & 0 & 3 & -9 & 4 & & & \\ & & & & & & & \\ -1/2 & 1/2 & 5 & 3/2 & & & & \\ & & & & & & & \\ 1/2 & -1/6 & 0 & 1/6 & & & & \end{array} \right] \begin{array}{l} \\ \\ \\ \\ \text{(iv)} \leftarrow \text{(iv)} - \frac{0}{5}(\text{iii}) \end{array}$$

Temos enfim

$$\left[ \begin{array}{cccc|cccc} 4 & -1 & 5 & 1 & & & & \\ \hline & 0 & 3 & -9 & 4 & & & \\ & & & & & & & \\ -1/2 & 1/2 & 5 & 3/2 & & & & \\ & & & & & & & \\ 1/2 & -1/6 & 0 & 1/6 & & & & \end{array} \right]$$

e

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1/2 & 1/2 & 1 & 0 \\ 1/2 & -1/6 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{U} = \begin{bmatrix} 4 & -1 & 5 & 1 \\ 0 & 3 & -9 & 4 \\ 0 & 0 & 5 & 3/2 \\ 0 & 0 & 0 & 1/6 \end{bmatrix}$$



se calcularmos o produto  $\mathbf{LU}$  obtemos

$$\mathbf{LU} = \begin{bmatrix} 4 & -1 & 5 & 1 \\ 0 & 3 & -9 & 4 \\ -2 & 2 & -2 & 3 \\ 2 & -1 & 4 & 0 \end{bmatrix} = \mathbf{PA}$$

que não é igual a matriz  $\mathbf{A}$ , porém o resultado é igual a  $\mathbf{A}$  com as mesmas trocas de linhas realizadas durante a fatoração.

No produto  $\mathbf{LU}$ ,  $\mathbf{P}$  é a matriz de permutação relativa às trocas de linhas e  $\mathbf{P} = \mathbf{P}_2\mathbf{P}_1$ .  $\mathbf{P}_1$  representa a troca das 1ª e 2ª linhas e  $\mathbf{P}_2$  representa a troca das 2ª e 4ª linhas. Na forma matricial temos

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{P}_2\mathbf{P}_1$$

Para usarmos a fatoração acima para resolver um sistema linear temos que permutar as linhas do segundo membro da mesma forma que foram feitas na matriz durante a fatoração. Logo inicialmente trocamos a primeira e a segunda linhas e a seguir trocamos a segunda e quarta linhas, resultando no seguinte segundo membro.

$$\mathbf{b}' = \begin{bmatrix} 9 \\ -2 \\ 1 \\ 5 \end{bmatrix}$$

## 2.4 Métodos Iterativos

Muitos problemas práticos requerem a solução de grandes sistemas lineares ( $\mathbf{Ax} = \mathbf{b}$ ) em que a matriz  $\mathbf{A}$  é esparsa, isto é, tem relativamente poucos elementos não nulos. Sistemas desse tipo surgem, por exemplo, em aplicações dos métodos de diferenças finitas ou elementos finitos para aproximar a solução de problemas de valor de contorno em equações diferenciais parciais. Os métodos de eliminação usuais, normalmente não podem ser empregados neste caso (exceção feita as matrizes do tipo *banda*, quando a largura de banda é pequena), pois eles tendem a gerar matrizes intermediárias densas e o número de operações necessárias para a solução torna-se muito grande, mesmo para os computadores modernos, além disso tais matrizes ocupariam uma memória, às vezes, não disponível, além disso existem os erros de arredondamento. Por estas e outras razões utiliza-se os métodos iterativos para resolver tais sistemas.

Nesta classe de métodos partimos de uma aproximação inicial  $\mathbf{x}^{(0)}$  para a solução do sistema linear, e a partir dela geramos por um processo recursivo (repetitivo) novas aproximações que definem uma seqüência. O método será bem sucedido se a seqüência convergir para a solução do sistema. Cada ciclo responsável pela geração de uma nova aproximação é chamado *iteração*. A cada iteração são usadas informações das iterações anteriores. Resumindo

$$\mathbf{x}^{(0)} \longrightarrow \mathbf{x}^{(1)} \longrightarrow \mathbf{x}^{(2)} \longrightarrow \mathbf{x}^{(3)} \longrightarrow \dots \longrightarrow \mathbf{x}$$

$$\lim_{i \rightarrow \infty} \mathbf{x}^{(i)} = \mathbf{x}$$

Obviamente é impossível realizar infinitas iterações para obter a solução do sistema. Por isso adotaremos um *critério de parada*.

### 2.4.1 Critério de Parada

Uma maneira de determinar se a seqüência gerada pelo método iterativo está convergindo é verificar se a diferença entre a aproximação e o vetor exato da solução está diminuindo. Seja  $\mathbf{x}^{(i)}$  o vetor aproximado obtido na iteração  $i$  e  $\bar{\mathbf{x}}$  o vetor solução, podemos verificar se

$$\|\mathbf{x}^{(i)} - \bar{\mathbf{x}}\| < \epsilon$$

isto é, se a diferença acima citada é menor que um valor *pequeno* ( $\epsilon$ ), onde  $\|\cdot\|$  denota uma norma do  $\mathbb{R}^n$ . Chamamos o valor escalar  $\epsilon$  de tolerância de parada ou simplesmente tolerância.

Entretanto, como não conhecemos a solução do sistema (queremos na verdade determinar este vetor), o teste acima não pode ser efetuado. Substituímos o teste acima pelo seguinte teste:

$$\|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| < \epsilon$$

A diferença agora é calculada entre duas aproximações sucessivas.

O critério empregando a diferença relativa entre aproximações sucessivas também é usado, isto é,

$$\frac{\|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|}{\|\mathbf{x}^{(i+1)}\|} < \epsilon$$

A seção a seguir se dedica à apresentar brevemente os métodos iterativos para resolução de sistemas lineares. Nos restringiremos aos métodos clássicos.

## 2.4.2 Método de Jacobi

No método de Jacobi partimos do sistema original abaixo

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + \cdots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + \cdots + a_{3n}x_n = b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + \cdots + a_{4n}x_n = b_4 \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + a_{n4}x_4 + \cdots + a_{nn}x_n = b_n \end{array} \right.$$

e o reescrevemos da maneira descrita a seguir.

Da primeira linha explicitamos a incógnita  $x_1$ , da segunda linha a incógnita  $x_2$ , da terceira linha a incógnita  $x_3$ , e assim por diante até a última equação do sistema, onde então, explicitamos a incógnita  $x_n$ . Resultando assim,

$$\left\{ \begin{array}{l} x_1 = \frac{1}{a_{11}} (b_1 - a_{12}x_2 - a_{13}x_3 - a_{14}x_4 - \cdots - a_{1n}x_n) \\ x_2 = \frac{1}{a_{22}} (b_2 - a_{21}x_1 - a_{23}x_3 - a_{24}x_4 - \cdots - a_{2n}x_n) \\ x_3 = \frac{1}{a_{33}} (b_3 - a_{31}x_1 - a_{32}x_2 - a_{34}x_4 - \cdots - a_{3n}x_n) \\ x_4 = \frac{1}{a_{44}} (b_4 - a_{41}x_1 - a_{42}x_2 - a_{43}x_3 - \cdots - a_{4n}x_n) \\ \vdots \quad \quad \quad \vdots \\ x_n = \frac{1}{a_{nn}} (b_n - a_{n1}x_1 - a_{n2}x_2 - a_{n3}x_3 - \cdots - a_{n,n-1}x_{n-1}) \end{array} \right. \quad (2.5)$$

A função de iteração é então definida como

$$\left\{ \begin{array}{l} x_1^{(i+1)} = \frac{1}{a_{11}} \left( b_1 - a_{12}x_2^{(i)} - a_{13}x_3^{(i)} - a_{14}x_4^{(i)} - \dots - a_{1n}x_n^{(i)} \right) \\ x_2^{(i+1)} = \frac{1}{a_{22}} \left( b_2 - a_{21}x_1^{(i)} - a_{23}x_3^{(i)} - a_{24}x_4^{(i)} - \dots - a_{2n}x_n^{(i)} \right) \\ x_3^{(i+1)} = \frac{1}{a_{33}} \left( b_3 - a_{31}x_1^{(i)} - a_{32}x_2^{(i)} - a_{34}x_4^{(i)} - \dots - a_{3n}x_n^{(i)} \right) \\ x_4^{(i+1)} = \frac{1}{a_{44}} \left( b_4 - a_{41}x_1^{(i)} - a_{42}x_2^{(i)} - a_{43}x_3^{(i)} - \dots - a_{4n}x_n^{(i)} \right) \\ \vdots \quad \quad \quad \vdots \\ x_n^{(i+1)} = \frac{1}{a_{nn}} \left( b_n - a_{n1}x_1^{(i)} - a_{n2}x_2^{(i)} - a_{n3}x_3^{(i)} - \dots - a_{n,n-1}x_{n-1}^{(i)} \right) \end{array} \right. \quad (2.6)$$

Para  $i = 0, 1, 2, \dots$

Para  $k = 1, 2, \dots, n$

$$x_k^{(i+1)} = \left\{ b_k - \sum_{j=1}^{k-1} a_{kj}x_j^{(i)} - \sum_{j=k+1}^n a_{kj}x_j^{(i)} \right\} / a_{kk}$$

Testa Criterio de Parada

com  $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)})$  dado

Assim se tivermos uma aproximação inicial para a solução do sistema, podemos gerar uma sequência de aproximações que esperamos convirja para a solução.

### Um Critério de Convergência (Jacobi)

Mostraremos a seguir uma condição suficiente para a convergência do Método de Jacobi.

Por ser um critério somente suficiente asseguramos a convergência se as condições do teorema são satisfeitas. Entretanto, nada podemos afirmar se as condições **não** são satisfeitas.

**Teorema 2.4.1 (Critério das Linhas ou Critério da Diagonal Dominante)** *Seja um sistema linear  $\mathbf{Ax} = \mathbf{b}$  e seja*

$$\alpha_1 = \frac{|a_{12}| + |a_{13}| + |a_{14}| + \dots + |a_{1n}|}{|a_{11}|}$$

$$\alpha_2 = \frac{|a_{21}| + |a_{23}| + |a_{24}| + \dots + |a_{2n}|}{|a_{22}|}$$

$$\alpha_3 = \frac{|a_{31}| + |a_{32}| + |a_{34}| + \cdots + |a_{3n}|}{|a_{33}|}$$

$$\vdots$$

$$\alpha_n = \frac{|a_{n1}| + |a_{n2}| + |a_{n3}| + \cdots + |a_{n,n-1}|}{|a_{nn}|}$$

isto é

$$\alpha_i = \left( \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \right) / |a_{ii}| \quad i = 1, \dots, n$$

Se

$$\alpha = \max_{1 \leq i \leq n} \alpha_i < 1$$

então o método de Jacobi converge para a solução do sistema linear, independentemente da aproximação inicial.

**Exemplo:**

$$\begin{cases} 4x_1 - x_2 + x_3 = 4 \\ x_1 + 6x_2 + 2x_3 = 9 \\ -x_1 - 2x_2 + 5x_3 = 2 \end{cases}$$

Pelo critério das linhas

$$\alpha_1 = \frac{|a_{12}| + |a_{13}|}{|a_{11}|} = \frac{1+1}{4} = \frac{1}{2} < 1$$

$$\alpha_2 = \frac{|a_{21}| + |a_{23}|}{|a_{22}|} = \frac{1+2}{6} = \frac{1}{2} < 1$$

$$\alpha_3 = \frac{|a_{31}| + |a_{32}|}{|a_{33}|} = \frac{1+2}{5} = \frac{3}{5} < 1$$

logo temos certeza de que o método de Jacobi convergirá. Reescrevendo o sistema linear

$$\begin{cases} x_1^{(k+1)} = \frac{1}{4} (4 + x_2^{(k)} - x_3^{(k)}) \\ x_2^{(k+1)} = \frac{1}{6} (9 - x_1^{(k)} - 2x_3^{(k)}) \\ x_3^{(k+1)} = \frac{1}{5} (2 + x_1^{(k)} + 2x_2^{(k)}) \end{cases}$$

com  $\mathbf{x}^{(0)} = \vec{0}$  e  $\epsilon = 0.01$

$$\begin{cases} x_1^{(1)} = \frac{1}{4} (4 + x_2^{(0)} - x_3^{(0)}) = \frac{1}{4} (4 + 0 - 0) = 1 \\ x_2^{(1)} = \frac{1}{6} (9 - x_1^{(0)} - 2x_3^{(0)}) = \frac{1}{6} (9 - 0 - 2 \cdot 0) = \frac{9}{6} \\ x_3^{(1)} = \frac{1}{5} (2 + x_1^{(0)} + 2x_2^{(0)}) = \frac{1}{5} (2 + 0 + 2 \cdot 0) = \frac{2}{5} \end{cases}$$

$$\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{9}{6} \\ \frac{2}{5} \end{bmatrix}$$

Temos que verificar se o critério de parada foi satisfeito

$$\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| < \epsilon \text{ ?????}$$

$$\left\| \begin{bmatrix} 1 \\ \frac{9}{6} \\ \frac{2}{5} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 1 \\ \frac{9}{6} \\ \frac{2}{5} \end{bmatrix} \right\| = \frac{9}{6} = 1.5 > \epsilon$$

faremos então mais uma iteração

$$\begin{cases} x_1^{(2)} = \frac{1}{4} (4 + x_2^{(1)} - x_3^{(1)}) = \frac{1}{4} (4 + \frac{9}{6} - \frac{2}{5}) = \frac{51}{40} \\ x_2^{(2)} = \frac{1}{6} (9 - x_1^{(1)} - 2x_3^{(1)}) = \frac{1}{6} (9 - 1 - 2 \cdot \frac{2}{5}) = \frac{36}{30} \\ x_3^{(2)} = \frac{1}{5} (2 + x_1^{(1)} + 2x_2^{(1)}) = \frac{1}{5} (2 + 1 + 2 \cdot \frac{9}{6}) = \frac{6}{5} \end{cases}$$

$$\begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{bmatrix} = \begin{bmatrix} \frac{51}{40} \\ \frac{36}{30} \\ \frac{6}{5} \end{bmatrix}$$

Pelo critério de parada

$$\|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}\| < \epsilon \text{ ?????}$$

$$\left\| \begin{bmatrix} \frac{51}{40} \\ \frac{36}{30} \\ \frac{6}{5} \end{bmatrix} - \begin{bmatrix} 1 \\ \frac{9}{6} \\ \frac{2}{5} \end{bmatrix} \right\| = \left\| \begin{bmatrix} \frac{11}{40} \\ \frac{-9}{30} \\ \frac{4}{5} \end{bmatrix} \right\| = \frac{4}{5} = 0.8 > \epsilon$$

mais uma iteração. E assim por diante.

**Outro Exemplo:**

$$\begin{cases} 10x_1 + 2x_2 + x_3 = 7 \\ x_1 + 5x_2 + x_3 = -8 \\ 2x_1 + 3x_2 + 10x_3 = 6 \end{cases}$$

Reescrevendo o sistema linear

$$\begin{cases} x_1^{(k+1)} = \frac{1}{10} (7 - 2x_2^{(k)} - x_3^{(k)}) \\ x_2^{(k+1)} = \frac{1}{5} (-8 - x_1^{(k)} - x_3^{(k)}) \\ x_3^{(k+1)} = \frac{1}{10} (6 - 2x_1^{(k)} - 3x_2^{(k)}) \end{cases}$$

com  $[\mathbf{x}^{(0)}]^T = (0.7; -1.6; 0.6)$  e  $\epsilon = 0.05$ , temos:

$$[\mathbf{x}^{(0)}]^T = (0.7000; -1.6000; 0.6000)$$

$$[\mathbf{x}^{(1)}]^T = (0.9780; -1.9800; 0.9660) \implies \|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}\| = 0.3800 > \epsilon$$

$$[\mathbf{x}^{(2)}]^T = (0.9994; -1.9888; 0.9984) \implies \|\mathbf{x}^{(3)} - \mathbf{x}^{(2)}\| = 0.0324 < \epsilon$$

**2.4.3 Método de Gauss-Seidel**

Novamente partindo do sistema linear reescrito na forma 2.5. Assim como no método do Jacobi as últimas aproximações para os componentes são usadas para calcular as novas aproximações. Entretanto, a medida que novos valores são obtidos para os componentes estes são usados ao invés daqueles da aproximação anterior.

$$\begin{cases} x_1^{(i+1)} = \frac{1}{a_{11}} (b_1 - a_{12}x_2^{(i)} - a_{13}x_3^{(i)} - a_{14}x_4^{(i)} - \dots - a_{1n}x_n^{(i)}) \\ x_2^{(i+1)} = \frac{1}{a_{22}} (b_2 - a_{21}x_1^{(i+1)} - a_{23}x_3^{(i)} - a_{24}x_4^{(i)} - \dots - a_{2n}x_n^{(i)}) \\ x_3^{(i+1)} = \frac{1}{a_{33}} (b_3 - a_{31}x_1^{(i+1)} - a_{32}x_2^{(i+1)} - a_{34}x_4^{(i)} - \dots - a_{3n}x_n^{(i)}) \\ x_4^{(i+1)} = \frac{1}{a_{44}} (b_4 - a_{41}x_1^{(i+1)} - a_{42}x_2^{(i+1)} - a_{43}x_3^{(i+1)} - \dots - a_{4n}x_n^{(i)}) \\ \vdots \\ x_n^{(i+1)} = \frac{1}{a_{nn}} (b_n - a_{n1}x_1^{(i+1)} - a_{n2}x_2^{(i+1)} - a_{n3}x_3^{(i+1)} - \dots - a_{n,n-1}x_{n-1}^{(i+1)}) \end{cases} \quad (2.7)$$

### Um Critério de Convergência (Gauss-Seidel)

Pode-se utilizar o Critério da Linhas ou Diagonal Dominante usado no método de Jacobi para o método de Gauss-Seidel. Mostraremos a seguir uma outra condição suficiente para a convergência do Método de Gauss-Seidel.

**Teorema 2.4.2 (Critério de Sassenfeld)** *Seja um sistema linear  $\mathbf{Ax} = \mathbf{b}$  e seja*

$$\beta_1 = \frac{|a_{12}| + |a_{13}| + |a_{14}| + \cdots + |a_{1n}|}{|a_{11}|}$$

$$\beta_2 = \frac{\beta_1 |a_{21}| + |a_{23}| + |a_{24}| + \cdots + |a_{2n}|}{|a_{22}|}$$

$$\beta_3 = \frac{\beta_1 |a_{31}| + \beta_2 |a_{32}| + |a_{34}| + \cdots + |a_{3n}|}{|a_{33}|}$$

$\vdots$

$$\beta_n = \frac{\beta_1 |a_{n1}| + \beta_2 |a_{n2}| + \beta_3 |a_{n3}| + \cdots + \beta_{n-1} |a_{n,n-1}|}{|a_{nn}|}$$

ou seja,

$$\beta_1 = \frac{|a_{12}| + |a_{13}| + |a_{14}| + \cdots + |a_{1n}|}{|a_{11}|}$$

e

$$\beta_i = \left( \sum_{j=1}^{i-1} \beta_j |a_{ij}| + \sum_{j=i+1}^n |a_{ij}| \right) / |a_{ii}| \quad i = 2, \dots, n$$

Se

$$\beta = \max_{1 \leq i \leq n} \beta_i < 1$$

então o método de Gauss-Seidel converge para a solução do sistema linear, independentemente da aproximação inicial.

**Exemplo:**

$$\begin{cases} 5x_1 + x_2 + x_3 = 5 \\ 3x_1 + 4x_2 + x_3 = 6 \\ 3x_1 + 3x_2 + 6x_3 = 0 \end{cases}$$

Pelo critério das linhas

$$\alpha_1 = \frac{|a_{12}| + |a_{13}|}{|a_{11}|} = \frac{1+1}{5} = \frac{2}{5} < 1$$

$$\alpha_2 = \frac{|a_{21}| + |a_{23}|}{|a_{22}|} = \frac{3+1}{4} = 1$$



$$\alpha_3 = \frac{|a_{31}| + |a_{32}|}{|a_{33}|} = \frac{3+3}{6} = 1$$

logo não temos certeza de que o método de Gauss-Seidel convergirá. Vamos testar o critério de Sassenfeld.

$$\beta_1 = \frac{|a_{12}| + |a_{13}|}{|a_{11}|} = \frac{1+1}{5} = \frac{2}{5} < 1$$

$$\beta_2 = \frac{\beta_1 |a_{21}| + |a_{23}|}{|a_{22}|} = \frac{\frac{2}{5} \cdot 3 + 1}{4} = \frac{11}{20} < 1$$

$$\beta_3 = \frac{\beta_1 |a_{31}| + \beta_2 |a_{32}|}{|a_{33}|} = \frac{\frac{2}{5} \cdot 3 + \frac{11}{20} \cdot 3}{6} = \frac{57}{120} < 1$$

que foi satisfeito.

Reescrevendo o sistema linear

$$\begin{cases} x_1^{(k+1)} &= \frac{1}{5} (5 - x_2^{(k)} - x_3^{(k)}) \\ x_2^{(k+1)} &= \frac{1}{4} (6 - 3x_1^{(k+1)} - x_3^{(k)}) \\ x_3^{(k+1)} &= \frac{1}{6} (0 - 3x_1^{(k+1)} - 3x_2^{(k+1)}) \end{cases}$$

com  $\mathbf{x}^{(0)} = \vec{0}$  e  $\epsilon = 0.05$ .

$$\begin{cases} x_1^{(1)} &= \frac{1}{5} (5 - x_2^{(0)} - x_3^{(0)}) &= \frac{1}{5} (5 - 0 - 0) = 1 \\ x_2^{(1)} &= \frac{1}{4} (6 - 3x_1^{(1)} - x_3^{(0)}) &= \frac{1}{4} (6 - 3 \cdot 1 - 0) = \frac{3}{4} \\ x_3^{(1)} &= \frac{1}{6} (0 - 3x_1^{(1)} - 3x_2^{(1)}) &= \frac{1}{6} (0 - 3 \cdot 1 - 3 \cdot \frac{3}{4}) = -\frac{21}{24} \end{cases}$$

$$\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{3}{4} \\ -\frac{21}{24} \end{bmatrix} = \begin{bmatrix} 1 \\ 0.75 \\ -0.875 \end{bmatrix}$$

Temos que verificar se o critério de parada foi satisfeito

$$\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| < \epsilon \text{ ?????}$$

$$\left\| \begin{bmatrix} 1 \\ 0.75 \\ -0.875 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 1 \\ 0.75 \\ -0.875 \end{bmatrix} \right\| = 1 > 0.05 = \epsilon$$

faremos então mais uma iteração

$$\begin{cases} x_1^{(2)} = \frac{1}{5} (5 - x_2^{(1)} - x_3^{(1)}) = \frac{1}{5} (5 - 1 + 0.875) = 1.025 \\ x_2^{(2)} = \frac{1}{4} (6 - 3x_1^{(2)} - x_3^{(1)}) = \frac{1}{4} (6 - 3 \cdot 1.025 + 0.875) = 0.95 \\ x_3^{(2)} = \frac{1}{6} (0 - 3x_1^{(2)} - 3x_2^{(2)}) = \frac{1}{6} (0 - 3 \cdot 1.025 - 3 \cdot 0.95) = -0.9875 \end{cases}$$

$$\begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{bmatrix} = \begin{bmatrix} 1.025 \\ 0.95 \\ -0.9875 \end{bmatrix}$$

Pelo critério de parada

$$\|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}\| < \epsilon \text{ ?????}$$

$$\left\| \begin{bmatrix} 1.025 \\ 0.95 \\ -0.9875 \end{bmatrix} - \begin{bmatrix} 1 \\ 0.75 \\ -0.875 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0.0250 \\ 0.2000 \\ -0.1125 \end{bmatrix} \right\| = 0.2000 > \epsilon$$

com mais uma iteração, obteremos,

$$\begin{bmatrix} x_1^{(3)} \\ x_2^{(3)} \\ x_3^{(3)} \end{bmatrix} = \begin{bmatrix} 1.0075 \\ 0.9912 \\ -0.9993 \end{bmatrix}$$

Pelo critério de parada

$$\|\mathbf{x}^{(3)} - \mathbf{x}^{(2)}\| < \epsilon \text{ ?????}$$

$$\left\| \begin{bmatrix} 1.0075 \\ 0.9912 \\ -0.9993 \end{bmatrix} - \begin{bmatrix} 1.025 \\ 0.95 \\ -0.9875 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0.0175 \\ 0.0412 \\ 0.0118 \end{bmatrix} \right\| = 0.0412 < \epsilon = 0.05$$



ou

$$\left\{ \begin{array}{l} u_1^{(i+1)} = u_1^{(i)} + \frac{\omega}{5} (-5u_1^{(i)} + 4u_2^{(i)} - u_3^{(i)}) \\ u_2^{(i+1)} = u_2^{(i)} + \frac{\omega}{6} (1 + 4u_1^{(i+1)} - 6u_2^{(i)} + 4u_3^{(i)} - u_4^{(i)}) \\ u_3^{(i+1)} = u_3^{(i)} + \frac{\omega}{6} (-u_1^{(i+1)} + 4u_2^{(i+1)} - 6u_3^{(i)} + 4u_4^{(i)}) \\ u_4^{(i+1)} = u_4^{(i)} + \frac{\omega}{5} (-u_2^{(i+1)} + 4u_3^{(i+1)} - 5u_4^{(i)}) \end{array} \right.$$

com uma estimativa inicial  $\mathbf{u}^{(0)} = \vec{0}$  e tolerância  $\epsilon = 0.001$ . Para o método de Gauss-Seidel ( $\omega = 1$ ) obteremos a convergência após 104 iterações e a aproximação para a solução é:

$$[\mathbf{u}^{(104)}]^T = (1.59; 2.59; 2.39; 1.39)$$

sendo que a solução exata é:

$$[\mathbf{u}]^T = (1.6; 2.6; 2.4; 1.4).$$

Variando-se o valor do parâmetro  $\omega$  temos a seguinte relação entre  $\omega$  e o número de iterações:

$\omega$	1,0	1,1	1,2	1,3	1,4	1,5	<u>1,6</u>	1,7	1,8	1,9
nº iterações	104	88	74	61	49	37	<u>23</u>	30	43	82

## Sistemas Lineares – Exercícios

1. Resolva os sistemas lineares abaixo por Eliminação de Gauss.

$$\text{a) } \begin{cases} 2x_1 + x_2 + 3x_3 = 7 \\ 4x_1 + 4x_2 + 3x_3 = 21 \\ 6x_1 + 7x_2 + 4x_3 = 32 \end{cases}$$

$$\text{b) } \begin{cases} 3x_1 + x_2 + 6x_3 = 2 \\ 2x_1 + x_2 + 3x_3 = 7 \\ x_1 + x_2 + x_3 = 4 \end{cases}$$

$$\text{c) } \begin{cases} 2x_1 + 2x_2 + x_3 + x_4 = 7 \\ x_1 - x_2 + 2x_3 - x_4 = 1 \\ 3x_1 + 2x_2 - 3x_3 - 2x_4 = 4 \\ 4x_1 + 3x_2 + 2x_3 + x_4 = 12 \end{cases}$$

$$\text{d) } \begin{cases} 4x_1 + 3x_2 + 2x_3 + x_4 = 10 \\ x_1 + 2x_2 + 3x_3 + 4x_4 = 5 \\ x_1 - x_2 - x_3 - x_4 = -1 \\ x_1 + x_2 + x_3 + x_4 = 3 \end{cases}$$

$$\text{e) } \begin{cases} 2x_1 + 3x_2 - x_3 = 5 \\ 4x_1 + 4x_2 - 3x_3 = 3 \\ 2x_1 - 3x_2 + 1x_3 = -1 \end{cases}$$

$$\text{Resp: } \mathbf{x} = [1 \ 2 \ 3]^T$$

**Solução:**

a)

$$\begin{cases} 2x_1 + x_2 + 3x_3 = 7 \\ 4x_1 + 4x_2 + 3x_3 = 21 \\ 6x_1 + 7x_2 + 4x_3 = 32 \end{cases}$$

$$\left( \begin{array}{ccc|c} 2 & 1 & 3 & 7 \\ 4 & 4 & 3 & 21 \\ 6 & 7 & 4 & 32 \end{array} \right) \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} - 4/2(\text{i}) \\ \text{(iii)} \leftarrow \text{(iii)} - 6/2(\text{i}) \end{array}$$

$$\left( \begin{array}{ccc|c} 2 & 1 & 3 & 7 \\ 0 & 2 & -3 & 7 \\ 0 & 4 & -5 & 11 \end{array} \right) \text{(iii)} \leftarrow \text{(iii)} - 4/2(\text{ii})$$

$$\left( \begin{array}{ccc|c} 2 & 1 & 3 & 7 \\ 0 & 2 & -3 & 7 \\ 0 & 0 & 1 & -3 \end{array} \right)$$

$$\text{Resp: } (17/2; -1; -3)$$

c)

$$\begin{cases} 2x_1 + 2x_2 + x_3 + x_4 = 7 \\ x_1 - x_2 + 2x_3 - x_4 = 1 \\ 3x_1 + 2x_2 - 3x_3 - 2x_4 = 4 \\ 4x_1 + 3x_2 + 2x_3 + x_4 = 12 \end{cases}$$

$$\left( \begin{array}{cccc|c} 2 & 2 & 1 & 1 & 7 \\ 1 & -1 & 2 & -1 & 1 \\ 3 & 2 & -3 & -2 & 4 \\ 4 & 3 & 2 & 1 & 12 \end{array} \right) \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} -1/2(\text{i}) \\ \text{(iii)} \leftarrow \text{(iii)} -3/2(\text{i}) \\ \text{(iv)} \leftarrow \text{(iv)} -2(\text{i}) \end{array}$$

$$\left( \begin{array}{cccc|c} 2 & 2 & 1 & 1 & 7 \\ 0 & -2 & 3/2 & -3/2 & -5/2 \\ 0 & -1 & -9/2 & -7/2 & -13/2 \\ 0 & -1 & 0 & -1 & -2 \end{array} \right) \begin{array}{l} \text{(iii)} \leftarrow \text{(iii)} -1/2(\text{ii}) \\ \text{(iv)} \leftarrow \text{(iv)} -1/2(\text{ii}) \end{array}$$

$$\left( \begin{array}{cccc|c} 2 & 2 & 1 & 1 & 7 \\ 0 & -2 & 3/2 & -3/2 & -5/2 \\ 0 & 0 & -21/42 & -11/4 & -21/4 \\ 0 & 0 & 0 & 1/7 & 0 \end{array} \right)$$

*Resp:* (1; 2; 1; 0)

2. Repita o exercício acima usando pivoteamento parcial.

**Solução:**

b) O elemento na diagonal da matriz é o maior da primeira coluna, logo não é necessário uma troca de linhas de escolha de pivô's.

$$\left( \begin{array}{ccc|c} 3 & 1 & 6 & 2 \\ 2 & 1 & 3 & 7 \\ 1 & 1 & 1 & 4 \end{array} \right) \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} -2/3(\text{i}) \\ \text{(iii)} \leftarrow \text{(iii)} -1/3(\text{i}) \end{array}$$

$$\left( \begin{array}{ccc|c} 3 & 1 & 6 & 2 \\ 0 & 1/3 & -1 & 17/3 \\ 0 & 2/3 & -1 & 10/3 \end{array} \right)$$

assim temos que trocar a segunda e terceiras linhas de

$$\left( \begin{array}{ccc|c} 3 & 1 & 6 & 2 \\ 0 & 2/3 & -1 & 10/3 \\ 0 & 1/3 & -1 & 17/3 \end{array} \right) \text{(iii)} \leftarrow \text{(iii)} -1/2(\text{ii})$$

$$\left( \begin{array}{ccc|c} 3 & 1 & 6 & 2 \\ 0 & 2/3 & -1 & 10/3 \\ 0 & 0 & -1/2 & 4 \end{array} \right)$$

com a retrossubstituição, obtemos

$$x_1 = 19$$

$$x_2 = -7$$

$$x_3 = -8$$

3. Ache a fatoração  $LU$  da matriz abaixo e calcule seu determinante

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 4 \\ 1 & -1 & 2 \\ 4 & 3 & 2 \end{pmatrix}$$

a seguir use esta fatoração para determinar a solução do sistema

$$\mathbf{Ax} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\begin{pmatrix} 3 & 2 & 4 \\ 1 & 1 & 2 \\ 4 & 3 & 2 \end{pmatrix} \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} - 1/3\text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} - 4/3\text{(i)} \end{array}$$

$$\begin{pmatrix} 3 & 2 & 4 \\ 0 & 1/3 & 2/3 \\ 0 & 1/3 & -10/3 \end{pmatrix} \text{(iii)} \leftarrow \text{(iii)} - \text{(ii)}$$

$$\begin{pmatrix} 3 & 2 & 4 \\ 0 & 1/3 & 2/3 \\ 0 & 0 & -12/3 \end{pmatrix}$$

Logo

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 4/3 & 1 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{U} = \begin{bmatrix} 3 & 2 & 4 \\ 0 & 1/3 & 2/3 \\ 0 & 0 & -4 \end{bmatrix}$$

Para resolver o sistema:

$$\mathbf{LUx} = \mathbf{b}$$

Chamando  $\mathbf{Ux} = \mathbf{y}$ , temos

$$\mathbf{Ly} = \mathbf{b}$$

ou para  $\mathbf{b} = (1, 2, 3)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 4/3 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

que resulta em

$$\begin{aligned} y_1 &= 4 \\ y_2 &= 5/3 \\ y_3 &= 0 \end{aligned}$$

Resolvendo agora o sistema triangular superior temos

$$\mathbf{U}\mathbf{x} = \mathbf{y}$$

ou

$$\begin{bmatrix} 3 & 2 & 4 \\ 0 & 1/3 & 2/3 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5/3 \\ 0 \end{bmatrix}$$

com a retrossubstituição, obtemos

$$\begin{aligned} x_1 &= -3 \\ x_2 &= 5 \\ x_3 &= 0 \end{aligned}$$

4. Dada a matriz  $\mathbf{A}$  abaixo

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & 6 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{pmatrix}$$

Ache a fatoração  $\mathbf{LR}$  de  $\mathbf{A}$  usando pivoteamento parcial. A seguir, com a fatoração encontrada resolva o sistema  $\mathbf{Ax} = \mathbf{b}$  onde

$$\mathbf{b} = \begin{bmatrix} 2 \\ 7 \\ 4 \end{bmatrix}$$

**Solução:**

O elemento na diagonal da matriz é o maior da primeira coluna, logo não é necessário uma troca de linhas de escolha de pivô's.

$$\begin{pmatrix} 3 & 1 & 6 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} - 2/3\text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} - 1/3\text{(i)} \end{array}$$



logo,

$$\mathbf{A}' = \begin{pmatrix} 3 & 1 & 6 \\ 0 & 1/3 & -1 \\ 0 & 2/3 & -1 \end{pmatrix} \quad \text{e} \quad \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 2/3 & 1 & 0 \\ 1/3 & ? & 1 \end{pmatrix} \quad \text{e} \quad \mathbf{R} = \begin{pmatrix} 3 & 1 & 6 \\ 0 & 1/3 & -1 \\ 0 & ? & ? \end{pmatrix}$$

assim temos que trocar a segunda e terceira linhas de

$$\mathbf{A}'' = \begin{pmatrix} 3 & 1 & 6 \\ 0 & 2/3 & -1 \\ 0 & 1/3 & -1 \end{pmatrix} \quad \text{e} \quad \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 2/3 & ? & 1 \end{pmatrix} \quad \text{e} \quad \mathbf{R} = \begin{pmatrix} 3 & 1 & 6 \\ 0 & 2/3 & -1 \\ 0 & ? & ? \end{pmatrix}$$

$$\begin{pmatrix} 3 & 1 & 6 \\ 0 & 2/3 & -1 \\ 0 & 1/3 & -1 \end{pmatrix} \quad (\text{iii}) \leftarrow (\text{iii}) - 1/2(\text{ii})$$

então

$$\mathbf{A}'' = \begin{pmatrix} 3 & 1 & 6 \\ 0 & 2/3 & -1 \\ 0 & 0 & -1/2 \end{pmatrix} \quad \text{e} \quad \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 2/3 & 1/2 & 1 \end{pmatrix} \quad \text{e} \quad \mathbf{R} = \begin{pmatrix} 3 & 1 & 6 \\ 0 & 2/3 & -1 \\ 0 & 0 & -1/2 \end{pmatrix}$$

Para utilizar a fatoração acima na resolução do sistema, precisamos considerar as trocas de linhas efetuadas na matriz e realizá-las também no segundo membro. Como trocamos a segunda e terceira linhas da matriz, devemos trocar a segunda e terceira linhas do segundo membro. Logo

$$\mathbf{b}' = \begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix}$$

$$\mathbf{LRx} = \mathbf{b}'$$

Chamando  $\mathbf{Rx} = \mathbf{y}$ , temos

$$\mathbf{Ly} = \mathbf{b}'$$

ou

$$\begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 2/3 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix}$$

que resulta em

$$y_1 = 2$$

$$y_2 = 10/3$$

$$y_3 = 4$$

Resolvendo agora o sistema triangular superior temos

$$\mathbf{R}\mathbf{x} = \mathbf{y}$$

ou

$$\begin{bmatrix} 3 & 1 & 6 \\ 0 & 2/3 & -1 \\ 0 & 0 & -1/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 10/3 \\ 4 \end{bmatrix}$$

com a retrossubstituição, obtemos

$$x_1 = 19$$

$$x_2 = -7$$

$$x_3 = -8$$

5. Suponha que desejamos resolver dois sistemas lineares que são idênticos exceto pelos segundos membros. Entre a Eliminação de Gauss e a Fatoração  $\mathbf{LU}$  qual método você escolheria? Por quê?

Usando o método escolhido por você resolva os dois sistemas abaixo.

$$\begin{cases} 2x_1 + x_2 + x_3 = 4 \\ 4x_1 + 4x_2 + 3x_3 = 11 \\ 6x_1 + 7x_2 + 4x_3 = 17 \end{cases}$$

e

$$\begin{cases} 2x_1 + x_2 + x_3 = 1 \\ 4x_1 + 4x_2 + 3x_3 = 1 \\ 6x_1 + 7x_2 + 4x_3 = 1 \end{cases}$$

### Solução:

Claramente escolheria a Fatoração  $\mathbf{LU}$ , visto que podemos resolver facilmente qualquer sistema no qual conhecemos sua fatoração.

Com o processo de eliminação de Gauss, temos

$$\begin{pmatrix} 2 & 1 & 1 \\ 4 & 4 & 3 \\ 6 & 7 & 4 \end{pmatrix} \begin{array}{l} \text{(i)} \\ \text{(ii)} \leftarrow \text{(i)} - 2\text{(i)} \\ \text{(iii)} \leftarrow \text{(i)} - 3\text{(i)} \end{array}$$

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 4 & 1 \end{pmatrix} \begin{array}{l} \text{(i)} \\ \text{(ii)} \\ \text{(iii)} \leftarrow \text{(ii)} - 2\text{(ii)} \end{array}$$

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

Logo

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{U} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

Para resolver os sistemas .....

$$\mathbf{LUx} = \mathbf{b}$$

Chamando  $\mathbf{Ux} = \mathbf{y}$ , temos

$$\mathbf{Ly} = \mathbf{b}$$

ou para  $\mathbf{b} = (4, 11, 17)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \\ 17 \end{bmatrix}$$

que resulta em

$$\begin{aligned} y_1 &= 4 \\ y_2 &= 3 \\ y_3 &= -1 \end{aligned}$$

Resolvendo agora o sistema triangular superior temos

$$\mathbf{Ux} = \mathbf{y}$$

ou

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ -1 \end{bmatrix}$$

com a retrossubstituição, obtemos

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 1 \\ x_3 &= 1 \end{aligned}$$

e para  $\mathbf{b} = (1, 1, 1)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

que resulta em

$$\begin{aligned} y_1 &= 1 \\ y_2 &= -1 \\ y_3 &= 0 \end{aligned}$$

Resolvendo agora o sistema triangular superior temos

$$\mathbf{U}\mathbf{x} = \mathbf{y}$$

ou

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

com a retrossubstituição, obtemos

$$\begin{aligned} x_1 &= 3/4 \\ x_2 &= -1/2 \\ x_3 &= 0 \end{aligned}$$

6. Seja a estrutura de treliças abaixo

Calcular os esforços nas barras da estrutura sabendo que cada uma possui comprimento  $L = 1 \text{ m}$ ,  $F = 1000 \text{ N}$ . Use o método da Eliminação de Gauss para resolver o sistema linear resultante.

7. Uma maneira de se calcular a inversa de uma matriz  $\mathbf{A}$  ( $n \times n$ ) é resolver os  $n$  sistemas lineares abaixo,

$$\mathbf{A}x = e_i \quad i = 1, 2, \dots, n$$

onde  $e_i$  são os vetores da base canônica, isto é,  $e_1 = (1, 0, 0, \dots, 0)$ ,  $e_2 = (0, 1, 0, \dots, 0)$ ,  $e_3 = (0, 0, 1, \dots, 0)$ ,  $\dots$ ,  $e_n = (0, 0, \dots, 0, 1)$ . As colunas de  $\mathbf{A}^{-1}$  são os vetores soluções desses sistemas. Entre a Eliminação de Gauss e a Fatoração  $\mathbf{LU}$  qual método você escolheria? Por quê?

Usando o método escolhido por você ache a inversa da matriz abaixo.

$$\mathbf{A} = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix}$$

### Solução:

Obviamente escolheríamos a fatoração  $\mathbf{LR}$  da matriz  $\mathbf{A}$  visto que temos que resolver vários sistemas lineares em a matriz dos coeficientes é fixa, só havendo modificações nos segundos membros.

Fatoração  $\mathbf{LR}$  de  $\mathbf{A}$

$$\begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix} \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} \quad -(-1)/(4)\text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} \quad -(0)/(4)\text{(i)} \end{array}$$

Obtemos

$$\begin{pmatrix} 4 & -1 & 0 \\ 0 & 15/4 & -1 \\ 0 & -1 & 4 \end{pmatrix}$$

Logo,

$$\Rightarrow \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ -1/4 & 1 & 0 \\ 0 & * & 1 \end{pmatrix} \quad \text{e} \quad \mathbf{R} = \begin{pmatrix} 4 & -1 & 0 \\ 0 & 15/4 & -1 \\ 0 & 0 & * \end{pmatrix}$$

Continuando a Fatoração

$$\begin{pmatrix} 4 & -1 & 0 \\ 0 & 15/4 & -1 \\ 0 & -1 & 4 \end{pmatrix} \text{(iii)} \leftarrow \text{(iii)} \quad -(-1)/(15/4)\text{(ii)}$$

$$\begin{pmatrix} 4 & -1 & 0 \\ 0 & 15/4 & -1 \\ 0 & 0 & 56/15 \end{pmatrix}$$

Logo,

$$\Rightarrow \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ -1/4 & 1 & 0 \\ 0 & -4/15 & 1 \end{pmatrix} \quad \text{e} \quad \mathbf{R} = \begin{pmatrix} 4 & -1 & 0 \\ 0 & 15/4 & -1 \\ 0 & 0 & 56/15 \end{pmatrix}$$

Usaremos agora a fatoração  $\mathbf{LR}$  para resolver os sistemas lineares.

Em todos os casos resolveremos os seguintes sistemas

$$\mathbf{Ax} = \mathbf{b}$$

$$\mathbf{LRx} = \mathbf{b}$$

$$\mathbf{L(Rx)} = \mathbf{b}$$

$$\mathbf{L(y)} = \mathbf{b}$$

resolvendo-se este sistema triangular inferior obtemos o vetor  $\mathbf{y}$  e a seguir podemos resolver o seguinte sistema linear triangular superior

$$\mathbf{Rx} = \mathbf{y}$$

Para o primeiro sistema

$$\mathbf{Ax} = \mathbf{e}_1$$

onde

$$\mathbf{e}_1 = (1, 0, 0)$$

temos:

$$\begin{pmatrix} 1 & 0 & 0 \\ -1/4 & 1 & 0 \\ 0 & -4/15 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/4 \\ 1/15 \end{pmatrix}$$

e

$$\begin{pmatrix} 4 & -1 & 0 \\ 0 & 15/4 & -1 \\ 0 & 0 & 56/15 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/4 \\ 1/15 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 15/56 \\ 1/14 \\ 1/56 \end{pmatrix}$$

Para o segundo sistema

$$\mathbf{Ax} = \mathbf{e}_2$$

onde

$$\mathbf{e}_2 = (0, 1, 0)$$

temos:

$$\begin{pmatrix} 1 & 0 & 0 \\ -1/4 & 1 & 0 \\ 0 & -4/15 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \implies \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 4/15 \end{pmatrix}$$

e

$$\begin{pmatrix} 4 & -1 & 0 \\ 0 & 15/4 & -1 \\ 0 & 0 & 56/15 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 4/15 \end{pmatrix} \implies \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1/14 \\ 2/7 \\ 1/14 \end{pmatrix}$$

Para o terceiro sistema

$$\mathbf{Ax} = \mathbf{e}_3$$

onde

$$\mathbf{e}_3 = (0, 0, 1)$$

temos:

$$\begin{pmatrix} 1 & 0 & 0 \\ -1/4 & 1 & 0 \\ 0 & -4/15 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \implies \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

e

$$\begin{pmatrix} 4 & -1 & 0 \\ 0 & 15/4 & -1 \\ 0 & 0 & 56/15 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \implies \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1/56 \\ 1/14 \\ 15/56 \end{pmatrix}$$

Finalmente temos

$$\mathbf{A}^{-1} = \begin{pmatrix} 15/56 & 1/14 & 1/56 \\ 1/14 & 2/7 & 1/14 \\ 1/56 & 1/14 & 15/56 \end{pmatrix}$$

8. Resolva os sistemas

$$\mathbf{Ax} = \mathbf{y}_1$$

$$\mathbf{Ax} = \mathbf{y}_2$$

usando a fatoração  $\mathbf{LU}$  de  $\mathbf{A}$ , onde:

$$\mathbf{A} = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix};$$

$$\mathbf{y}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix};$$

$$\mathbf{y}_2 = \begin{pmatrix} 12 \\ 15 \\ 20 \end{pmatrix};$$

9. Calcule o determinante da matriz de coeficientes do sistema linear abaixo e o resolva por Eliminação de Gauss.

$$\begin{cases} -x_1 + 2x_2 + 3x_3 + x_4 = 5 \\ 2x_1 - 4x_2 - 5x_3 - x_4 = -8 \\ -3x_1 + 8x_2 + 8x_3 + x_4 = 14 \\ x_1 + 2x_2 - 6x_3 + 4x_4 = 1 \end{cases}$$

**Solução:**

$$\left( \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 5 \\ 2 & -4 & -5 & -1 & -8 \\ -3 & 8 & 8 & 1 & 14 \\ 4 & 2 & -6 & 4 & 1 \end{array} \right) \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} - (2)/(-1)\text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} - (-3)/(-1)\text{(i)} \\ \text{(iv)} \leftarrow \text{(iv)} - (1)/(-1)\text{(i)} \end{array}$$

$$\left( \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 5 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 2 & -1 & -2 & -1 \\ 0 & 4 & -3 & 5 & 6 \end{array} \right) \begin{array}{l} \text{Trocando a (ii)} \\ \text{e (iii) linhas} \end{array}$$

$$\left( \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 5 \\ 0 & 2 & -1 & -2 & -1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 4 & -3 & 5 & 6 \end{array} \right) \begin{array}{l} \text{(iii)} \leftarrow \text{(iii)} - 0/2\text{(ii)} \\ \text{(iv)} \leftarrow \text{(iv)} - 4/2\text{(ii)} \end{array}$$



$$\left( \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 5 \\ 0 & 2 & -1 & -2 & -1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 9 & 8 \end{array} \right) \quad (\text{iv}) \leftarrow (\text{iv}) - (-1)/1(\text{iii})$$

$$\left( \begin{array}{cccc|c} -1 & 2 & 3 & 1 & 5 \\ 0 & 2 & -1 & -2 & -1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 10 & 10 \end{array} \right)$$

Com a retrossubstituição obtemos:

$$x_1 = 1; \quad x_2 = 1; \quad x_3 = 1; \quad x_4 = 1$$

O determinante da matriz será então o produto dos elementos da diagonal multiplicado por  $(-1)$  elevado ao número de trocas de linhas efetuadas, portantoo:

$$\det \mathbf{A} = (-1)^1 \times (-1) \times (2) \times (1) \times (10) = 20$$

10. Considere o sistema linear:

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 19 \\ 19 \\ -3 \\ -12 \end{bmatrix}$$

- Calcule a solução por Eliminação de Gauss;
- Calcule os vetores  $\mathbf{u}^{(k)}$ ,  $k \leq 10$ , obtidos pelos Métodos Iterativos de Jacobi e Gauss-Seidel, e relaxação para  $\omega = 1, 1; 1, 2; 1, 3; \dots; 1, 9$ . Use como estimativa inicial o vetor nulo.

11. Considere o sistema linear:

$$\begin{bmatrix} 3 & 2 & 0 & 1 \\ 9 & 8 & -3 & 4 \\ -6 & 4 & -8 & 0 \\ 3 & -8 & 3 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \\ -16 \\ 18 \end{bmatrix}$$

Resolva este sistema utilizando a fatoração  $\mathbf{LU}$  com pivoteamento parcial.

**Solução:**

Vamos calcular a fatoração  $\mathbf{PA} = \mathbf{LU}$ . As trocas de linhas efetuadas na matriz serão armazenadas em um vetor  $\mathbf{t}$  ao invés de utilizarmos matrizes de permutação. Inicialmente, este vetor terá a seguinte configuração:

$$\mathbf{t} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

a cada troca de linhas permutaremos as linhas de  $\mathbf{t}$  de forma idêntica.

$$\begin{bmatrix} 3 & 2 & 0 & 1 \\ 9 & 8 & -3 & 4 \\ -6 & 4 & -8 & 0 \\ 3 & -8 & 3 & -4 \end{bmatrix}$$

Com a estratégia de pivoteamento parcial temos que troca a primeira e segunda linhas da matriz, logo

$$\begin{bmatrix} 9 & 8 & -3 & 4 \\ 3 & 2 & 0 & 1 \\ -6 & 4 & -8 & 0 \\ 3 & -8 & 3 & -4 \end{bmatrix}$$

e o vetor  $\mathbf{t}$  de trocas passará a

$$\mathbf{t} = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 4 \end{bmatrix}$$

Podemos agora fatorar a matriz

$$\begin{bmatrix} 9 & 8 & -3 & 4 \\ 3 & 2 & 0 & 1 \\ -6 & 4 & -8 & 0 \\ 3 & -8 & 3 & -4 \end{bmatrix} \begin{array}{l} (ii) \leftarrow (ii) - (3)/(9) \quad (i) \rightarrow l_{21} = 3/9 \\ (iii) \leftarrow (iii) - (-6)/(9) \quad (i) \rightarrow l_{31} = -6/9 \\ (iv) \leftarrow (iv) - (3)/(9) \quad (i) \rightarrow l_{41} = 3/9 \end{array}$$

que resulta em

$$\begin{bmatrix} 9 & 8 & -3 & 4 \\ 0 & -6/9 & 1 & -3/9 \\ 0 & 84/9 & -10 & 24/9 \\ 0 & -96/9 & 4 & -48/9 \end{bmatrix}$$

Novamente devemos permutar a segunda e a quarta linha de forma que o pivo seja o elemento com maior valor absoluto na coluna, logo

$$\begin{bmatrix} 9 & 8 & -3 & 4 \\ 0 & -96/9 & 4 & -48/9 \\ 0 & 84/9 & -10 & 24/9 \\ 0 & -6/9 & 1 & -3/9 \end{bmatrix}$$

e o vetor  $\mathbf{t}$  de trocas passará a

$$\mathbf{t} = \begin{bmatrix} 2 \\ 4 \\ 3 \\ 1 \end{bmatrix}$$

Fatorando, temos

$$\begin{bmatrix} 9 & 8 & -3 & 4 \\ 0 & -96/9 & 4 & -48/9 \\ 0 & 84/9 & -10 & 24/9 \\ 0 & -6/9 & 1 & -3/9 \end{bmatrix} \begin{array}{l} (iii) \leftarrow (iii) - (84/9)/(-96/9) (ii) \rightarrow l_{32} = -7/8 \\ (iv) \leftarrow (iv) - (-6/9)/(-96/9) (ii) \rightarrow l_{42} = 1/16 \end{array}$$

Assim

$$\begin{bmatrix} 9 & 8 & -3 & 4 \\ 0 & -96/9 & 4 & -48/9 \\ 0 & 0 & -13/2 & -2 \\ 0 & 0 & 3/4 & 0 \end{bmatrix}$$

Na próxima etapa da fatoração não há necessidade de trocas de linhas logo

$$\begin{bmatrix} 9 & 8 & -3 & 4 \\ 0 & -96/9 & 4 & -48/9 \\ 0 & 0 & -13/2 & -2 \\ 0 & 0 & 3/4 & 0 \end{bmatrix} (iv) \leftarrow (iv) - (3/4)/(-13/2) (iii) \rightarrow l_{43} = -3/26$$

Temos enfim

$$\mathbf{U} = \begin{bmatrix} 9 & 8 & -3 & 4 \\ 0 & -32/3 & 4 & -16/3 \\ 0 & 0 & -13/2 & -2 \\ 0 & 0 & 0 & -3/13 \end{bmatrix} \quad \text{e} \quad \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/3 & 1 & 0 & 0 \\ -2/3 & -7/8 & 1 & 0 \\ 1/3 & 1/16 & -3/26 & 1 \end{bmatrix}$$

Como a forma final do vetor de trocas é:

$$\mathbf{t} = \begin{bmatrix} 2 \\ 4 \\ 3 \\ 1 \end{bmatrix}$$

antes de efetuarmos a resolução do sistema devemos realizar as trocas nas linhas do segundo membro, conseqüentemente

$$\mathbf{b} = \begin{bmatrix} 3 \\ 6 \\ -16 \\ 18 \end{bmatrix} \quad \longrightarrow \quad \mathbf{b}' = \mathbf{Pb} = \begin{bmatrix} 6 \\ 18 \\ -16 \\ 3 \end{bmatrix}$$

Resolvendo o sistema triangular inferior obtemos,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/3 & 1 & 0 & 0 \\ -2/3 & -7/8 & 1 & 0 \\ 1/3 & 1/16 & -3/26 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 6 \\ 18 \\ -16 \\ 3 \end{bmatrix}$$

ou

$$z = \begin{bmatrix} 6 \\ 16 \\ 2 \\ 3/13 \end{bmatrix}$$

Resolvendo agora o sistema triangular superior

$$U = \begin{bmatrix} 9 & 8 & -3 & 4 \\ 0 & -32/3 & 4 & -16/3 \\ 0 & 0 & -13/2 & -2 \\ 0 & 0 & 0 & -3/13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 6 \\ 16 \\ 2 \\ 3/13 \end{bmatrix}$$

que fornece

$$x = \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix}$$

12. a) Usando o critério de Sassenfeld, verifique para quais valores positivos de  $k$  se tem garantia de que o método de Gauss-Seidel vai gerar uma sequência convergente para a solução do sistema,

$$\begin{cases} kx_1 + 3x_2 + x_3 = 1 \\ kx_1 + 6x_2 + x_3 = 2 \\ x_1 + 6x_2 + 7x_3 = 3 \end{cases}$$

- b) Escolha o menor inteiro positivo maior que  $k$  e faça duas iterações do método de Gauss-Seidel.

**Solução:**

a) Pelo critério de Sassenfeld temos que para

$$\begin{aligned}\beta_1 &= \frac{|a_{12}| + |a_{13}| + |a_{14}| + \dots + |a_{1n}|}{|a_{11}|} \\ \beta_2 &= \frac{\beta_1 |a_{21}| + |a_{23}| + |a_{24}| + \dots + |a_{2n}|}{|a_{22}|} \\ \beta_3 &= \frac{\beta_1 |a_{31}| + \beta_2 |a_{32}| + |a_{34}| + \dots + |a_{3n}|}{|a_{33}|} \\ \beta_4 &= \frac{\beta_1 |a_{41}| + \beta_2 |a_{42}| + \beta_3 |a_{43}| + |a_{45}| + \dots + |a_{4n}|}{|a_{44}|} \\ &\vdots \\ \beta_n &= \frac{\beta_1 |a_{n1}| + \beta_2 |a_{n2}| + \beta_3 |a_{n3}| + \dots + \beta_{n-1} |a_{n-1,n}|}{|a_{nn}|}\end{aligned}$$

Se

$$\beta = \max_{1 \leq i \leq n} \beta_i < 1$$

então o Método de Gauss-Seidel converge para a solução do sistema.

Logo, nesse exercício temos:

$$\beta_1 = \frac{3+1}{k} = \frac{4}{k} < 1 \implies k > 4$$

$$\beta_2 = \frac{k\frac{4}{k} + 1}{6} = \frac{5}{6} < 1 \implies k > 4$$

$$\beta_3 = \frac{1\frac{4}{k} + 6\frac{5}{6}}{7} = \frac{4+5k}{7k} < 1 \implies k > 2$$

Logo  $k > 4$ .

b) Usando  $k = 5$  teremos:

$$\begin{cases} 5x_1 + 3x_2 + x_3 = 1 \\ 5x_1 + 6x_2 + x_3 = 2 \\ x_1 + 6x_2 + 7x_3 = 3 \end{cases}$$

Método de Gauss-Seidel, temos:

$$\begin{cases} x_1^{(i+1)} = \frac{1}{5} (1 - 3x_2^{(i)} - x_3^{(i)}) \\ x_2^{(i+1)} = \frac{1}{6} (2 - 5x_1^{(i+1)} - x_3^{(i)}) \\ x_3^{(i+1)} = \frac{1}{7} (3 - x_1^{(i+1)} - 6x_2^{(i+1)}) \end{cases}$$

Efetuada as iterações temos:

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{x}^{(1)} = \begin{bmatrix} 0,200000000 \\ 0,166666667 \\ 0,257142858 \end{bmatrix} \quad \dots\dots\dots$$

13. Compare as soluções dos sistemas lineares

a)

$$\begin{cases} x_1 - x_2 = 1 \\ x_1 - 1,00001x_2 = 0 \end{cases}$$

b)

$$\begin{cases} x_1 - x_2 = 1 \\ x_1 - 0,99999x_2 = 0 \end{cases}$$

O que aconteceu e por quê?

**Solução:**

a)

$$\left( \begin{array}{cc|c} 1 & -1 & 1 \\ 1 & -1,00001 & 0 \end{array} \right) \text{ (ii) } \leftarrow \text{ (ii) } -1/1(\text{i})$$

$$\left( \begin{array}{cc|c} 1 & -1 & 1 \\ 0 & -0,00001 & -1 \end{array} \right)$$

$$x_2 = 100000$$

$$x_1 = 100001$$

b)

$$\left( \begin{array}{cc|c} 1 & -1 & 1 \\ 1 & -0,99999 & 0 \end{array} \right) \text{ (ii) } \leftarrow \text{ (ii) } -1/1(\text{i})$$

$$\left( \begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 0,00001 & -1 \end{array} \right)$$

$$x_2 = -100000$$

$$x_1 = -99999$$

As soluções são bastantes diferentes, embora os coeficientes do sistema sejam quase idênticos. Esta instabilidade é fácil de interpretar. Cada sistema linear pode ser visto como a determinação da interseção entre duas retas quase paralelas. Naturalmente uma pequena variação de uma das retas pode mudar significativamente a interseção entre elas.

14. Reorganize o sistema

$$\begin{cases} x_1 + 10x_2 = 2 \\ x_2 + 10x_4 = 10 \\ 10x_1 - x_3 = 10 \\ 2x_3 - x_4 = 19 \end{cases}$$

de forma que sempre convirja o método de Jacobi.

Faça duas iterações a partir de  $x_1 = x_2 = x_3 = x_4 = 0$ . Use como tolerância  $\epsilon = 0,01$ .

**Solução:**

Para termos certeza que o Método de Jacobi converge, temos que reorganizar o sistema de forma que sua matriz se torne diagonal dominante (se for possível).

$$\begin{bmatrix} 1 & 10 & 0 & 0 \\ 0 & 1 & 0 & 10 \\ 10 & 0 & -1 & 0 \\ 0 & 0 & 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 10 \\ 10 \\ 19 \end{bmatrix}$$

Trocando linhas do sistema, podemos obter

$$\begin{bmatrix} 10 & 0 & -1 & 0 \\ 1 & 10 & 0 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 1 & 0 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 19 \\ 10 \end{bmatrix}$$

Que é diagonal dominante, logo, certamente, o Método de Jacobi convergirá:

Método de Jacobi:

$$\begin{cases} x_1^{(i+1)} = \frac{1}{10} (10 + x_3^{(i)}) \\ x_2^{(i+1)} = \frac{1}{10} (2 - x_1^{(i)}) \\ x_3^{(i+1)} = \frac{1}{2} (19 + x_4^{(i)}) \\ x_4^{(i+1)} = \frac{1}{10} (10 - x_2^{(i)}) \end{cases} \quad (2.9)$$

Efetuando as iterações temos:

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \mathbf{x}^{(1)} = \begin{bmatrix} 1,0 \\ 0,2 \\ 9,5 \\ 1,0 \end{bmatrix}; \quad \mathbf{x}^{(2)} = \begin{bmatrix} 1,95 \\ 0,10 \\ 10,00 \\ 0,90 \end{bmatrix}; \quad \dots\dots\dots$$

15. Repita o exercício acima para o Método de Gauss-Seidel.

Método de Gauss-Seidel:

$$\begin{cases} x_1^{(i+1)} = \frac{1}{10} (10 + x_3^{(i)}) \\ x_2^{(i+1)} = \frac{1}{10} (2 - x_1^{(i+1)}) \\ x_3^{(i+1)} = \frac{1}{2} (19 + x_4^{(i)}) \\ x_4^{(i+1)} = \frac{1}{10} (10 - x_1^{(i+1)}) \end{cases}$$

Efetutando as iterações temos:

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \mathbf{x}^{(1)} = \begin{bmatrix} 1 \\ 1/10 \\ 19/2 \\ 9/10 \end{bmatrix}; \quad \mathbf{x}^{(2)} = \begin{bmatrix} 39/20 \\ 1/200 \\ 199/20 \\ 161/200 \end{bmatrix}; \quad \dots\dots\dots$$

16. Verifique se a matriz do sistema abaixo satisfaz o critério de Sassenfeld. Faça três iterações do método de Gauss-Seidel.

$$\begin{bmatrix} 5 & 2 & 2 \\ 2 & 5 & 3 \\ 2 & 3 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 10 \\ 10 \end{bmatrix}$$

17. Dado o sistema linear abaixo:

$$\begin{bmatrix} 10 & 1 & 1 \\ 1 & 10 & 1 \\ 1 & 1 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 12 \\ 12 \end{bmatrix}$$

a) Verifique se o Critério de Sassenfeld é satisfeito.

b) Resolva por Gauss-Seidel se possível

**Solução:**

a)



$$\beta_1 = \frac{|a_{12}| + |a_{13}|}{|a_{11}|} = \frac{1+1}{10} = \frac{1}{5} < 1$$

$$\beta_2 = \frac{\beta_1 |a_{21}| + |a_{23}|}{|a_{22}|} = \frac{\frac{1}{5}1 + 1}{10} = \frac{6}{50} < 1$$

$$\beta_3 = \frac{\beta_1 |a_{31}| + \beta_2 |a_{32}|}{|a_{33}|} = \frac{\frac{1}{5}1 + \frac{6}{50}1}{10} = \frac{4}{125} < 1$$

Então, o critério é satisfeito

b)

Método de Gauss-Seidel:

$$\begin{cases} x_1^{(i+1)} = \frac{1}{10} \{12 - x_2^{(i)} - x_3^{(i)}\} \\ x_2^{(i+1)} = \frac{1}{10} \{12 - x_1^{(i+1)} - x_3^{(i)}\} \\ x_3^{(i+1)} = \frac{1}{10} \{12 - x_1^{(i+1)} - x_2^{(i+1)}\} \end{cases}$$

logo

$$\begin{cases} x_1^{(1)} = \frac{1}{10} \{12 - x_2^{(0)} - x_3^{(0)}\} = \frac{1}{10} \{12 - 0 - 0\} \implies x_1^{(1)} = 1,2 \\ x_2^{(1)} = \frac{1}{10} \{12 - x_1^{(1)} - x_3^{(0)}\} = \frac{1}{10} \{12 - 1,2 - 0\} \implies x_2^{(1)} = 1,08 \\ x_3^{(1)} = \frac{1}{10} \{12 - x_1^{(1)} - x_2^{(1)}\} = \frac{1}{10} \{12 - 1,2 - 1,08\} \implies x_3^{(1)} = 0,972 \end{cases}$$

$$\mathbf{x}^{(1)} = \begin{bmatrix} 1,2 \\ 1,08 \\ 0,972 \end{bmatrix}$$

$$\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|_{\infty} = \left\| \begin{bmatrix} 1,2 \\ 1,08 \\ 0,972 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\|_{\infty} = \left\| \begin{bmatrix} 1,2 \\ 1,08 \\ 0,972 \end{bmatrix} \right\|_{\infty} = 1,2 > \epsilon$$

$$\left\{ \begin{array}{l} x_1^{(2)} = \frac{1}{10} \{12 - x_2^{(1)} - x_3^{(1)}\} = \frac{1}{10} \{12 - 1,08 - 0,972\} \implies x_1^{(2)} = 0,9948 \\ x_2^{(2)} = \frac{1}{10} \{12 - x_1^{(2)} - x_3^{(1)}\} = \frac{1}{10} \{12 - 0,9948 - 0,972\} \implies x_2^{(2)} = 1,00 \\ x_3^{(2)} = \frac{1}{10} \{12 - x_1^{(2)} - x_2^{(2)}\} = \frac{1}{10} \{12 - 0,9948 - 1,00\} \implies x_3^{(2)} = 0,9925 \end{array} \right.$$

$$\mathbf{x}^{(2)} = \begin{bmatrix} 0,9948 \\ 1,00 \\ 0,9925 \end{bmatrix}$$

$$\|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}\|_{\infty} = \left\| \begin{bmatrix} 0,9948 \\ 1,00 \\ 0,9925 \end{bmatrix} - \begin{bmatrix} 1,2 \\ 1,08 \\ 0,972 \end{bmatrix} \right\|_{\infty} = \left\| \begin{bmatrix} -0,2 \\ -0,08 \\ 0,02 \end{bmatrix} \right\|_{\infty} = 0,2 > \epsilon$$

$$\left\{ \begin{array}{l} x_1^{(3)} = \frac{1}{10} \{12 - x_2^{(2)} - x_3^{(2)}\} = \frac{1}{10} \{12 - 1,00 - 0,9925\} \implies x_1^{(3)} = 1,00 \\ x_2^{(3)} = \frac{1}{10} \{12 - x_1^{(3)} - x_3^{(2)}\} = \frac{1}{10} \{12 - 1,00 - 0,9925\} \implies x_2^{(3)} = 1,00 \\ x_3^{(3)} = \frac{1}{10} \{12 - x_1^{(3)} - x_2^{(3)}\} = \frac{1}{10} \{12 - 1,00 - 1,00\} \implies x_3^{(3)} = 1,00 \end{array} \right.$$

$$\mathbf{x}^{(3)} = \begin{bmatrix} 1,00 \\ 1,00 \\ 1,00 \end{bmatrix}$$

$$\|\mathbf{x}^{(3)} - \mathbf{x}^{(2)}\|_{\infty} = \left\| \begin{bmatrix} 1,00 \\ 1,00 \\ 1,00 \end{bmatrix} - \begin{bmatrix} 0,9948 \\ 1,00 \\ 0,9925 \end{bmatrix} \right\|_{\infty} = \left\| \begin{bmatrix} 0,0052 \\ 0,00 \\ 0,0075 \end{bmatrix} \right\|_{\infty} = 0,0075 < \epsilon$$

18. Seja o diagrama de um circuito

A corrente entre dois nós  $p$  e  $q$  de uma rede elétrica é

$$I_{pq} = \frac{V_p - V_q}{R_{pq}}$$

onde  $V_p$  e  $V_q$  são respectivamente as voltagens em  $p$  e  $q$ , e  $R_{pq}$  é a resistência entre  $p$  e  $q$ . A soma das correntes que chegam a cada nó é nula.

Calcular as voltagens em cada nó do circuito.

19. Considere o seguinte sistema de equações lineares:

$$\begin{cases} 2x_1 + x_2 + 3x_3 = 1 \\ -x_2 + x_3 = 2 \\ x_1 + 3x_3 = 3 \end{cases}$$

- (a) Reescreva o sistema de modo que o critério de Sassenfeld seja satisfeito.  
 (b) A seguir ache a solução do sistema pelo método de Gauss-Seidel, use como tolerância  $\epsilon = 0.05$ . (Caso o número de iterações necessárias para a convergência seja maior que três, efetue somente três iterações)

**Solução:**

- (a) Se permutamos a primeira e terceira linhas e a primeira e terceira colunas, obtemos

$$\begin{cases} 3x_3 + x_1 = 3 \\ x_3 - x_2 = 2 \\ 3x_3 + x_2 + 2x_1 = 1 \end{cases}$$

ou

$$\begin{pmatrix} 3 & 0 & 1 \\ 1 & -1 & 0 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_3 \\ x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

Assim o critério de Sassenfeld torna-se satisfeito como visto a seguir:

$$\beta_1 = \frac{|a_{12}| + |a_{13}|}{|a_{11}|} = \frac{|0| + |1|}{|3|} = \frac{1}{3} < 1$$

$$\beta_2 = \frac{\beta_1 |a_{21}| + |a_{23}|}{|a_{22}|} = \frac{(1/3) |1| + |0|}{|-1|} = \frac{1}{3} < 1$$

$$\beta_3 = \frac{\beta_1 |a_{31}| + \beta_2 |a_{32}|}{|a_{33}|} = \frac{(1/3) |3| + (1/3) |1|}{|2|} = \frac{4}{6} = \frac{2}{3} < 1$$

- (b) Usando como aproximação inicial o vetor nulo  $\mathbf{x}^{(0)} = (x_3^{(0)}, x_2^{(0)}, x_1^{(0)}) = (0, 0, 0)$ , temos os seguintes resultados:

Iteração	$x_3$	$x_2$	$x_1$
1	1.0000000000000000	-1.0000000000000000	-0.5000000000000000
2	1.1666666666666667	-0.8333333333333333	-0.8333333333333333
3	1.2777777777777778	-0.7222222222222221	-1.0555555555555556
4	1.351851851851852	-0.6481481481481481	-1.203703703703704
5	1.401234567901235	-0.5987654320987652	-1.302469135802470
6	1.434156378600823	-0.5658436213991767	-1.368312757201647
7	1.456104252400549	-0.5438957475994510	-1.412208504801098

20. Resolva o sistema linear abaixo pelo método de Gauss-Seidel, continuando o processo iterativo até que a diferença máxima entre os valores sucessivos de  $x$ ,  $y$  e  $z$  seja menor que 0,02. Isto significa que a solução aproximada está a uma distância menor que 0,02 da solução exata?

$$\begin{cases} 10x + 2y + 6z = 28 \\ x + 10y + 9z = 7 \\ 2x - 7y - 10z = -17 \end{cases}$$

**Solução:**

Usaremos como estimativa inicial,  $x_0 = 0$ ,  $y_0 = 0$  e  $z_0 = 0$ .

Resultado das Aproximações

Iteração	$x$	$y$	$z$
1	2.8000000000000000	0.4200000000000000	1.9660000000000000
2	1.5364000000000000	-1.2230400000000000	2.8634080000000000
3	1.3265632000000000	-2.0097235200000000	3.3721191040000000
4	1.1786732416000000	-2.4527745177600000	3.652676810752001
5	1.098948817100800	-2.697304011386881	3.807902571390977
6	1.054719259442790	-2.832584240196158	3.893752820025868
7	1.030265156023711	-2.907404053625652	3.941235868742699
8	1.016739289479511	-2.948786210816380	3.967498205467368
9	1.009258318882855	-2.971674216808917	3.982023615542813
10	1.005120674036096	-2.984333321392141	3.990057459781718

21. Escreva um procedimento computacional, em pseudocódigo ou usando alguma linguagem de programação, para transformar uma matriz quadrada em uma matriz triangular superior usando o processo de Eliminação de Gauss (Isto é, não considere as operações sobre o segundo membro nem a retrossubstituição).
22. Mostre que o número de operações aritméticas necessárias para se triangularizar uma matriz quadrada pelo Processo de Eliminação de Gauss é aproximadamente  $n^3/3$  operações, sendo  $n$  a ordem da matriz.

**Solução:**

**Custo da Eliminação Gaussiana**

*Quantas operações aritméticas são necessárias para se resolver um sistema de  $n$  equações e  $n$  incógnitas?*

Vejamos em primeiro lugar somente as operações na matriz (ignoraremos operações no segundo membro).

Essas operações são de dois tipos.

Uma operação é a divisão pelo pivô, para encontrar o múltiplo da linha pivotal que deverá ser subtraído. E a seguir, efetuamos realmente essa subtração.

Consideremos cada divisão e cada multiplicação–subtração uma única operação.

No início, quando a primeira equação tem  $n$  elementos na matriz, são necessárias  $n$  operações para cada zero(0) obtido na primeira coluna de cada linha ( 1 operação para determinarmos o múltiplo (divisão) e  $n - 1$  para os outros elementos da linha fatorada (multiplicações e subtrações)). Existem  $n - 1$  linhas abaixo da primeira linha, assim a primeira etapa da eliminação necessita de  $n(n - 1) = n^2 - n$  operações (outra abordagem é a seguinte: todos os  $n^2$  elementos precisam ser modificados, menos os  $n$  elementos da primeira linha  $\rightarrow n^2 - n$ ). Agora, observe que as etapas seguintes são mais "rápidas", porque as equações se tornam progressivamente menores. Quando a eliminação é feita em  $k$  equações ( $k < n$ ), somente  $k^2 - k$  operações são necessárias para zerar a coluna abaixo do pivô (pela mesma razão usada na primeira linha).

Juntas, o número de operações na matriz é a soma de  $k^2 - k$  operações para  $k$  variando de 1 até  $n$ . Isto é,

$$\begin{aligned} \sum_{k=1}^n k^2 - k &= \sum_{k=1}^n k^2 - \sum_{k=1}^n k = \\ &= (1^2 + 2^2 + \dots + n^2) - (1 + 2 + \dots + n) = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2} = \frac{n^3 - n}{3} \approx \frac{n^3}{3} \end{aligned}$$

#### Observação:

Se o  $n$  é muito grande, uma estimativa do número de operações é  $\frac{1}{3}n^3$ .

No segundo membro na primeira etapa temos  $n - 1$  operações (uma multiplicação–subtração para cada elemento do segundo membro). Na segunda etapa como temos um sistema  $(n - 1) \times (n - 1)$ , serão necessárias  $n - 2$  operações, e assim por diante. Logo, somando todas as operações, teremos

$$\begin{aligned} (n - 1) + (n - 2) + \dots + 2 + 1 &= 1 + 2 + \dots + (n - 1) = \frac{(n - 1)((n - 1) + 1)}{2} = \frac{n(n - 1)}{2} = \\ &= \frac{n^2 - n}{2} \approx \frac{n^2}{2} \end{aligned}$$

#### Observação:

Se o  $n$  é muito grande, uma estimativa do número de operações no segundo membro é  $\frac{1}{2}n^2$ .

A retro-substituição é bem mais "rápida".

A última incógnita ( $x_n$ ) é determinada com somente uma operação (uma divisão). A penúltima incógnita requer 2 operações (uma subtração–multiplicação e uma divisão) e assim por diante. Logo o número total de operações para a retrosubstituição é

$$1 + 2 + \dots + n = \frac{n(n+1)}{2} = \frac{n^2 + n}{2} \approx \frac{n^2}{2}$$

Somando todas as operações efetuadas, temos:

$$\frac{n^3 - n}{3} + \frac{n^2 - n}{2} + \frac{n^2 + n}{2} = \frac{n^3 - n}{3} + n^2$$

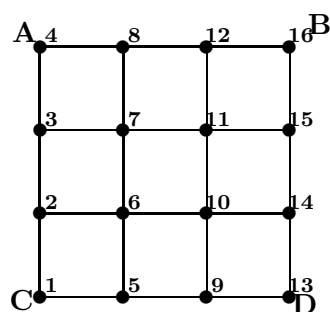
**Observação:**

Novamente, podemos dizer, se  $n$  for muito grande que o número de operações necessários para se resolver um sistema por Eliminação de Gauss é  $\frac{1}{3}n^3$ .

23. Seja a placa mostrada na figura abaixo, os lados AC e BD são mantidos à temperatura de  $20^\circ\text{C}$ , o lado AB, a  $40^\circ\text{C}$  e CD a  $10^\circ\text{C}$ . Para determinar as temperaturas  $T_i$  dos pontos interiores da placa, pode-se supor que a temperatura em cada ponto é igual à média aritmética dos quatros pontos contíguos. Por exemplo:

$$T_7 = \frac{T_3 + T_8 + T_{11} + T_6}{4}$$

As relações desse tipo permitem formar um sistema de 4 equações e 4 incógnitas, envolvendo as temperaturas dos pontos interiores da placa. Monte o sistema de equações acima citado e determine  $T_6$ ,  $T_7$ ,  $T_{10}$  e  $T_{11}$  usando o método de eliminação de Gauss.

**Solução:**

Para o ponto 7 temos:

$$T_7 = \frac{T_3 + T_8 + T_{11} + T_6}{4}$$

Para o ponto 11 temos:

$$T_{11} = \frac{T_7 + T_{12} + T_{15} + T_{10}}{4}$$

Para o ponto 6 temos:

$$T_6 = \frac{T_2 + T_7 + T_{10} + T_5}{4}$$

Para o ponto 10 temos:

$$T_{10} = \frac{T_6 + T_{11} + T_{14} + T_9}{4}$$

mas

$$T_2 = T_3 = T_{14} = T_{15} = 20$$

$$T_8 = T_{12} = 40$$

$$T_5 = T_9 = 10$$

Substituindo nas equações acima:

$$4T_7 = T_3 + T_8 + T_{11} + T_6 = 20 + 40 + T_{11} + T_6$$

$$4T_{11} = T_7 + T_{12} + T_{15} + T_{10} = T_7 + 40 + 20 + T_{10}$$

$$4T_6 = T_2 + T_7 + T_{10} + T_5 = 20 + T_7 + T_{10} + 10$$

$$4T_{10} = T_6 + T_{11} + T_{14} + T_9 = T_6 + T_{11} + 20 + 10$$

ou

$$\begin{cases} 4T_6 - T_7 - T_{10} = 30 \\ 4T_7 - T_{11} - T_6 = 60 \\ 4T_{10} - T_6 - T_{11} = 30 \\ 4T_{11} - T_7 - T_{10} = 60 \end{cases}$$

ou ainda

$$\begin{cases} 4T_6 - T_7 - T_{10} = 30 \\ -T_6 + 4T_7 - T_{11} = 60 \\ -T_6 + 4T_{10} - T_{11} = 30 \\ -T_7 - T_{10} + 4T_{11} = 60 \end{cases}$$

Como uma matriz aumentada

$$\left[ \begin{array}{cccc|c} 4 & -1 & -1 & 0 & 30 \\ -1 & 4 & 0 & -1 & 60 \\ -1 & 0 & 4 & -1 & 30 \\ 0 & -1 & -1 & 4 & 60 \end{array} \right]$$

$$\left[ \begin{array}{cccc|c} 4 & -1 & -1 & 0 & 30 \\ -1 & 4 & 0 & -1 & 60 \\ -1 & 0 & 4 & -1 & 30 \\ 0 & -1 & -1 & 4 & 60 \end{array} \right] \begin{array}{l} \text{(ii)} \leftarrow \text{(ii)} -(-1)/4(\text{i}) \\ \text{(iii)} \leftarrow \text{(iii)} -(-1)/4(\text{i}) \\ \text{(iv)} \leftarrow \text{(iv)} \end{array}$$

$$\left[ \begin{array}{cccc|c} 4 & -1 & -1 & 0 & 30 \\ 0 & 15/4 & -1/4 & -1 & 270/4 \\ 0 & -1/4 & 15/4 & -1 & 150/4 \\ 0 & -1 & -1 & 4 & 60 \end{array} \right] \begin{array}{l} \text{(iii)} \leftarrow \text{(iii)} -(-1/4)/(15/4)(\text{ii}) \\ \text{(iv)} \leftarrow \text{(iv)} -(-1)/(15/4)(\text{ii}) \end{array}$$

$$\left[ \begin{array}{cccc|c} 4 & -1 & -1 & 0 & 30 \\ 0 & 15/4 & -1/4 & -1 & 270/4 \\ 0 & 0 & 224/60 & -16/15 & 2520/60 \\ 0 & 0 & -16/15 & 56/15 & 1170/15 \end{array} \right] \quad (\text{iv}) \leftarrow (\text{iv}) - (-16/15)/(224/60)(\text{iii})$$

$$\left[ \begin{array}{cccc|c} 4 & -1 & -1 & 0 & 30 \\ 0 & 15/4 & -1/4 & -1 & 270/4 \\ 0 & 0 & 224/60 & -16/15 & 2520/60 \\ 0 & 0 & 0 & 24/7 & 90 \end{array} \right]$$

Com a retrosubstituição

24. Considere o seguinte sistema de equações lineares:

$$\begin{cases} x_1 + 10x_2 = 2 \\ x_2 + 10x_4 = 10 \\ 10x_1 - x_3 = 10 \\ 2x_3 - x_4 = 19 \end{cases}$$

- Determine a matriz ampliada do sistema.
- Resolva o sistema pelo Método de Gauss, sem pivoteamento, definindo todas as operações elementares utilizadas na triangularização.
- Determine a fatoração **LU** da matriz dos coeficientes do sistema.
- Determine o determinante da matriz dos coeficientes do sistema, usando a fatoração **LU** desta matriz.

25. Considere o seguinte sistema de equações lineares:

$$\begin{cases} x_1 + 10x_2 = 2 \\ x_2 + 10x_4 = 10 \\ 10x_1 - x_3 = 10 \\ 2x_3 - x_4 = 19 \end{cases}$$

- Coloque o sistema acima em uma forma tal que seja possível garantir a convergência dos métodos de Jacobi e de Gauss-Seidel.
- Escreva a expressão para o valor de  $x_j^{(i+1)}$  da  $j$ -ésima incógnita,  $1 \leq j \leq 4$ , numa iteração  $(i+1)$  obtida pelos métodos de Jacobi e de Gauss-Seidel.
- Quais os critérios de parada que você usaria na resolução desse sistema pelos métodos de Jacobi e de Gauss-Seidel?
- Faça duas iterações do Método de Jacobi, usando  $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = x_4^{(0)} = 0$  (trabalhe com três dígitos decimais).

26. Responda as seguintes questões, justificando integralmente suas respostas.



- (a) Se o Método da Eliminação de Gauss com pivoteamento parcial fosse usado para resolver um sistema linear, como seria calculado o determinante da matriz de coeficientes do sistema?
- (b) Diga quais as diferenças entre os métodos diretos e os métodos iterativos para a solução de sistemas lineares. Quando devemos usar os métodos iterativos?
- (c) Tanto nos métodos diretos quanto nos métodos iterativos para a resolução de sistemas lineares existem erros de arredondamento, ambos associados a precisão da máquina. Em quais métodos os erros de arredondamento são irrelevantes?
- (d) Para que serve o pivoteamento parcial ou total no processo de Eliminação de Gauss?
- (e) Dado um sistema linear  $\mathbf{Ax} = \mathbf{b}$ , se a matriz  $\mathbf{A}$  não satisfaz o critério de Sassenfeld, o que podemos dizer sobre a utilização do método de Gauss-Seidel para resolver esse sistema?
- (f) Se o determinante de uma matriz  $\mathbf{A}$  é diferente de zero podemos garantir que  $\mathbf{A}$  tem fatoração  $\mathbf{LU}$ ? Comente sua resposta.
- (g) Quando devemos usar um método iterativo ao invés de um método direto na resolução de um sistema linear?
- (h) No problema de resolver um sistema linear, podemos analisar o condicionamento da matriz de coeficientes do sistema. O que nos diz o estado desse condicionamento? O que é o número de condição da matriz do sistema ( $\text{Cond}(\mathbf{A})$  ou  $\kappa(\mathbf{A})$ ).
- (i) Quando podemos usar a fatoração de Cholesky para resolver um sistema linear?

27. Comente, confirmando ou contestando, as seguintes afirmações:

- (a) Custo computacional da eliminação gaussiana é proporcional a  $n^4$  para sistemas  $n \times n$ .
- (b) Método de Jacobi converge sempre (condição suficiente) que a diagonal for maior que 1 (em módulo) em todas as linhas.

# Chapter 3

## Zeros de Funções

### 3.1 Introdução

Equações não-lineares são aquelas que contém potências ou funções transcendentais da variável independente. Tais equações surgem freqüentemente em engenharia, física ou química. Nosso objetivo é encontrar os zeros de uma dada função  $f(x)$  que é achar os pontos  $x$  tais que  $f(x) = 0$ . Graficamente temos,

Por exemplo, determinar os zeros de um polinômio (ou raízes do polinômio),

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

é um problema clássico.

Exemplo:

$$x^3 - 4x^2 + 8x - 2 = 0$$

$$e^x - \cos x = 0$$

Obter os zeros de uma função qualquer analiticamente é muito difícil ou impossível. Por isso, devemos tentar encontrar apenas aproximações para esses zeros.

A idéia geral dos métodos numéricos que estudaremos é obter, a partir de uma aproximação inicial para um zero, uma seqüência de melhores aproximações através de um processo iterativo. Isto é,

$$x_0 \longrightarrow x_1 \longrightarrow x_2 \longrightarrow x_3 \longrightarrow \cdots \longrightarrow \bar{x}$$

ou

$$\lim_{i \rightarrow \infty} x_i = \bar{x}$$

Todos os métodos que apresentaremos são iterativos.

Um método iterativo consiste de uma seqüência de instruções que são repetidas em ciclos.

A execução de um ciclo chama-se iteração. Cada iteração utiliza resultados das iterações anteriores.

$$\underbrace{x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_4 \rightarrow \cdots}_{\text{aproximações}} \rightarrow \underbrace{\bar{x}}_{\text{zero}}$$

Surgem algumas questões:

1. Como encontrar uma aproximação inicial?
2. Como interromper a geração de novas aproximações?
3. Como gerar as novas aproximações?
4. Quão rapidamente a seqüência se aproxima do limite?

## Isolamento dos zeros - Aproximação inicial

Veremos mais adiante que diferentemente dos métodos iterativos para sistemas lineares a aproximação inicial pode ser importante para a convergência.

A melhor maneira de se isolar um determinado zero de uma função é construir seu gráfico.

Exemplo:

Entretanto construir o gráfico de uma função pode ser muito difícil, tanto quanto resolver um problema de zeros de funções. Nesse caso, podemos utilizar o seguinte resultado.

**Teorema 3.1.1** *Seja  $f(x)$  uma função contínua em um intervalo  $[a, b]$ . Se  $f(a) \cdot f(b) < 0$  então existe pelo menos um ponto entre  $a$  e  $b$  que é zero de  $f(x)$ .*

E então podemos construir uma tabela de pontos com os valores da função e verificar onde existam trocas de sinal da função.

Exemplo:

Seja a função  $f(x) = x^3 - 9x + 3$ , então:

$x$	-100	-10	-5	-3	-1	0	1	2	3	4
$f(x)$	-999097	-907	-77	3	11	3	-5	-7	3	11
sinal	-	-	-	+	+	+	-	-	+	+

Pode-se então afirmar que existem zeros nos seguintes intervalos:  $(-5, -3)$ ;  $(0, 1)$ ;  $(2, 3)$ .

Conhecendo-se onde se encontram aproximadamente os zeros da função parte-se, então, para a melhoria da aproximação do zero, para isto apresentaremos diferentes métodos numéricos para obtenção de zeros de funções. Observe que a técnica acima não nos dá uma estimativa inicial para as raízes complexas de uma função.

## Critério de Parada

Quando parar o processo iterativo?  $\rightarrow$  Critério de Parada.

Uma maneira de determinar se a seqüência gerada pelo método iterativo está convergindo é verificar se a diferença entre a aproximação e o valor exato do zero está diminuindo. Seja  $x_i$  o valor aproximado obtido na iteração  $i$  e  $\bar{x}$  o valor exato do zero, podemos verificar se

$$|x_i - \bar{x}| < \epsilon$$

isto é, se a diferença acima citada é menor que um valor *pequeno* ( $\epsilon$ ). Chamamos  $\epsilon$  de tolerância de parada ou simplesmente tolerância.

Entretanto, como não conhecemos o valor exato do zero da função (queremos na verdade determinar este valor), o teste acima não pode ser efetuado. Substituímos o teste acima pelo seguinte teste:

$$|x_{i+1} - x_i| < \epsilon$$

A diferença agora é calculada entre duas aproximações sucessivas.

O critério empregando a diferença relativa entre aproximações sucessivas também é usado, isto é,

$$\frac{|x_{i+1} - x_i|}{|x_{i+1}|} < \epsilon$$

Outro teste que também pode ser realizado é

$$|f(x_i)| < \epsilon$$

Porém, os dois testes nem sempre são satisfeitos simultaneamente.

## Métodos de Quebra

Nestes métodos a aproximação do zero é obtida partindo-se de um intervalo inicial que contenha um zero por divisões sucessivas desse intervalo, como se houvesse uma *quebra* do intervalo em subintervalos.

### 3.2 Método da Bissecção

Seja  $[a, b]$  um intervalo que contenha um zero de uma função contínua  $f(x)$ , isto é,  $f(a) \cdot f(b) < 0$ .

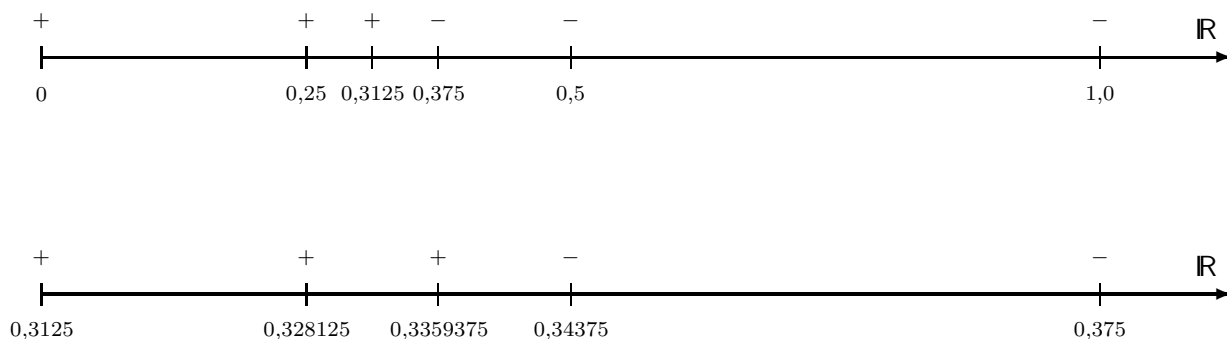
O objetivo do método é reduzir o tamanho do intervalo inicial que contém o zero até que seja atingida uma precisão desejada. A redução do intervalo é conseguida através da divisão sucessiva ao meio do intervalo. O ponto que divide o intervalo em dois subintervalos, é,

$$x = \frac{a + b}{2}$$

A divisão do intervalo, define dois subintervalos, sendo que um contém o zero e o outro não. Repetimos o mesmo processo sempre com o intervalo que contém o zero, como mostra a figura abaixo. O novo intervalo que contém o zero será aquele onde existe troca de sinal da função.

**Exemplo:**

$f(x) = x^3 - 9x + 3$  com  $I = [0, 1]$  e  $\epsilon = 0.01$



Iteração	$x = (a + b)/2$	$f(x)$	$b - a$	$ f(x) $
1	$x_1 = 0,5000000$	$-1,375000000$	$0,5000000$	$1,375000000$
2	$x_2 = 0,2500000$	$0,765625000$	$0,2500000$	$0,765625000$
3	$x_3 = 0,3750000$	$-0,322265250$	$0,1250000$	$0,322265250$
4	$x_4 = 0,3125000$	$0,218017578$	$0,0625000$	$0,218017578$
5	$x_5 = 0,3437500$	$-0,0531311035$	$0,0312500$	$0,0531311035$
6	$x_6 = 0,3281250$	$0,0822029114$	$0,0156250$	$0,0822029114$
7	$x_7 = 0,3359375$	$0,0144743919$	$0,0078125$	$0,0144743919$

logo  $\bar{x} = 0,3359375$  é uma boa aproximação para um zero de  $f(x)$ .

### 3.2.1 Estimativa do número de iterações

No método da bissecção, se utilizarmos como critério de parada, o tamanho do intervalo, podemos estimar *a priori* o número de iterações necessárias para satisfazer o critério de parada.

$$\begin{aligned}
 b_i - a_i &= \frac{b_{i-1} - a_{i-1}}{2} = \frac{1}{2} \frac{b_{i-2} - a_{i-2}}{2} = \frac{b_{i-2} - a_{i-2}}{2^2} = \frac{1}{2^2} \frac{b_{i-3} - a_{i-3}}{2} = \\
 &= \frac{b_{i-3} - a_{i-3}}{2^3} = \frac{1}{2^3} \frac{b_{i-4} - a_{i-4}}{2} = \frac{b_{i-4} - a_{i-4}}{2^4} = \dots = \frac{b_0 - a_0}{2^i}
 \end{aligned}$$

ou

$$b_i - a_i = \frac{b_0 - a_0}{2^i}$$

Como queremos que  $b_i - a_i < \epsilon$  podemos escrever

$$\frac{b_0 - a_0}{2^i} < \epsilon$$

$$2^i > \frac{b_0 - a_0}{\epsilon}$$

usando logaritmo temos

$$\log(2^i) > \log\left(\frac{b_0 - a_0}{\epsilon}\right)$$

$$i \log(2) > \log(b_0 - a_0) - \log(\epsilon)$$

$$i > \frac{\log(b_0 - a_0) - \log(\epsilon)}{\log(2)}$$

No exemplo resolvido teríamos

$$i > \frac{\log(1 - 0) - \log(0.01)}{\log(2)} = 6,64$$

Portanto serão necessárias 7 iterações.

**Observações:**

- O método da bissecção sempre converge;
- As iterações são facilmente realizadas;
- A convergência é muito lenta.

### 3.3 Método da Falsa Posição

Este também é um método de quebra. Seja  $[a, b]$  um intervalo que contenha um zero da função contínua  $f(x)$ , isto é,  $f(a) \cdot f(b) < 0$ .

Assim como no Método da Bissecção o intervalo que contém um zero é dividido em dois subintervalos, porém, esta divisão não é mais feita pela média aritmética dos limites do intervalo. O ponto divisor do intervalo nesse método é determinado pela média ponderada entre  $a$  e  $b$  com pesos respectivamente iguais a  $|f(b)|$  e  $|f(a)|$ . Logo, este ponto é calculado pela expressão:

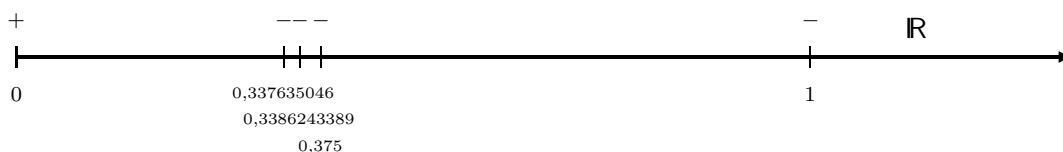
$$x = \frac{a |f(b)| + b |f(a)|}{|f(b)| + |f(a)|} = \frac{a f(b) - b f(a)}{f(b) - f(a)}$$

O método consiste em gerar uma seqüência de aproximações  $\{x_i\}$  a partir de um intervalo inicial que contenha um zero de  $f(x)$ .

A figura abaixo mostra geometricamente como são realizadas as iterações nesse método.

**Exemplo:**

$$f(x) = x^3 - 9x + 3 \text{ com } I = [0, 1] \text{ e } \epsilon = 0,0005$$



Iteração	$x = \frac{a f(b) - b f(a)}{f(b) - f(a)}$	$f(x)$	$ f(x) $
1	$x_1 = 0,375000000$	$-0,322265625$	$0,322265625$
2	$x_2 = 0,338624339$	$-0,008790100$	$0,008790100$
3	$x_3 = 0,337635046$	$-0,000225880$	$0,000225880$

logo  $\bar{x} = 0,337635046$  é uma boa aproximação para um zero de  $f(x)$ .

#### Observações:

- O método da falsa posição sempre converge;
- A convergência é mais rápida que no método da bissecção.

## Métodos de Ponto Fixo

### 3.4 Método Iterativo Linear

Nesse método partimos da equação

$$f(x) = 0$$

Reescrevemos a equação acima da seguinte maneira

$$x = g(x)$$

e usamos  $g(x)$  como função de iteração. Desta forma, a partir de uma aproximação inicial  $x_0$ , geramos uma seqüência de novas aproximações como mostrado abaixo,

$$x_{i+1} = g(x_i)$$

$$x_0$$

$$x_1 = g(x_0)$$

$$x_2 = g(x_1)$$

$$x_3 = g(x_2)$$

$$x_4 = g(x_3)$$

$$\vdots$$

$$x_{i+1} = g(x_i)$$

$$\vdots$$

$g(x)$  é chamada função de iteração para  $f(x) = 0$ .



Por exemplo:

$$x - \cos x = 0$$

$$f(x) = x - \cos x = 0$$

ou

$$x = \cos x$$

$$g(x) = \cos x$$

A figura abaixo mostra graficamente um exemplo

Outro exemplo:

Seja a seguinte equação:

$$x^2 + x - 6 = 0$$

que pode ser reescrita como  $x = g(x)$  de maneiras diferentes

$$g(x) = 6 - x^2$$

$$g(x) = \pm\sqrt{6 - x}$$

$$g(x) = \frac{6}{x} - 1$$

$$g(x) = \frac{6}{x+1}$$

$$g(x) = x^2 + 2x - 6$$

Entretanto, nem sempre o Método Iterativo Linear converge.

Ex:

$$x^2 + x - 6 = 0$$

e

$$g(x) = 6 - x^2; \quad x_0 = 1.5; \quad \epsilon = 0.0001$$

logo

$$x_1 = g(x_0) = 6 - x_0^2 = 6 - (1.5)^2 = 3.75 \longrightarrow |x_1 - x_0| = 2.25$$

$$x_2 = g(x_1) = 6 - x_1^2 = 6 - (3.75)^2 = -8.0625 \longrightarrow |x_2 - x_1| = 11.8125$$

$$x_3 = g(x_2) = 6 - x_2^2 = 6 - (-8.0625)^2 = -59.003906 \longrightarrow |x_3 - x_2| = 50.941406$$

$$x_4 = g(x_3) = 6 - x_3^2 = 6 - (-59.003906)^2 = -3475.4609 \longrightarrow |x_4 - x_3| = 3416.456994$$

⋮

não converge

Ex:

$$x^2 + x - 6 = 0$$

e

$$g(x) = \sqrt{6-x}; \quad x_0 = 1.5; \quad \epsilon = 0.0001$$

logo

$$x_1 = g(x_0) = \sqrt{6-x_0} = \sqrt{6-(1.50000)} = 2.12132 \longrightarrow |x_1 - x_0| = 0.62132$$

$$x_2 = g(x_1) = \sqrt{6-x_1} = \sqrt{6-(2.12132)} = 1.96944 \longrightarrow |x_2 - x_1| = 0.15188$$

$$x_3 = g(x_2) = \sqrt{6-x_2} = \sqrt{6-(1.96944)} = 2.00763 \longrightarrow |x_3 - x_2| = 0.03819$$

$$x_4 = g(x_3) = \sqrt{6-x_3} = \sqrt{6-(2.00763)} = 1.99809 \longrightarrow |x_4 - x_3| = 0.00954$$

$$x_5 = g(x_4) = \sqrt{6-x_4} = \sqrt{6-(1.99809)} = 2.00048 \longrightarrow |x_5 - x_4| = 0.00239$$

$$x_6 = g(x_5) = \sqrt{6-x_5} = \sqrt{6-(2.00048)} = 1.99989 \longrightarrow |x_6 - x_5| = 0.00059$$

$$x_7 = g(x_6) = \sqrt{6-x_6} = \sqrt{6-(1.99989)} = 2.00003 \longrightarrow |x_7 - x_6| = 0.00014$$

$$x_8 = g(x_7) = \sqrt{6-x_7} = \sqrt{6-(2.00003)} = 1.99999 \longrightarrow |x_8 - x_7| = 0.00004$$

$$\vdots$$

converge para 2

**Teorema 3.4.1** *Seja  $\bar{x}$  um zero de uma função  $f(x)$  em um intervalo  $I$  e seja  $g(x)$  uma função de iteração para  $f(x) = 0$ . Então o método iterativo linear converge se:*

1.  $|g'(x)| < 1$ , para todo  $x$  pertencente ao intervalo  $I$ , e
2.  $x_0$  também pertence ao intervalo  $I$ .

**Observação:** Quanto menor for o valor de  $|g'(x)|$  mais rapidamente a seqüência converge.

Nos exemplos:

$$g(x) = 6 - x^2 \longrightarrow g'(x) = -2x$$

$$|g'(x)| < 1 \longrightarrow |-2x| < 1 \longrightarrow -1 < 2x < 1 \longrightarrow -\frac{1}{2} < x < \frac{1}{2} \longrightarrow I = \left(-\frac{1}{2}, \frac{1}{2}\right)$$

mas

$$x_0 = 1.5 \quad \text{e} \quad \bar{x} = 2 \quad \text{ou} \quad \bar{x} = -3$$

portanto

$$x_0 \notin I \quad \text{e} \quad \bar{x} \notin I$$

e

$$g(x) = \sqrt{6-x} \longrightarrow g'(x) = -\frac{1}{2\sqrt{6-x}}$$

$$|g'(x)| < 1 \longrightarrow \left| -\frac{1}{2\sqrt{6-x}} \right| < 1 \longrightarrow \frac{1}{2} < \sqrt{6-x} \longrightarrow$$

$$\longrightarrow \frac{1}{4} < 6-x \longrightarrow x < 6 - \frac{1}{4} = 5.75 \longrightarrow I = (-\infty, 5.75)$$

mas

$$x_0 = 1.5 \quad \text{e} \quad \bar{x} = 2 \quad \text{ou} \quad \bar{x} = -3$$

portanto

$$x_0 \in I \quad \text{e} \quad \bar{x} \in I$$

### 3.5 Método de Newton-Raphson

Funções de iteração podem ser obtidas sistematicamente da seguinte maneira. Se  $\xi$  é o zero de uma função  $f : \mathbb{R} \rightarrow \mathbb{R}$ , e se  $f$  é suficientemente diferenciável em uma vizinhança  $\mathcal{N}(\xi)$  em torno do zero  $\xi$  de  $f(x)$ , então expandindo a função em série de Taylor em torno de  $x_0 \in \mathcal{N}(\xi)$  temos

$$f(\xi) = 0 = f(x_0) + (\xi - x_0)f'(x_0) + \frac{(\xi - x_0)^2}{2!}f''(x_0) + \dots + \frac{(\xi - x_0)^k}{k!}f^{(k)}(x_0 + \vartheta(\xi - x_0))$$

Se as maiores potências  $(\xi - x_0)^\nu$  são desprezadas, chegamos a equações que expressam o ponto  $\xi$  aproximadamente em termos de um ponto dado  $x_0$ , por exemplo:

$$0 = f(x_0) + (\bar{\xi} - x_0)f'(x_0)$$

ou

$$0 = f(x_0) + (\bar{\xi}^* - x_0)f'(x_0) + \frac{(\bar{\xi}^* - x_0)^2}{2!}f''(x_0)$$

que resultam, respectivamente

$$\bar{\xi} = x_0 - \frac{f(x_0)}{f'(x_0)}$$

e

$$\bar{\xi}^* = x_0 - \frac{f'(x_0) \pm \sqrt{(f'(x_0))^2 - 2f(x_0)f''(x_0)}}{f''(x_0)}$$

as expressões acima fornecem simplesmente valores próximos do zero desejado, mas elas foram deduzidas para fornecerem funções de iteração. Dessa maneira, chega-se aos seguintes processos iterativos.

$$x_{i+1} = \Phi(x_i), \quad \Phi(x) = x - \frac{f(x)}{f'(x)}$$

e

$$x_{i+1} = \Phi_{\pm}(x_i), \quad \Phi_{\pm}(x) = x - \frac{f'(x) \pm \sqrt{(f'(x))^2 - 2f(x)f''(x)}}{f''(x)}$$

O primeiro é o clássico método de Newton-Raphson. O segundo é uma extensão óbvia do primeiro (também chamado Método de Cauchy).

Graficamente temos a seguinte representação

O método de Newton-Raphson é obtido *linearizando-se*  $f$ .

Resumindo temos:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

No método de Newton-Raphson a nova aproximação ( $x_{i+1}$ ) é obtida a partir da reta tangente à função usando-se a aproximação anterior. A interseção da reta tangente e o eixo  $x$  define a nova aproximação, como mostra a figura abaixo.

### Observações:

(a) O método de Newton pode ser interpretado com uma tentativa de garantir a convergência do método iterativo linear, para isso escolhe-se a função de iteração de forma que  $g'(x) = 0$ . Garantindo assim que  $|g'(x)| \leq M < 1$  e que a convergência será mais rápida.

(b) Aproximação Inicial e Convergência

Exemplo:

$$x^2 + x - 6 = 0$$

e

$$x_0 = 1.5; \quad \epsilon = 0.001$$

$$g(x) = x - \frac{f(x)}{f'(x)} = x - \frac{x^2 + x - 6}{2x + 1} = \frac{x^2 + 6}{2x + 1}$$

logo

$$x_1 = g(x_0) = \frac{x_0^2 + 6}{2x_0 + 1} = \frac{1.5^2 + 6}{2 \cdot 1.5 + 1} = 2.06250 \longrightarrow |x_1 - x_0| = 0.56250$$

$$x_2 = g(x_1) = \frac{x_1^2 + 6}{2x_1 + 1} = \frac{2.06250^2 + 6}{2 \cdot 2.06250 + 1} = 2.00076 \longrightarrow |x_2 - x_1| = 0.06174$$

$$x_3 = g(x_2) = \frac{x_2^2 + 6}{2x_2 + 1} = \frac{2.00076^2 + 6}{2 \cdot 2.00076 + 1} = 2.00000 \longrightarrow |x_3 - x_2| = 0.00076$$

⋮

converge para 2

Outro Exemplo:

$$f(x) = x^3 - 9x + 3$$

já vimos que as raízes estão nos intervalos  $(-4, -3)$ ;  $(0, 1)$ ;  $(2, 3)$

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

seja  $x_0 = 1.5$  e  $\epsilon = 2 \times 10^{-2}$ , então

$$x_0 = 1.50000000$$

$$x_1 = -1.66666666$$

$$x_2 = 18.38888888$$

$$x_3 = 12.36601040$$

$$x_4 = 8.40230675$$

$$x_5 = 5.83533816$$

$$x_6 = 4.23387355$$

$$x_7 = 3.32291026$$

$$x_8 = 2.91733893$$

$$x_9 = 2.82219167$$

$$x_{10} = 2.81692988$$

Note que neste exemplo o processo iterativo tende a divergir inicialmente. Isto se deve ao fato de que  $f'(x_1) \approx 0$  ( $\sqrt{3}$  é raiz de  $f'(x)$ ). Havendo neste ponto uma divisão por um número próximo de zero.

### 3.6 Método da Secante

Este método substitui a avaliação da derivada da função no ponto  $x_i$  por uma aproximação desta derivada. Isto é,

$$f'(x_i) \approx \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}$$

Substituindo-se a expressão acima na função de iteração do método de Newton-Raphson, obtemos a função de iteração para o Método da Secante.

$$x_{i+1} = \frac{x_{i-1} \cdot f(x_i) - x_i \cdot f(x_{i-1})}{f(x_i) - f(x_{i-1})}$$

Nota-se claramente que para se obter  $x_{i+1}$  precisa-se de dois valores  $x_{i-1}$  e  $x_i$ , portanto necessita-se de duas estimativas iniciais para se iniciar o método. Graficamente temos

Exemplo:

$$x^2 + x - 6 = 0$$

e

$$x_0 = 1.5; \quad x_1 = 1.7 \quad \epsilon = 0.01$$

logo

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)} = \frac{1.5 \cdot (-1.41) - 1.7 \cdot (-2.25)}{(-1.41) - (-2.25)} = 2.03571 \longrightarrow |x_2 - x_1| = 0.33571$$

$$x_3 = \frac{x_1 f(x_2) - x_2 f(x_1)}{f(x_2) - f(x_1)} = 1.99774 \longrightarrow |x_3 - x_2| = 0.03797$$

$$x_4 = \frac{x_2 f(x_3) - x_3 f(x_2)}{f(x_3) - f(x_2)} = 1.99999 \longrightarrow |x_4 - x_3| = 0.00225$$

⋮

converge para 2

**Observações:** Pode divergir se  $f(x_i) \approx f(x_{i+1})$



## Ordem de Convergência

**Definição 3.6.1** *Seja uma seqüência  $x_0, x_1, x_2, \dots$  que converge para  $\bar{x}$ . Seja  $e_i = |x_i - \bar{x}|$ . Se existir um número  $p > 1$  e uma constante  $C \neq 0$  tal que*

$$\lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i^p} = C$$

*então,  $p$  é dito ordem de convergência da seqüência e  $C$  é a constante assintótica de erro. Se  $p = 1$  dizemos que a convergência é linear, neste caso para que haja convergência  $C < 1$ .*

### Método Iterativo Linear:

$$x_{i+1} = g(x_i)$$

e

$$\bar{x} = g(\bar{x})$$

subtraindo as duas relações

$$x_{i+1} - \bar{x} = g(x_i) - g(\bar{x})$$

mas usando o Teorema do Valor Médio temos

$$x_{i+1} - \bar{x} = g(x_i) - g(\bar{x}) = g'(c_i)(x_i - \bar{x}) \quad c_i \in (x_i, \bar{x})$$

ou

$$\frac{e_{i+1}}{e_i} = \frac{x_{i+1} - \bar{x}}{x_i - \bar{x}} = g'(c_i)$$

que por continuidade

$$\lim_{i \rightarrow \infty} \frac{e_{i+1} - \bar{x}}{x_i - \bar{x}} = \lim_{i \rightarrow \infty} g'(c_i) = g'(\lim_{i \rightarrow \infty} c_i) = g'(\xi) = C$$

$$\lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i} = g'(\xi) = C$$

logo o Método Iterativo Linear tem convergência pelo menos linear.

### Newton-Raphson:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

$$x_{i+1} - \bar{x} = x_i - \bar{x} - \frac{f(x_i)}{f'(x_i)}$$

$$e_{i+1} = e_i - \frac{f(x_i)}{f'(x_i)}$$

Desenvolvendo em série de Taylor para  $x = x_i$

$$f(x) = f(x_i) + f'(x_i)(x - x_i) + \frac{f''(c_i)}{2}(x - x_i)^2 \quad c_i \in (x, x_i)$$

$$x = \bar{x}$$

$$0 = f(\bar{x}) = f(x_i) - f'(x_i)(x_i - \bar{x}) + \frac{f''(c_i)}{2}(x_i - \bar{x})^2$$

$$\implies f(x_i) = f'(x_i)(x_i - \bar{x}) - \frac{f''(c_i)}{2}(x_i - \bar{x})^2$$

que dividindo por  $f'(x_i)$  ( $\neq 0$ ) por hipótese

$$\frac{f(x_i)}{f'(x_i)} = x_i - \bar{x} - \frac{f''(c_i)}{2f'(x_i)}(x_i - \bar{x})^2$$

$$\frac{f''(c_i)}{2f'(x_i)}e_i^2 = e_i - \frac{f(x_i)}{f'(x_i)} = e_{i+1}$$

visto que

$$\left( x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \right) \implies x_{i+1} - \bar{x} = x_i - \bar{x} - \frac{f(x_i)}{f'(x_i)} \implies e_{i+1} = e_i - \frac{f(x_i)}{f'(x_i)}$$

Assim

$$\frac{e_{i+1}}{e_i^2} = \frac{1}{2} \frac{f''(c_i)}{f'(x_i)}$$

$$\lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i^2} = \frac{1}{2} \lim_{i \rightarrow \infty} \frac{f''(c_i)}{f'(x_i)} = \frac{1}{2} \frac{\lim_{i \rightarrow \infty} f''(c_i)}{\lim_{i \rightarrow \infty} f'(x_i)} = \frac{1}{2} \frac{f''(\lim_{i \rightarrow \infty} c_i)}{f'(\lim_{i \rightarrow \infty} x_i)} = \frac{1}{2} \frac{f''(\bar{x})}{f'(\bar{x})} = C$$

logo

$$\lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i^2} = C$$

Portanto o Método de Newton-Raphson tem convergência quadrática

**Bisseccão:**

Resumindo

Método	Ordem de Convergência	
Bisseccão	1	Linear
Falsa Posição	—	—
Iterativo Linear	1	Linear
Newton-Raphson	2	Quadrática
Secante	1,618	—

# Sistemas Não-Lineares

## Resolução de Sistemas de Equações Não-Lineares

De uma maneira geral, estudaremos o problema de determinar os zeros de uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , que é descrita por  $n$  funções reais  $f_k(x_1, x_2, \dots, x_n)$ ;  $k = 1, 2, 3, \dots, n$  de  $n$  variáveis reais  $x_1, x_2, \dots, x_n$ . Ou seja,

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ f_3(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

O problema de resolver  $\mathbf{f}(\mathbf{x}) = \vec{0}$  torna-se o de resolver um sistema de equações não-lineares:

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0 \\ f_2(x_1, x_2, \dots, x_n) = 0 \\ f_3(x_1, x_2, \dots, x_n) = 0 \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) = 0 \end{cases}$$

Normalmente não é possível determinar um zero  $\boldsymbol{\xi}$  de uma função  $\mathbf{f}$  explicitamente, então temos que procurar outros métodos que aproximam a solução. Estes métodos são normalmente iterativos e tem a seguinte forma:

- Partindo-se de um valor inicial  $\mathbf{x}^{(0)}$ , novas aproximações sucessivas  $\mathbf{x}^{(i)}$ ,  $i = 1, 2, \dots$  são calculadas com a ajuda de uma função de iteração  $\Phi(\mathbf{x})$

$$\mathbf{x}^{(i+1)} = \Phi(\mathbf{x}^{(i)}), \quad i = 0, 1, 2, \dots$$

Espera-se que no limite a seqüência de aproximações convirja para a solução do sistema, isto é

$$\lim_{i \rightarrow \infty} \mathbf{x}^{(i)} = \boldsymbol{\xi}$$

Assim como no caso uni-dimensional, algumas questões surgem:

1. Como interromper a geração de novas aproximações?
2. Como podemos encontrar uma função de iteração adequada?

3. Sob que condições a seqüência  $\{\mathbf{x}^{(i)}\}$  converge?
4. Quão rapidamente a seqüência  $\{\mathbf{x}^{(i)}\}$  converge?

### Critério de Parada

Quando parar o processo iterativo?

Uma maneira de se determinar se a seqüência de aproximações gerada pelo método iterativo está convergindo é verificar se a diferença entre a aproximação e o valor exato do zero está diminuindo. Seja  $\mathbf{x}^{(i)}$  o valor aproximado obtido na iteração  $i$  e  $\bar{\mathbf{x}}$  o valor exato do zero, podemos verificar se

$$\|\mathbf{x}^{(i)} - \bar{\mathbf{x}}\| < \epsilon$$

isto é, se a diferença acima citada é menor que um valor *pequeno*. Chamamos  $\epsilon$  de tolerância de parada ou simplesmente tolerância.  $\|\cdot\|$  denota uma norma definida no  $\mathbb{R}^n$ .

Entretanto, como não conhecemos o valor exato do zero da função (queremos na verdade determinar este zero), o teste acima não pode ser efetuado. Substituímos o teste acima pelo seguinte teste.

$$\|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| < \epsilon$$

A diferença agora é calculada entre duas aproximações sucessivas.

Outro teste que também pode ser realizado é

$$\frac{\|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|}{\|\mathbf{x}^{(i+1)}\|} < \epsilon$$

ou ainda

$$\|\mathbf{f}(\mathbf{x}^{(i)})\| < \epsilon$$

Em geral a norma do  $\mathbb{R}^n$  usada é a norma do máximo, isto é

$$\|\mathbf{x}\| = \max_{1 \leq k \leq n} |x_k|$$

### Funções de Iteração

Vamos examinar como as funções de iterações podem ser construídas.

#### Método Iterativo Linear

Algumas vezes as funções de iteração são sugeridas pela formulação do problema. Por exemplo, se a equação é  $x - \cos x = 0$ , então é natural tentar um processo iterativo com:

$$x_{i+1} = \cos x_i, \quad i = 0, 1, 2, \dots$$

ou seja,

$$\Phi(x) = \cos x$$

analogamente ao caso escalar, em que transformávamos a equação  $f(x) = 0$  em  $x = g(x)$  e utilizávamos  $g(x)$  como função de iteração.

Mais precisamente, dado um sistema não-linear

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0 \\ f_2(x_1, x_2, \dots, x_n) = 0 \\ f_3(x_1, x_2, \dots, x_n) = 0 \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) = 0 \end{cases}$$

Da primeira equação do sistema explicitamos a incógnita  $x_1$ , da segunda equação do sistema a incógnita  $x_2$ , da terceira equação a incógnita  $x_3$ , e assim por diante até a última equação do sistema, onde então, explicitamos a incógnita  $x_n$ . Assim, reescrevemos o sistema como:

$$\begin{cases} x_1 = g_1(x_1, x_2, \dots, x_n) \\ x_2 = g_2(x_1, x_2, \dots, x_n) \\ x_3 = g_3(x_1, x_2, \dots, x_n) \\ \vdots \\ x_n = g_n(x_1, x_2, \dots, x_n) \end{cases}$$

A função de iteração é então definida como

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(x_1, x_2, \dots, x_n) \\ g_2(x_1, x_2, \dots, x_n) \\ g_3(x_1, x_2, \dots, x_n) \\ \vdots \\ g_n(x_1, x_2, \dots, x_n) \end{bmatrix}$$

Donde

$$\mathbf{x}^{(i+1)} = \mathbf{g}(\mathbf{x}^{(i)})$$

ou

$$\begin{cases} x_1^{(i+1)} = g_1(x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}) \\ x_2^{(i+1)} = g_2(x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}) \\ x_3^{(i+1)} = g_3(x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}) \\ \vdots \\ x_n^{(i+1)} = g_n(x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}) \end{cases}$$

Logo, se tivermos uma aproximação inicial para a solução do sistema, podemos gerar uma sequência de aproximações que *esperamos* convirja para a solução. Isto é,

$$\lim_{i \rightarrow \infty} \mathbf{x}^{(i)} = \boldsymbol{\xi}$$

**Observação:** Assim como no caso unidimensional existem infinitas funções de iteração para  $\mathbf{f}(\mathbf{x}) = \vec{0}$ .

**Exemplo:**

Seja o sistema de equações não-lineares dado abaixo.

$$\begin{cases} x^2 + y^2 = 1 \\ x - 2y = 0 \end{cases}$$

ou

$$\begin{cases} x_1^2 + x_2^2 = 1 \\ x_1 - 2x_2 = 0 \end{cases}$$

que pode ser reescrito da forma

$$\begin{cases} x_1 = \sqrt{1 - x_2^2} \\ x_2 = \frac{x_1}{2} \end{cases}$$

ou

$$\mathbf{x} = \mathbf{g}(\mathbf{x}) = \begin{bmatrix} \sqrt{1 - x_2^2} \\ \frac{x_1}{2} \end{bmatrix}$$

Como

$$\mathbf{x}^{(i+1)} = \mathbf{g}(\mathbf{x}^{(i)})$$

$$\begin{cases} x_1^{(i+1)} = \sqrt{1 - (x_2^{(i)})^2} \\ x_2^{(i+1)} = \frac{x_1^{(i)}}{2} \end{cases}$$

Usando como aproximação inicial  $\mathbf{x}^{(0)} = (0.5; 0.5)$ , com uma tolerância  $\epsilon = 0.01$ , temos

**Primeira iteração:**

$$\mathbf{x}^{(1)} = \mathbf{g}(\mathbf{x}^{(0)}) = \begin{bmatrix} \sqrt{1 - (x_2^{(0)})^2} \\ \frac{x_1^{(0)}}{2} \end{bmatrix} = \begin{bmatrix} \sqrt{1 - (0.5)^2} \\ \frac{0.5}{2} \end{bmatrix} = \begin{bmatrix} \sqrt{0.75} \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0.866025404 \\ 0.25 \end{bmatrix}$$

Critério de Parada

$$\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| = \left\| \begin{bmatrix} 0.866025404 \\ 0.25 \end{bmatrix} - \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0.366025404 \\ -0.25 \end{bmatrix} \right\| = 0.366025404 > \epsilon$$

**Segunda iteração:**

$$\mathbf{x}^{(2)} = \mathbf{g}(\mathbf{x}^{(1)}) = \begin{bmatrix} \sqrt{1 - (x_2^{(1)})^2} \\ \frac{x_1^{(1)}}{2} \end{bmatrix} = \begin{bmatrix} \sqrt{1 - (0.25)^2} \\ \frac{0.866025404}{2} \end{bmatrix} = \begin{bmatrix} \sqrt{0.9375} \\ 0.433012702 \end{bmatrix} = \begin{bmatrix} 0.968245837 \\ 0.433012702 \end{bmatrix}$$

Critério de Parada

$$\|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}\| = \left\| \begin{bmatrix} 0.968245837 \\ 0.433012702 \end{bmatrix} - \begin{bmatrix} 0.866025404 \\ 0.25 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0.102220433 \\ 0.183012702 \end{bmatrix} \right\| = 0.183012702 > \epsilon$$

Resumindo

Iteração	$x_1^{(i)}$	$x_2^{(i)}$	$\ \mathbf{x}^{(i)} - \mathbf{x}^{(i-1)}\ $
1	0.8660254	0.2500000	0.366025400
2	0.9682459	0.4330127	0.183012700
3	0.9013878	0.4841229	0.066858050
4	0.8750000	0.4506939	0.033429030
5	0.8926786	0.4375000	0.017678560
6	0.8992184	0.4463393	0.008839279

Portanto, uma aproximação para o zero do sistema de equações dado é  $\bar{\mathbf{x}} = (0.8992184; 0.4463393)$ .

## Método de Newton

Outra maneira de criar funções de iteração é gerá-las sistematicamente, como descrito a seguir:

Seja  $\mathcal{N}(\xi)$  uma vizinhança em torno do zero  $\xi$  de  $f(x)$ , então expandindo a função em série de Taylor em torno de  $x_0 \in \mathcal{N}(\xi)$  temos

$$f(\xi) = 0 = f(x_0) + (\xi - x_0)f'(x_0) + \frac{(\xi - x_0)^2}{2!}f''(x_0) + \dots + \frac{(\xi - x_0)^k}{k!}f^{(k)}(x_0) + \vartheta(\xi - x_0)$$

Se as maiores potências  $(\xi - x_0)^\nu$  são desprezadas, chegamos a equações que expressam o ponto  $\xi$  aproximadamente em termos de um ponto dado  $x_0$ , por exemplo:

$$0 = f(x_0) + (\bar{\xi} - x_0)f'(x_0)$$

ou

$$0 = f(x_0) + (\bar{\xi}^* - x_0)f'(x_0) + \frac{(\bar{\xi}^* - x_0)^2}{2!}f''(x_0)$$

que fornecem, respectivamente

$$\bar{\xi} = x_0 - \frac{f(x_0)}{f'(x_0)}$$

e

$$\bar{\xi}^* = x_0 - \frac{f'(x_0) \pm \sqrt{(f'(x_0))^2 - 2f(x_0)f''(x_0)}}{f''(x_0)}$$

as expressões acima fornecem simplesmente valores próximos do zero desejado, mas elas foram deduzidas para fornecerem funções de iteração. Dessa maneira, chega-se aos seguintes processos iterativos.

$$x_{i+1} = \Phi(x_i), \quad \Phi(x) = x - \frac{f(x)}{f'(x)}$$

e

$$x_{i+1} = \Phi_{\pm}(x_i), \quad \Phi_{\pm}(x) = x - \frac{f'(x) \pm \sqrt{(f'(x))^2 - 2f(x)f''(x)}}{f''(x)}$$

O primeiro é o clássico método de Newton-Raphson. O segundo é uma extensão óbvia do primeiro (também chamado Método de Cauchy).

Graficamente temos a seguinte representação

O método de Newton-Raphson é obtido *linearizando-se*  $f$ . A *linearização* é também um meio de construir métodos iterativos para resolver sistemas da forma

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ f_3(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{bmatrix} = \vec{0}$$



Analogamente ao processo usado na obtenção da função de iteração do método de Newton-Raphson, temos

$$\mathbf{f}(\boldsymbol{\xi}) = \vec{0} \approx \mathbf{f}(\mathbf{x}^{(0)}) + \mathbf{Df}(\mathbf{x}^{(0)})(\boldsymbol{\xi} - \mathbf{x}^{(0)})$$

onde

$$\mathbf{Df}(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}) & \frac{\partial f_n}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{bmatrix}$$

e

$$\boldsymbol{\xi} - \mathbf{x}^{(0)} = \begin{bmatrix} \xi_1 - x_1^{(0)} \\ \xi_2 - x_2^{(0)} \\ \vdots \\ \xi_n - x_n^{(0)} \end{bmatrix}$$

Se  $\mathbf{Df}(\mathbf{x}^{(0)})$  é não-singular

$$\mathbf{f}(\mathbf{x}^{(0)}) + \mathbf{Df}(\mathbf{x}^{(0)})(\mathbf{x}^{(1)} - \mathbf{x}^{(0)}) = \vec{0}$$

pode ser resolvida para  $\mathbf{x}^{(1)}$ :

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - [\mathbf{Df}(\mathbf{x}^{(0)})]^{-1} \mathbf{f}(\mathbf{x}^{(0)})$$

e, generalizando para melhores aproximações do zero

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - [\mathbf{Df}(\mathbf{x}^{(i)})]^{-1} \mathbf{f}(\mathbf{x}^{(i)}) \quad i = 0, 1, 2, \dots$$

A resolução da equação acima envolve a obtenção da inversa da matriz  $\mathbf{Df}(\mathbf{x}^{(i)})$  a cada iteração, o que é uma tarefa dispendiosa. Podemos contornar este cálculo reescrevendo a equação acima

$$\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)} = -[\mathbf{Df}(\mathbf{x}^{(i)})]^{-1} \mathbf{f}(\mathbf{x}^{(i)}) \quad i = 0, 1, 2, \dots$$

premultiplicando ambos os lados por  $\mathbf{Df}(\mathbf{x}^{(i)})$ , temos

$$\mathbf{Df}(\mathbf{x}^{(i)}) (\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}) = -\mathbf{Df}(\mathbf{x}^{(i)}) [\mathbf{Df}(\mathbf{x}^{(i)})]^{-1} \mathbf{f}(\mathbf{x}^{(i)}) \quad i = 0, 1, 2, \dots$$

$$\mathbf{Df}(\mathbf{x}^{(i)}) (\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}) = -\mathbf{f}(\mathbf{x}^{(i)}) \quad i = 0, 1, 2, \dots$$

ou

$$\mathbf{Df}(\mathbf{x}^{(i)}) \Delta \mathbf{x} = -\mathbf{f}(\mathbf{x}^{(i)}) \quad i = 0, 1, 2, \dots$$

onde

$$\Delta \mathbf{x} = \mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}$$

Desta forma podemos dado  $\mathbf{x}^{(i)}$  resolver o sistema linear

$$D\mathbf{f}(\mathbf{x}^{(i)})\Delta \mathbf{x} = -\mathbf{f}(\mathbf{x}^{(i)}) \quad i = 0, 1, 2, \dots$$

para  $\Delta \mathbf{x}$  e a seguir calcular a nova aproximação  $\mathbf{x}^{(i+1)}$  por

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \Delta \mathbf{x}$$

Note que a cada iteração avaliamos a matriz  $D\mathbf{f}(\mathbf{x}^{(i)})$  e resolvemos um sistema linear.

**Exemplo:**

Seja o seguinte sistema de equações não-lineares,

$$\begin{cases} x^2 + y^2 = 1 \\ x - 2y = 0 \end{cases}$$

ou

$$\begin{cases} x_1^2 + x_2^2 = 1 \\ x_1 - 2x_2 = 0 \end{cases}$$

ou ainda

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2^2 - 1 \\ x_1 - 2x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}$$

logo

$$D\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 & 2x_2 \\ 1 & -2 \end{bmatrix}$$

Usaremos como no exemplo anterior  $\mathbf{x}^{(0)} = (0.5; 0.5)$  e  $\epsilon = 0.01$ .

**Primeira iteração:**

$$D\mathbf{f}(\mathbf{x}^{(0)})\Delta \mathbf{x} = -\mathbf{f}(\mathbf{x}^{(0)})$$

$$D\mathbf{f}(\mathbf{x}^{(0)}) = \begin{bmatrix} 2x_1^{(0)} & 2x_2^{(0)} \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 2 \times 0.5 & 2 \times 0.5 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$$

$$\mathbf{f}(\mathbf{x}^{(0)}) = \begin{bmatrix} (x_1^{(0)})^2 + (x_2^{(0)})^2 - 1 \\ x_1^{(0)} - 2x_2^{(0)} \end{bmatrix} = \begin{bmatrix} 0.5^2 + 0.5^2 - 1 \\ 0.5 - 2 \times 0.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} =$$

Logo

$$\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = - \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

Resolvendo o sistema linear obtemos

$$\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.0 \end{bmatrix}$$

Podemos agora calcular  $\mathbf{x}^{(1)}$

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \Delta \mathbf{x} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.0 \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}$$

Critério de Parada

$$\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| = \|\Delta \mathbf{x}\| = \left\| \begin{bmatrix} 0.5 \\ 0.0 \end{bmatrix} \right\| = 0.5 > \epsilon$$

**Segunda iteração:**

$$D\mathbf{f}(\mathbf{x}^{(1)})\Delta \mathbf{x} = -\mathbf{f}(\mathbf{x}^{(1)})$$

$$D\mathbf{f}(\mathbf{x}^{(1)}) = \begin{bmatrix} 2x_1^{(1)} & 2x_2^{(1)} \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 2 \times 1.0 & 2 \times 0.5 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}$$

$$\mathbf{f}(\mathbf{x}^{(1)}) = \begin{bmatrix} (x_1^{(1)})^2 + (x_2^{(1)})^2 - 1 \\ x_1^{(1)} - 2x_2^{(1)} \end{bmatrix} = \begin{bmatrix} 1.0^2 + 0.5^2 - 1 \\ 1.0 - 2 \times 0.5 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.0 \end{bmatrix} =$$

Logo

$$\begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = - \begin{bmatrix} 0.25 \\ 0.0 \end{bmatrix} = \begin{bmatrix} -0.25 \\ 0.0 \end{bmatrix}$$

Resolvendo o sistema linear obtemos

$$\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} -0.10 \\ -0.05 \end{bmatrix}$$

Podemos agora calcular  $\mathbf{x}^{(2)}$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \Delta \mathbf{x} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix} + \begin{bmatrix} -0.10 \\ -0.05 \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.45 \end{bmatrix}$$

Critério de Parada

$$\|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}\| = \|\Delta \mathbf{x}\| = \left\| \begin{bmatrix} -0.10 \\ -0.05 \end{bmatrix} \right\| = 0.1 > \epsilon$$

Resumindo

Iteração	$x_1^{(i)}$	$x_2^{(i)}$	$\ \mathbf{x}^{(i)} - \mathbf{x}^{(i-1)}\ $
1	1.0000000	0.5000000	0.500000000
2	0.9000000	0.4500000	0.100000000
3	0.8944445	0.4472222	0.005555511

Portanto, uma aproximação para o zero do sistema de equações dado é  $\tilde{\mathbf{x}} = (0.8944445; 0.4472222)$ .

**Observação:**

- (a) As velocidades de convergência dos métodos multidimensionais são análogas àquelas do caso unidimensional. Isto é, a convergência do método iterativo linear é linear (Ordem de Convergência igual a 1) e a convergência do método de Newton é quadrática (Ordem de Convergência igual a 2).
- (b) Métodos quasi-Newton.

Como o método de Newton necessita a cada iteração recalculer os elementos de  $D\mathbf{f}(\mathbf{x}^{(i)})$  e resolver o sistema linear. Existem variações do método de Newton que propõem manter a matriz  $D\mathbf{f}(\mathbf{x}^{(i)})$  fixa durante algumas iterações. No caso unidimensional significa manter a derivada fixa...

# Zero de Funções – Exercícios

1. Determine a raiz quadrada negativa de 0,5 até quatro casas decimais escrevendo  $f(x) = x^2 - 0,5$  e resolvendo por  $x = x^2 + x - 0,5$  pelo Método Iterativo Linear usando  $x_0 = -0,6$ . Poderia a raiz quadrada positiva ser determinada por este método?

## Solução:

Com 4 casas decimais  $\rightarrow |x_{i+1} - x_i| < \epsilon = 0,0001$

$$x_{i+1} = g(x_i) = x_i^2 + x_i - 0,5$$

efetuando as iterações, obtemos:

$x_0 = -0,6$	
$x_1 = g(x_0) = -0,74$	$ x_1 - x_0  = 0,140000000$
$x_2 = g(x_1) = -0,6924$	$ x_2 - x_1  = 0,047600000$
$x_3 = g(x_2) = -0,712982240$	$ x_3 - x_2  = 0,020582240$
$x_4 = g(x_3) = -0,704638565$	$ x_4 - x_3  = 0,008343675$
$x_5 = g(x_4) = -0,708123058$	$ x_5 - x_4  = 0,003484493$
$x_6 = g(x_5) = -0,706684793$	$ x_6 - x_5  = 0,001438265$
$x_7 = g(x_6) = -0,707281396$	$ x_7 - x_6  = 0,000596603$
$x_8 = g(x_7) = -0,707034423$	$ x_8 - x_7  = 0,000246973$
$x_9 = g(x_8) = -0,707136748$	$ x_9 - x_8  = 0,000102325$
$x_{10} = g(x_9) = -0,707094368$	$ x_{10} - x_9  = 0,000042380$

2. Repita as iterações do exemplo anterior usando o Método de Newton-Raphson.

## Solução:

A função de iteração do Método de Newton-Raphson é:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

Assim com  $x_0 = -0,6$  geramos a seguinte seqüência de aproximações;

Iteração	$x_{i+1}$	$ x_{i+1} - x_i $
1	-0,7166666	$ x_1 - x_0  = 0,116666600$
2	-0,7071705	$ x_2 - x_1  = 0,009496093$
3	-0,7071068	$ x_3 - x_2  = 0,000063777$

3. Determine com até três casas decimais o zero de  $0,1x^2 - x \ln x = 0$  entre 1 e 2.

**Solução:**

$$f(x) = 0,1x^2 - x \ln x = 0 \text{ em } [1, 2]$$

Com 3 casas decimais  $\rightarrow |x_{i+1} - x_i| < \epsilon = 0,001$

Temos que  $f(1) = 0,1$  e  $f(2) = -0,986294361 \rightarrow f(1) \cdot f(2) < 0$ .

No Método da Bisseção o ponto divisor do intervalo é

$$x = \frac{a + b}{2}$$

Logo,

Iteração	$x_i$	$f(x_i)$	$a_i$	$b_i$	$ b_i - a_i $
1	1,500000	-0,38319770	1,000000	1,500000	0,50000000
2	1,250000	-0,12267940	1,000000	1,250000	0,25000000
3	1,125000	-0,00594341	1,000000	1,125000	0,12500000
4	1,062500	0,04847696	1,062500	1,125000	0,06250000
5	1,093750	0,02161561	1,093750	1,125000	0,03125000
6	1,109375	0,00792172	1,109375	1,125000	0,01562500
7	1,117188	0,00101037	1,117188	1,125000	0,00781250
8	1,121094	-0,00246124	1,117188	1,121094	0,00390625
9	1,119141	-0,00072411	1,117188	1,119141	0,00195313
10	1,118164	0,00014346	1,118164	1,119141	0,00097656

Portanto, uma aproximação para o zero é 1,11816400.

4. Repita o exemplo anterior usando o Método da Falsa Posição.

**Solução:**

No Método da Falsa Posição o ponto que divide o intervalo em dois subintervalos é determinado pela expressão:

$$x = \frac{a f(b) - b f(a)}{f(b) - f(a)}$$

Efetuada as iterações temos.

Iteração	$x_i$	$f(x_i)$	$a_i$	$b_i$	$ x_{i+1} - x_i $
1	1,092056	0,02308979	1,092056	2,0000	0,09205604
2	1,112825	0,00487458	1,112825	2,0000	0,02076936
3	1,117189	0,00100942	1,117189	2,0000	0,00436318
4	1,118091	0,00020824	1,118091	2,0000	0,00090253

Portanto, uma aproximação para o zero é 1,11809100.

5. Use o Método da Secante para encontrar uma aproximação para um zero da função  $f(x) = x \ln x - 1$  com uma tolerância de  $10^{-7}$ . Comece o processo iterativo com as aproximações iniciais  $x_0 = 2,3$  e  $x_1 = 2,7$ .

**Solução:**

No Método da Secante uma nova aproximação é obtida a partir de duas aproximações anteriores pela seguinte função de iteração:

$$x_{i+1} = \frac{x_{i-1}f(x_i) - x_i f(x_{i-1})}{f(x_i) - f(x_{i-1})}$$

Iteração	$x_{i+1}$	$ x_{i+1} - x_i $
2	1,821888	0,8781122000
3	1,770545	0,0513432000
4	1,763299	0,0072458980
5	1,763223	0,0000758171
6	1,763223	0,0000001192

6. Um modelo para crescimento da população de bactérias supõe que a taxa de mudança da população ( $p$ ) é proporcional à população existente num determinado instante de tempo, logo,

$$\frac{dp}{dt} = kp$$

se existe uma limitação de população, por falta de alimento, por exemplo, a equação acima transforma-se em,

$$\frac{dp}{dt} = kp(p_{max} - p)$$

resolvendo-se a equação diferencial analiticamente, obtemos:

$$p(t) = \frac{p_{max}}{1 + \left(\frac{p_{max}}{p_0} - 1\right) e^{-kp_{max}t}} \quad p(t=0) = p_0$$

Suponha que queremos estudar o crescimento de uma população de bactérias num lago.

Na primavera ( $t = 0$ ) a população é pequena,  $p = 10$  células/litro. É conhecido que a população atinge uma densidade de 15.000 células/litro quando  $t = 60$  dias e que  $k = 2 \times 10^{-6}$  litros por célula por dia. Deseja-se saber a densidade da população quando  $t = 90$  dias. Se a população de bactérias exceder 40.000 células/litro, para manter a qualidade da água é necessário um tratamento para a proteção dos banhistas.

**Solução:**

Com as informações acima,

$$15000 = \frac{p_{max}}{1 + \left(\frac{p_{max}}{10} - 1\right) e^{-2 \times 10^{-6}(p_{max})60}}$$

Para a determinação de  $p_{max}$  → Métodos Numéricos.

Usando-se

$$f(p_{max}) = \frac{15.000}{p_{max}} \left[ 1 + \left( \frac{p_{max}}{10} - 1 \right) e^{-1,2 \times 10^{-4} p_{max}} \right] - 1, 0$$

Logo, com chutes iniciais 60.000 e 70.000 células/litro e com tolerância de 7 unidades.

$$f(60.000) = 0,36969223$$

$$f(70.000) = -0,4484614$$

Método da Bissecção					
Iteração	$x_i$	$f(x_i)$	$a_i$	$b_i$	$ b_i - a_i $
1	65000,00	-0,15472270	60000,00	65000,00	5000,000000
2	62500,00	0,06949397	62500,00	65000,00	2500,000000
3	63750,00	-0,05075156	62500,00	63750,00	1250,000000
4	63125,00	0,00718261	63125,00	63750,00	625,000000
5	63437,50	-0,02231143	63125,00	63437,50	312,500000
6	63281,25	-0,00769861	63125,00	63281,25	156,250000
7	63203,13	-0,00029186	63125,00	63203,13	78,125000
8	63164,06	0,00343687	63164,06	63203,13	39,062500
9	63183,59	0,00157038	63183,59	63203,13	19,531250
10	63193,36	0,00063873	63193,36	63203,13	9,765625
11	63198,24	0,00017330	63198,24	63203,13	4,882813

Método da Falsa Posição					
Iteração	$x_i$	$f(x_i)$	$a_i$	$b_i$	$ x_{i+1} - x_i $
1	64518,62	-0,11645870	60000	64518,62	4518,617000
2	63436,17	-0,02218834	60000	63436,17	1082,445000
3	63241,62	-0,00394959	60000	63241,62	194,554700
4	63207,35	-0,00069430	60000	63207,35	34,265630
5	63201,34	-0,00012183	60000	63201,34	6,011719

Método da Secante		
Iteração	$x_{i+1}$	$ x_{i+1} - x_i $
1	64518,62	5481,383000
2	62595,88	1922,742000
3	63247,04	651,168000
4	63201,70	45,343750
5	63200,05	1,644531

Resumindo, temos:



Método	$p_{max}$	Iterações
Bissecção	63.198	11
Falsa Posição	63.201	5
Secante	63.200	5

Logo,  $p_{max} = 63.200$

Substituindo-se novamente

$$p(90) = \frac{63.200}{1 + \left(\frac{63.200}{10} - 1\right) e^{-2 \times 10^{-6} (63.200)(90)}} = 58930 \text{ células/litro} > 40.000$$

7. Deseja-se encontrar uma aproximação da única raiz real da função  $f(x) = x^3 + 3x - 1$ , através do Método Iterativo Linear. Verificou-se que a raiz da equação  $f(x) = 0$  está no intervalo  $[0, 1]$ .

(a) Analise a convergência da sequência  $\{x_k\}$ , gerada pela seguinte fórmula de recorrência, proposta para resolver o problema:

$$\begin{cases} x_{k+1} = \frac{1 - x_k^3}{3}, & k \geq 0 \\ x_0 = 0,25 \end{cases}$$

(b) Efetue três iterações com a fórmula de recorrência do item anterior.

**Solução:**

$$(a) \quad x_{i+1} = \frac{1 - x_i^3}{3}$$

$$g(x) = \frac{1 - x^3}{3} \longrightarrow g'(x) = -x^2$$

$$|g'(x)| < 1$$

$$|-x^2| < 1$$

$$|x^2| < 1$$

$$x^2 < 1$$

$$-1 < x < 1$$

Como o zero pertence ao intervalo  $[0, 1]$  e a aproximação inicial pertence a  $(-1, 1)$ , temos certeza que a sequência gerada convergirá para o zero de  $f(x)$ .

(b)

$$x_1 = g(x_0) = \frac{1 - x_0^3}{3} = \frac{1 - (0,25)^3}{3} = 0,328125000$$

$$x_2 = g(x_1) = \frac{1 - x_1^3}{3} = \frac{1 - (0,328125)^3}{3} = 0,321557363$$

$$x_3 = g(x_2) = \frac{1 - x_2^3}{3} = \frac{1 - (0,321557363)^3}{3} = 0,322250415$$

8. A força e o momento em uma seção  $D$  de uma viga são chamados *força cortante* e *momento fletor*. Seus valores na viga são obtidos no caso de vigas isostáticas pelas equações de equilíbrio.

Para uma viga isostática, submetida a um carregamento parabólico, tem-se a seguinte equação para o esforço cortante:

$$V_D = \frac{qL}{12} \left\{ 3 - 12\frac{x}{L} + 12\left(\frac{x}{L}\right)^2 - 4\left(\frac{x}{L}\right)^3 \right\}$$

Os pontos de esforço cortante nulo correspondem aos de momento fletor máximo, que são usados no dimensionamento da viga. Calcular estes pontos para  $q = 1000 \text{ N}$ ;  $L = 1 \text{ m}$ .

**Solução:**

Momento Máximo  $\rightarrow V_D = 0$

Logo,

$$\frac{qL}{12} \left\{ 3 - 12\frac{x}{L} + 12\left(\frac{x}{L}\right)^2 - 4\left(\frac{x}{L}\right)^3 \right\} = 0$$

ou

$$3 - 12\frac{x}{L} + 12\left(\frac{x}{L}\right)^2 - 4\left(\frac{x}{L}\right)^3 = 0$$

para  $L = 1$  m.

$$3 - 12x + 12x^2 - 4x^3 = 0$$

Usando o método de Newton-Raphson, com  $x_0 = 0,1$  e  $\epsilon = 0,001$  temos,

Iteração	$x_{i+1}$	$ x_{i+1} - x_i $
1	0,2971193	0,1971194000
2	0,3627360	0,0656166700
3	0,3699561	0,0072200890
4	0,3700395	0,0000833571

$$\bar{x} = 0,370039375$$

E as outras raízes são imaginárias.

9. Ache um zero da função  $p(x) = x^3 - 5x^2 + 17x + 21$  no intervalo  $[-1, 0]$ , usando o Método da Bissecção, com uma tolerância  $\epsilon = 1 \times 10^{-4}$ .

**Solução:**

Iteração	$x = (a + b)/2$	$f(x)$	$a$	$b$	$ b - a $
1	-0,5000000	11,12500000	-1,0000000	-0,5000000	0,50000000
2	-0,7500000	5,01562500	-1,0000000	-0,7500000	0,25000000
3	-0,8750000	1,62695300	-1,0000000	-0,8750000	0,12500000
4	-0,9375000	-0,15600590	-0,9375000	-0,8750000	0,06250000
5	-0,9062500	0,74301150	-0,9375000	-0,9062500	0,03125000
6	-0,9218750	0,29539870	-0,9375000	-0,9218750	0,01562500
7	-0,9296875	0,07017183	-0,9375000	-0,9296875	0,00781250
8	-0,9335938	-0,04279798	-0,9335938	-0,9296875	0,00390625
9	-0,9316406	0,01371666	-0,9335938	-0,9316406	0,00195313
10	-0,9326172	-0,01453322	-0,9326172	-0,9316406	0,00097656
11	-0,9321289	-0,00040642	-0,9321289	-0,9316406	0,00048828
12	-0,9318848	0,00665558	-0,9321289	-0,9318848	0,00024414
13	-0,9320068	0,00312470	-0,9321289	-0,9320068	0,00012207
14	-0,9320679	0,00135917	-0,9321289	-0,9320679	0,00006104

Assim, podemos dizer que o valor aproximado para o zero é  $-0,9320679$ .

10. Repita o exercício anterior usando o Método da Falsa Posição.

**Solução:**

Iteração	$x_i$	$f(x_i)$	$a$	$b$
1	-0,9130435	0,54886110	-1,0000	-0,91304350
2	-0,9317684	0,01002247	-1,0000	-0,93176840
3	-0,9321086	0,00018154	-1,0000	-0,93210860
4	-0,9321147	0,00000394	-1,0000	-0,93211470

A solução é  $\bar{x} = -0,93211470$ .

11. Ache um zero da função  $f(x) = x - 2\text{sen } x$  no intervalo  $[\frac{\pi}{2}, \frac{2\pi}{3}]$  com  $\epsilon = 1 \times 10^{-4}$ . Mostre que a função de iteração escolhida satisfaz as condições para a convergência.

**Solução:**

Podemos escolher como função de Iteração  $g(x) = 2\text{sen } x$ . Portanto devemos verificar as condições para a convergência da sequência gerada por esta função de iteração.

$$|g'(x)| < 1 \quad ???$$

$$g'(x) = 2\cos x$$

$$|2\cos x| < 1 \quad ???$$

$$|\cos x| < \frac{1}{2}$$

$$-\frac{1}{2} < \cos x < \frac{1}{2}$$

que sempre satisfeita no intervalo considerado, assim, a sequência gerada convergirá para o zero da função no intervalo. Efetuando-se as iterações temos, com  $x_0 = 2 \in [\frac{\pi}{2}, \frac{2\pi}{3}]$ :

Iteração	$x_{i+1}$
1	1,81859500
2	1,93891000
3	1,86601600
4	1,91347700
5	1,88371500
6	1,90287800
7	1,89073100
8	1,89851200
9	1,89356000
10	1,89672500
11	1,89470800
12	1,89599500
13	1,89517400
14	1,89569800
15	1,89536400
16	1,89557700
17	1,89544100
18	1,89552800

12. Resolva a equação  $xe^{-x} - 0,2 = 0$  pelo Método de Newton-Raphson, com  $x_0 = 0$  e  $\epsilon = 1 \times 10^{-6}$ .

**Solução:**

Método de Newton-Raphson:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

$$x_{i+1} = x_i - \frac{x_i e^{-x_i} - 0,2}{(1 - x_i)e^{-x_i}}$$

$$x_0 = 0$$

$$|x_{i+1} - x_i| < \epsilon = 1 \times 10^{-6}$$

Iteração	$x_i$
1	0,200000000
2	0,255350689
3	0,257632981
4	0,259168327
5	0,259171101
6	0,259171101

A aproximação para o zero é  $\bar{x} = 0,259171101$

13. Determine as raízes reais de

$$f(x) = -2,1 + 6,21x - 3,9x^2 + 0,667x^3$$

usando o Método da Bisseção para encontrar a menor raiz. Use como estimativa inicial  $a_0 = 0,4$  e  $b_0 = 0,6$  e faça iterações até o erro seja menor que 4%.

Logo,

$$\frac{|x_{i+1} - x_i|}{|x_{i+1}|} < 0,04$$

Iteração	$x$	$f(x)$	$a_i$	$b_i$	$\frac{ x_{i+1} - x_i }{ x_{i+1} }$
1	0,5000000	0,11337510	0,4000000	0,5000000	
2	0,4500000	-0,03446957	0,4500000	0,5000000	0,11111111
3	0,4750000	0,04129621	0,4500000	0,4750000	0,05263157
4	0,4625000	0,00387805	0,4500000	0,4625000	0,02702703

14. Repita o exercício anterior usando o Método de Newton-Raphson, com  $x_0 = 0,5$ .

Iteração	$x_{i+1}$	$\frac{ x_{i+1} - x_i }{ x_{i+1} }$
1	0,4596566	0,087768565
2	0,4612195	0,003388625

15. Determine a raiz real de

$$f(x) = x^3 - 100$$

usando o Método da Falsa Posição com tolerância de 0,1 %.

Usando como estimativas iniciais  $[a, b] = [4, 5]$ .

Iteração	$x_i$	$f(x_i)$	$a_i$	$b_i$	$\frac{ x_{i+1} - x_i }{ x_{i+1} }$
1	4,590164	-3,28707400	4,590164	5,0	
2	4,637788	-0,24543920	4,637788	5,0	0,01026881
3	4,641310	-0,01803770	4,641310	5,0	0,00075871
4	4,641568	-0,00133465	4,641568	5,0	0,00005568
5	4,641587	-0,00010188	4,641587	5,0	0,00000410

16. Repita o exercício anterior usando o Método da Secante.

Iteração	$x_{i+1}$	$ x_{i+1} - x_i $	$\frac{ x_{i+1} - x_i }{ x_{i+1} }$
1	4,59016	0,4098363	0,0892858
2	4,63779	0,0476246	0,0102688
3	4,64163	0,0038428	0,0008278
4	4,64159	0,0000424	0,0000091

17. Repita o exercício anterior usando o Método da Bissecção com  $\epsilon = 0,05$ .

**Solução:**

Iteração	$x = (a + b)/2$	$f(x)$	$a$	$b$	$ b - a $
1	4,50000000	-8,87500000	4,50000000	5,00000000	0,50000000
2	4,75000000	7,17187500	4,50000000	4,75000000	0,25000000
3	4,62500000	-1,06835900	4,62500000	4,75000000	0,12500000
4	4,68750000	2,99682600	4,62500000	4,68750000	0,06250000
5	4,65625000	0,95059200	4,62500000	4,65625000	0,03125000

18. Ache uma raiz da função  $f(x) = x^3 - 9x + 3$  no intervalo  $[0, 1]$ , usando o Método da Bissecção com uma tolerância de  $\epsilon = 0,01$ .

Iteração	$x_i$	$f(x_i)$	$a_i$	$b_i$	$ b_i - a_i $
1	0,5000000	-1.37500000	0,0000000	0,50000	0,5000000
2	0,2500000	0.76562500	0,2500000	0,50000	0,2500000
3	0,3750000	-0.32226560	0,2500000	0,37500	0,1250000
4	0,3125000	0.21801760	0,3125000	0,37500	0,0625000
5	0,3437500	-0.05313110	0,3125000	0,34375	0,0312500
6	0,3281250	0.08220291	0,3281250	0,34375	0,0156250
7	0,3359375	0.01447439	0,3359375	0,34375	0,0078125

19. A equação  $x^2 - x = 0$  possui raízes  $x = 0$  e  $x = 1$ . Qual das raízes teremos certeza de encontrar, com a iteração  $x_{i+1} = (x_i)^2$  e o valor inicial  $x_0 = 1/2$ ? Justifique sua resposta. Faça três iterações usando a função acima e  $\epsilon = 0,01$ .

**Solução:**

Queremos determinar se a seqüência gerada por esta função de iteração convergirá?

$$g(x) = x^2$$

$$|g'(x)| < 1$$

$$|2x| < 1$$

$$-1 < 2x < 1$$

$$-\frac{1}{2} < x < \frac{1}{2}$$

a raiz 0 pertence ao intervalo  $(-1/2, 1/2)$  assim como a aproximação inicial  $x_0 = 1/2$ .

$$x_0 = 0,5$$

$$x_1 = 0,25$$

$$x_2 = 0,0625$$

$$x_3 = 0,00390625$$

20. O potencial a que uma partícula está submetida é  $V(x) = \frac{1}{x^2} - \frac{2}{x}$ ,  $x > 0$ . O ponto de equilíbrio corresponde a um mínimo do potencial. Realize 3 iterações de Newton-Raphson com o intuito de encontrar tal valor ( $x_0 = 1,5$ ).
21. Use o Método de Newton-Raphson para resolver a equação  $f(x) = e^x - 3x = 0$ , use como estimativa inicial  $x_0 = 0$  e tome  $\epsilon = 0,02$ .

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

$$x_{i+1} = x_i - \frac{e^{x_i} - 3x_i}{e^{x_i} - 3}$$

Iteração	$x_{i+1}$	$ x_{i+1} - x_i $
1	0,5000000	0,5000000
2	0,6100597	0,1100597
3	0,6189968	0,00893712

22. Repita o exercício anterior usando o Método da Secante.

Usando como aproximações iniciais  $x_0 = 0$  e  $x_1 = 1$

Iteração	$x_{i+1}$	$ x_{i+1} - x_i $
2	0,78020	0,2197973
3	0,49668	0,2835242
4	0,63595	0,1392737
5	0,62056	0,0153918

23. Em Vibrações de corpos elásticos é comum a ocorrência da equação  $(4 - 3x^2) \sin x = 4x \cos x$ . Determinar as três menores raízes positivas.

*Resp:* 2,56334163; 6,058670084; 9,279861114.

24. Em cada caso abaixo determinar o valor da única raiz fisicamente possível:

- (a) uma bóia esférica de raio  $R$  e densidade específica  $\rho$ , ao flutuar na água, afunda de uma quantidade  $x$ , dada por  $x^3 + 2Rx^2 - 4\rho R^3 = 0$ . Achar o afundamento quando  $R = 3$  e o material é cortiça ( $\rho = 0,25$ );
- (b) tem-se duas cargas elétricas  $A$  e  $B$  à distância de  $2 m$  uma da outra. As duas têm o mesmo sinal. Coloca-se uma terceira carga  $C$ , de sinal oposto entre  $A$  e  $B$ . A distância  $x$ , com que  $C$  ficará em equilíbrio de  $A$  e  $C$ , é dada por:  $x^4 - 4x^3 + 8x^2 - 32x + 32 = 0$ . Determinar um valor  $x$ .

*Resp:* a) 1,854101966; b) 1,212169077

25. Método de Newton-Raphson Modificado:

Existe uma modificação no Método de Newton-Raphson na qual a função de iteração é dada por:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_0)}$$

onde  $x_0$  é a aproximação inicial e é tal que  $f'(x_0) \neq 0$ .

- (a) Com o auxílio de um gráfico, escreva a interpretação geométrica deste método;
- (b) Use este método para achar um zero de

$$f(x) = e^x - 4x^2$$

no intervalo  $(0, 1)$ , com  $\epsilon = 10^{-3}$ .



**Solução:**

(a) Gráfico

(b) Usando como aproximação inicial  $x_0 = 0.5$ . Temos

Iteração	$x_i$	$ x_{i+1} - x_i $
1	0,7759015000	0,2759015000
2	0,6757250000	0,1001765000
3	0,7348589000	0,0591338300
4	0,7030113000	0,0318476000
5	0,7212679000	0,0182566000
6	0,7111161000	0,0101518000
7	0,7168654000	0,0057493450
8	0,7136414000	0,0032240150
9	0,7154596000	0,0018182400
10	0,7144374000	0,0010222200
11	0,7150131000	0,0005757213

O valor aproximado para o zero é 0,71501310.

26. Determine o valor de  $\pi$  com erro menor que  $10^{-4}$ . (Sugestão:  $\pi$  é um zero de  $\sin x = 0$ , use  $x_0 = 3$  e o método de Newton-Raphson.)

27. Resolva a equação abaixo pelo método da Bisseção com uma tolerância  $\epsilon = 0,005$  e  $[a_0, b_0] =$

$[0,5;1,0]$ . Estime, a priori, o número de iterações necessário.

$$e^{-x} - x = 0$$

28. Deduza o Método de Newton Raphson a partir de sua interpretação geométrica.

29. Demonstre que se

$$x_{i+1} = x_i(2 - ax_i)$$

converge, converge então para  $1/a$  quando  $i \rightarrow \infty$ . Este é um método iterativo para fazer divisão ou determinar inversos desde que exista uma estimativa inicial razoavelmente boa. Que limites deve  $x_0$  satisfazer para assegurar convergência?

**Solução:**

Se converge teremos,

$$x = g(x)$$

$$x = x(2 - ax)$$

$$1 = 2 - ax$$

$$x = \frac{1}{a}$$

Verificaremos agora o intervalo para a convergência:

$$|g'(x)| < 1$$

$$-1 < 2 - 2ax < 1$$

$$-3 < -2ax < -1$$

$$3 > 2ax > 1$$

$$\frac{1}{2a} < x < \frac{3}{2a}$$

30. Calcule  $1/13$  usando uma calculadora fictícia que só soma, subtrai e multiplica.

(Sugestão: Observe que  $1/a$  é solução da equação  $a - 1/x = 0$  e aplique o Método de Newton-Raphson.)

**Solução:**

$$f(x) = a - \frac{1}{x}$$

$$f'(x) = \frac{1}{x^2}$$

Método de Newton-Raphson

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

substituindo

$$x_{i+1} = x_i - \frac{a - \frac{1}{x_i}}{\frac{1}{x_i^2}}$$

ou

$$x_{i+1} = -ax_i^2 + 2x_i$$

que é uma função de iteração que não envolve divisões com  $x_0 = \frac{1}{10} = 0,1$  e  $\epsilon = 10^{-15}$  temos

Iteração	$x_i$	$ x_i - x_{i-1} $
1	0,070000000000000000	$ x_1 - x_0  = 0,03000000000000000000000000000000$
2	0,076300000000000000	$ x_2 - x_1  = 0,006300000000000000000000000000$
3	0,076918030000000000	$ x_3 - x_2  = 0,00061803000000000055000000$
4	0,07692307659194830	$ x_4 - x_3  = 0,0000050465919483044800000$
5	0,07692307692307693	$ x_5 - x_4  = 0,0000000003311286245200051$
6	0,07692307692307693	$ x_6 - x_5  = 0,00000000000000000000000000$

logo,

$$\frac{1}{13} = 0,07692307692307693$$

31. O preço à vista ( $PV$ ) de uma mercadoria é R\$ 312,00, mas pode ser financiado com uma entrada ( $E$ ) de R\$ 91,00 e 12 ( $N$ ) prestações mensais ( $PM$ ) de R\$ 26,00. Calcular os juros ( $j$ ) sabendo que,

$$\frac{1 - (1 + j)^{-N}}{j} = \frac{PV - E}{PM}$$

*Resp: 5,75 %*

32. Determinar o comprimento ( $L$ ) de um cabo suspenso em dois pontos de mesma altura e distantes 400 m ( $2x$ ), com uma flecha ( $f$ ) de 100 m, sabendo que

$$L = 2a \operatorname{senh} \frac{x}{a}$$

sendo  $a$  a raiz da equação

$$a \left( \cosh \frac{x}{a} - 1 \right) - f = 0$$

*Resp: 460,316 m*

33. Mostrar que a equação

$$x_{i+1} = \frac{1}{p} \left( (p-1)x_i + \frac{a}{x_i^{p-1}} \right)$$

pode ser usada para calcular  $\sqrt[p]{a}$ ,  $a \geq 0$ .

34. Calcular  $\sqrt[3]{8}$ .

35. Responda as seguintes questões, justificando integralmente suas respostas.

- a. Qual a principal vantagem do método de Newton-Raphson em relação aos outros que estudamos? E do método da Bissecção?
- b. Diga as desvantagens dos métodos citados no item acima.

36. Mostre que a função de iteração abaixo, que é usada para determinar a raiz quadrada de um número  $Q$ , é um caso especial do método de Newton-Raphson.

$$x_{i+1} = \frac{1}{2} \left( x_i + \frac{Q}{x_i} \right)$$

**Solução:**

Com  $f(x) = x^2 - Q$ , é claro que  $f(x) = 0$  é equivalente a encontrar um raiz quadrada de  $Q$ . Como  $f'(x) = 2x$ , a expressão para o método de Newton-Raphson fornece:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{x_i^2 - Q}{2x_i} = \frac{1}{2} \left( x_i + \frac{Q}{x_i} \right)$$

37. Aplique o método acima para determinar a raiz quadrada de 2.

**Solução:**

Efetuada as iterações com  $x_0 = 1$  temos:

$i$	$x_i$
1	1,500000000
2	1,416666667
3	1,414215686
4	1,414216562
5	1,414216562

38. Considere a função  $f$  definida por

$$f(x) = x^3 - x - 1$$

Sabe-se que a equação  $f(x) = 0$  tem apenas uma raiz no intervalo  $[1, 2]$ .

- (a) Determine o número mínimo de iterações para que  $|\bar{x} - x_{i+1}| < 10^{-3}$ , onde  $x_{i+1}$  é o valor de uma aproximação de  $\bar{x}$  calculada pelo Método da Bissecção com  $I_0 = [1, 2]$ .

- (b) Calcule três aproximações para  $\bar{x}$  pelo Método da Bissecção com  $I_0 = [1, 2]$  e utilize dois critérios para estabelecer o erro das aproximações encontradas (trabalhe com 3 dígitos decimais).
- (c) Se aplicarmos o Método da Falsa Posição para calcular aproximações de  $\bar{x}$  com  $I_0 = [1, 2]$ , podemos garantir que  $I_{i+1} = [x_{i+1}, b]$  para todo  $i$ ,  $i = 0, 1, 2, \dots$ ?

39. Considere a função  $f$  definida por

$$f(x) = x^3 - x - 1$$

- (a) Sabendo que a equação  $f(x) = 0$  tem apenas uma raiz real  $\bar{x}$ , verifique, usando argumentos gráficos ou teóricos que  $\bar{x} \in [0; 0, 5]$ .
- (b) Pretende-se utilizar o Método Iterativo Linear (MIL) para calcular aproximações de  $\bar{x}$  com a função de iteração  $g(x)$  definida por

$$g(x) = \frac{1 + x - x^3}{4}$$

Verifique que  $g(x)$  é realmente uma função de iteração.

- (c) Podemos afirmar que: se a aproximação inicial  $x_0$  de  $\bar{x}$  pertencer ao intervalo  $[0; 0, 5]$ , então, o MIL definido no item anterior converge?
- (d) Faça três iterações usando esta função de iteração com  $x_0 = 0, 25$ . Use um critério para estabelecer o erro das aproximações encontradas (trabalhe com 3 dígitos decimais).
40. Ache um zero da função  $f(x) = 8x^3 - 6x^2 - 2x - 1$  no intervalo  $[1, 2]$  com tolerância  $\epsilon = 0, 0001$ , usando os métodos da Bissecção e da Falsa Posição.

41. Deduza uma fórmula de iteração de Newton-Raphson para determinação da raiz cúbica de um número positivo  $C$ . Ache  $\sqrt[3]{2}$  usando esta fórmula. **Solução:**

$$x^3 = C \implies f(x) = x^3 - C$$

Newton-Raphson

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{x_i^3 - C}{3x_i^2} = \frac{2x_i^3 + C}{3x_i^2}$$

para  $C = 2$ ,

$$x_{i+1} = \frac{2x_i^3 + 2}{3x_i^2}$$

com  $x_0 = 1, 5$

$$x_1 = 1, 94\dots$$

42. Use o método de Newton para resolver o sistema não-linear abaixo

$$\begin{cases} xy = 0 \\ x^2 + y^2 = 1 \end{cases}$$

com  $x^{(0)} = 0,5$  e  $y^{(0)} = 0,1$  e tolerância  $\epsilon = 1 \times 10^{-4}$ .

43. Repita o exercício anterior para

$$\begin{cases} 3x^2y - y^3 = 4 \\ x^2 + xy^3 = 9 \end{cases}$$

resolva para 4 estimativas iniciais

a)  $x_0 = (1, 2 ; 2, 5)$

b)  $x_0 = (-2, 0 ; 2, 5)$

c)  $x_0 = (-1, 2 ; -2, 5)$

d)  $x_0 = (2, 0 ; -2, 5)$

44. Resolva o sistema não-linear dado abaixo pelo Método de Newton-Raphson com  $x^{(0)} = 2$  e  $y^{(0)} = 2$  e tolerância  $\epsilon = 1 \times 10^{-4}$ .

$$\begin{cases} xy = 1 \\ x^2 + y^2 - 4x + 2y = 4 \end{cases}$$

45. Resolva o sistema não-linear dado abaixo pelo Método de Newton.

$$\begin{cases} x = \text{sen}(x + y) \\ y = \text{cos}(x - y) \end{cases}$$

**Solução:**

$$\vec{F}(\vec{x}) = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} x_1 - \text{sen}(x_1 + x_2) \\ x_2 - \text{cos}(x_1 - x_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}$$

Método de Newton:

$$\vec{x}^{(i+1)} = \vec{x}^{(i)} - \left( D\vec{F}(\vec{x}^{(i)}) \right)^{-1} \vec{F}(\vec{x}^{(i)})$$

ou

$$D\vec{F}(\vec{x}^{(i)}) \left[ \vec{x}^{(i+1)} - \vec{x}^{(i)} \right] = -\vec{F}(\vec{x}^{(i)})$$

mas

$$D\vec{F}(\vec{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}) & \frac{\partial f_1}{\partial x_2}(\vec{x}) \\ \frac{\partial f_2}{\partial x_1}(\vec{x}) & \frac{\partial f_2}{\partial x_2}(\vec{x}) \end{bmatrix} = \begin{bmatrix} 1 - \cos(x_1 + x_2) & -\cos(x_1 + x_2) \\ \sin(x_1 - x_2) & 1 - \sin(x_1 - x_2) \end{bmatrix}$$

E a solução é:

$$\vec{x} = (0.935082, 0.9980200)$$

Aproximação Inicial	No. de iterações
$\vec{x}^{(0)} = (0.1, 0.1)$	diverge
$\vec{x}^{(0)} = (0.5, 0.5)$	10 iterações
$\vec{x}^{(0)} = (0.9, 0.9)$	4 iterações
$\vec{x}^{(0)} = (1.0, 1.0)$	3 iterações

46. Seja o sistema de equações não lineares abaixo

$$\begin{cases} x + 2y = 3 \\ 3x^2 + y^2 = 7 \end{cases}$$

Faça duas iterações para determinar um zero desse sistema usando o método de Newton. Use  $x^{(0)} = 2$  e  $y^{(0)} = 1$  com  $\epsilon = 0,01$ .

47. Seja o sistema de equações não lineares abaixo

$$\begin{cases} 2x + y = 1 \\ x^2 + y^2 = 4 \end{cases}$$

Faça duas iterações do Método de Newton para determinar um zero desse sistema. Faça o gráfico das curvas no plano para estimar bons valores para a aproximação inicial  $x^{(0)}$  e  $y^{(0)}$ .

**Solução:**

Avaliando a estimativa inicial pelo gráfico usaremos,  $x^{(0)} = -1.4$  e  $y^{(0)} = 1.8$ . Usaremos como tolerância  $\epsilon = 10^{-5}$ . Logo, pelo Método de Newton

$$\left[ DF(z^{(i)}) \right] \Delta z = -F(z^{(i)})$$

onde

$$\left[ DF(z^{(i)}) \right] = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}$$

$$F(z) = \begin{bmatrix} 2x + y - 1 \\ x^2 + y^2 - 4 \end{bmatrix}$$

e

$$z = \begin{bmatrix} x \\ y \end{bmatrix}$$

Resultado das Aproximações

Iteração	$x^{(i)}$	$y^{(i)}$
1	-5.600000E-01	2.120000
2	-4.758333E-01	1.951667
3	-4.717892E-01	1.943578
4	-4.717798E-01	1.943560
5	-4.717798E-01	1.943560

#### 48. Método de Newton-Raphson Modificado para Sistemas Não Lineares

Existe uma modificação no Método de Newton-Raphson para Sistemas Não Lineares na qual a função de iteração é dada por:

$$\vec{x}^{(i+1)} = \vec{x}^{(i)} - \left( D\vec{F}(\vec{x}^{(0)}) \right)^{-1} \vec{F}(\vec{x}^{(i)})$$

ou

$$D\vec{F}(\vec{x}^{(0)}) \left[ \vec{x}^{(i+1)} - \vec{x}^{(i)} \right] = -\vec{F}(\vec{x}^{(i)})$$

onde  $\vec{x}^{(0)}$  é a aproximação inicial.

Use este método para achar um zero de

$$\begin{cases} x = \text{sen}(x + y) \\ y = \text{cos}(x - y) \end{cases}$$

com  $\vec{x}^{(0)} = (1, 0; 1, 0)$ . Faça no máximo 3 iterações.



**Solução:**

$$\vec{F}(\vec{x}) = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} x_1 - \text{sen}(x_1 + x_2) \\ x_2 - \text{cos}(x_1 - x_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}$$

mas

$$D\vec{F}(\vec{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}) & \frac{\partial f_1}{\partial x_2}(\vec{x}) \\ \frac{\partial f_2}{\partial x_1}(\vec{x}) & \frac{\partial f_2}{\partial x_2}(\vec{x}) \end{bmatrix} = \begin{bmatrix} 1 - \text{cos}(x_1 + x_2) & -\text{cos}(x_1 + x_2) \\ \text{sen}(x_1 - x_2) & 1 - \text{sen}(x_1 - x_2) \end{bmatrix}$$

em  $\vec{x} = \vec{x}^{(0)}$

$$D\vec{F}(\vec{x}^{(0)}) = \begin{bmatrix} 1 - \text{cos}(2, 0) & -\text{cos}(2, 0) \\ \text{sen}(0, 0) & 1 - \text{sen}(0, 0) \end{bmatrix} = \begin{bmatrix} 1,416146837 & 0,416146837 \\ 0,0 & 1,0 \end{bmatrix}$$

Como neste método a matriz permanece fixa em todas as iterações temos:

**Primeira iteração:**

$$D\vec{F}(\vec{x}^{(0)}) [\vec{x}^{(1)} - \vec{x}^{(0)}] = -\vec{F}(\vec{x}^{(0)})$$

$$\begin{bmatrix} 1 - \text{cos}(2, 0) & -\text{cos}(2, 0) \\ \text{sen}(0, 0) & 1 - \text{sen}(0, 0) \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = - \begin{bmatrix} 1, 0 - \text{sen}(1, 0 + 1, 0) \\ 1, 0 - \text{cos}(1, 0 - 1, 0) \end{bmatrix}$$

$$\begin{bmatrix} 1,416146837 & 0,416146837 \\ 0,0 & 1,0 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = - \begin{bmatrix} 0,090702573 \\ 0,0 \end{bmatrix}$$

Resolvendo o sistema obtemos

$$\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} -0,064048848 \\ 0,0 \end{bmatrix}$$

e

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \Delta \vec{x} = \begin{bmatrix} 1,0 \\ 1,0 \end{bmatrix} + \begin{bmatrix} -0,064048848 \\ 0,0 \end{bmatrix} = \begin{bmatrix} 0,9359512 \\ 1,0000000 \end{bmatrix}$$

**Segunda iteração:**

$$D\vec{F}(\vec{x}^{(0)}) [\vec{x}^{(2)} - \vec{x}^{(1)}] = -\vec{F}(\vec{x}^{(1)})$$

$$\begin{bmatrix} 1 - \cos(2,0) & -\cos(2,0) \\ \sin(0,0) & 1 - \sin(0,0) \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = - \begin{bmatrix} 0,9359512 - \sin(0,9359512 + 1,0) \\ 1,0 - \cos(0,9359512 - 1,0) \end{bmatrix}$$

$$\begin{bmatrix} 1,416146837 & 0,416146837 \\ 0,0 & 1,0 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = - \begin{bmatrix} 0,001882732 \\ 0,002050423 \end{bmatrix}$$

Resolvendo o sistema obtemos

$$\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} -0,000726941 \\ -0,002050423 \end{bmatrix}$$

e

$$\vec{x}^{(2)} = \vec{x}^{(1)} + \Delta \vec{x} = \begin{bmatrix} 0,9359512 \\ 1,0 \end{bmatrix} + \begin{bmatrix} -0,000726941 \\ -0,002050423 \end{bmatrix} = \begin{bmatrix} 0,9352242 \\ 0,9979496 \end{bmatrix}$$

Terceira iteração:

$$D\vec{F}(\vec{x}^{(0)}) [\vec{x}^{(3)} - \vec{x}^{(2)}] = -\vec{F}(\vec{x}^{(2)})$$

$$\begin{bmatrix} 1 - \cos(2, 0) & -\cos(2, 0) \\ \sin(0, 0) & 1 - \sin(0, 0) \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = - \begin{bmatrix} 0,9352242 - \sin(0,9352242 + 0,9979496) \\ 0,9979496 - \cos(0,9352242 - 0,9979496) \end{bmatrix}$$

$$\begin{bmatrix} 1,416146837 & 0,416146837 \\ 0,0 & 1,0 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = - \begin{bmatrix} 0,000167623 \\ -0,000083811 \end{bmatrix}$$

Resolvendo o sistema obtemos

$$\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} -0,000142994 \\ 0,000083811 \end{bmatrix}$$

e

$$\vec{x}^{(3)} = \vec{x}^{(2)} + \Delta \vec{x} = \begin{bmatrix} 0,9352242 \\ 0,9979496 \end{bmatrix} + \begin{bmatrix} -0,000142994 \\ 0,000083811 \end{bmatrix} = \begin{bmatrix} 0,9350812 \\ 0,9980334 \end{bmatrix}$$

49. Resolva o sistema de equações não lineares abaixo pelo método de Newton

$$\begin{cases} x = x^2 + y^2 \\ y = x^2 - y^2 \end{cases}$$

com  $\mathbf{x}^{(0)} = (0, 8; 0, 4)$  e  $\epsilon = 1 \times 10^{-3}$ .

Resp:  $\mathbf{x} \approx (0, 772; 0, 420)$

50. Resolva o sistema de equações não lineares abaixo usando o método de Newton

$$\begin{cases} y = \sqrt{x} \\ x^2 + y^2 = 1 \end{cases}$$

use como estimativa inicial  $\mathbf{x}^{(0)} = (0, 8; 0, 8)$ , com uma tolerância  $\epsilon = 1 \times 10^{-4}$ .

## Chapter 4

# Interpolação

Considere o seguinte problema: Seja uma função  $f(x)$  conhecida somente em alguns pontos de abcissas  $x_i$ ,  $i = 0, 1, 2, 3, \dots, n$ ;  $x_i \neq x_k$  para  $i \neq k$ . Seja  $f(x_i)$  os valores conhecidos da função nos pontos  $x_i$  e queremos determinar uma função  $p(x)$  que aproxime  $f(x)$ , isto é,  $f(x) \approx p(x)$ . Resumindo, conhecemos alguns pontos discretos de  $f(x)$ ,  $(x_i, f(x_i))$ , chamados pontos de suporte, chamamos os  $x_i$  de abcissas de suporte e as  $f(x_i)$  de ordenadas de suporte.

O problema geral de interpolação consiste em dada uma família de funções  $\Phi(x; a_0, a_1, \dots, a_n)$  encontrar os parâmetros  $a_0, a_1, \dots, a_n$  tais que para os  $(n + 1)$  pontos  $(x_i, f(x_i))$  dados, tenha-se

$$\Phi(x; a_0, a_1, \dots, a_n) = f(x_i), \quad i = 0, 1, 2, \dots, n.$$

O problema acima é chamado um problema de interpolação linear se  $\Phi(x; a_0, a_1, \dots, a_n)$  depende linearmente dos parâmetros  $a_0, a_1, \dots, a_n$ , ou seja,

$$\Phi(x; a_0, a_1, \dots, a_n) = a_0\Phi_0(x) + a_1\Phi_1(x) + a_2\Phi_2(x) + \dots + a_n\Phi_n(x)$$

Nesta classe de problema encontra-se a clássica interpolação polinomial, onde

$$\Phi(x; a_0, a_1, \dots, a_n) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

assim como, a interpolação trigonométrica, onde

$$\Phi(x; a_0, a_1, \dots, a_n) = a_0 + a_1e^{xi} + a_2e^{2xi} + \dots + a_n e^{nxi} \quad (i^2 = -1).$$

Encontra-se também na classe de interpolação linear a interpolação por splines (pedaços de polinômios).

Como exemplos de interpolação não-linear podemos citar, a interpolação racional, onde

$$\Phi(x; a_0, a_1, \dots, a_n, b_0, b_1, \dots, b_m) = \frac{a_0 + a_1x + a_2x^2 + \dots + a_nx^n}{b_0 + b_1x + b_2x^2 + \dots + b_mx^m}$$

e a interpolação exponencial, onde

$$\Phi(x; a_0, a_1, \dots, a_n, \lambda_0, \lambda_1, \dots, \lambda_n) = a_0e^{\lambda_0x} + a_1e^{\lambda_1x} + \dots + a_n e^{\lambda_nx}$$

## 4.1 Interpolação Polinomial

Vejamos a interpolação polinomial. Temos  $(n + 1)$  pontos de suporte  $(x_i, f(x_i))$ ,  $i = 0, 1, 2, \dots, n$ .

**Teorema 4.1.1** *Para  $(n + 1)$  pontos de suporte  $(x_i, f(x_i))$ ,  $i = 0, 1, 2, \dots, n$ ,  $x_i \neq x_k$  para  $i \neq k$  existe um único polinômio  $p(x)$  de grau menor ou igual a  $n$ , tal que:*

$$p(x_i) = f(x_i) \quad i = 0, 1, 2, \dots, n.$$

**Demonstração:**

**Unicidade:**

Sejam dois polinômios  $p_1(x)$  e  $p_2(x)$  de grau menor ou igual a  $n$  com

$$p_1(x_i) = p_2(x_i) = f(x_i), \quad i = 0, 1, 2, \dots, n$$

o polinômio  $p(x) = p_1(x) - p_2(x)$  tem no máximo grau  $n$  (porque é a diferença entre dois polinômios de grau  $n$ ), e tem  $(n + 1)$  zeros diferentes, a saber  $x_i$ ,  $i = 0, 1, 2, \dots, n$ . Logo,  $p(x)$  é necessariamente nulo.

Então,  $p(x) = 0$  e portanto,  $p_1(x) = p_2(x)$ .

**Existência:**

Construíremos o polinômio interpolador  $p(x)$  explicitamente com a ajuda dos polinômios  $L_i(x)$ ,  $i = 0, 1, 2, \dots, n$ , de grau menor ou igual a  $n$ , para os quais

$$L_i(x_k) = \begin{cases} 1 & i = k \\ 0 & i \neq k. \end{cases}$$

Os seguintes polinômios, chamados polinômios de Lagrange, satisfazem as condições acima:

$$L_i(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}$$

$$L_i(x) = \frac{(x - x_0)}{(x_i - x_0)} \frac{(x - x_1)}{(x_i - x_1)} \dots \frac{(x - x_{i-1})}{(x_i - x_{i-1})} \frac{(x - x_{i+1})}{(x_i - x_{i+1})} \dots \frac{(x - x_n)}{(x_i - x_n)}$$

$$L_i(x) = \prod_{\substack{k=0 \\ k \neq i}}^n \frac{(x - x_k)}{(x_i - x_k)}$$

A solução  $p(x)$  do problema de interpolação pode ser agora expressa diretamente em termos dos polinômios  $L_i(x)$ .

$$p(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x) + \dots + f(x_n)L_n(x) = \sum_{i=0}^n f(x_i) \prod_{\substack{k=0 \\ k \neq i}}^n \frac{(x - x_k)}{(x_i - x_k)}$$

**Exemplo:** Dados os seguintes pontos de suporte encontrar o polinômio interpolador.

$x$	-1	0	2
$f(x)$	4	1	-1

Pela forma de Lagrange temos,

$$p(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x)$$

onde

$$L_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} = \frac{(x - 0)(x - 2)}{(-1 - 0)(-1 - 2)} = \frac{x(x - 2)}{3}$$

$$L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} = \frac{(x + 1)(x - 2)}{(0 + 1)(0 - 2)} = \frac{(x + 1)(x - 2)}{-2}$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} = \frac{(x + 1)(x - 0)}{(2 + 1)(2 - 0)} = \frac{x(x + 1)}{6}$$

e o polinômio interpolador

$$p(x) = 4\frac{x(x - 2)}{3} + 1\frac{(x + 1)(x - 2)}{-2} + (-1)\frac{x(x + 1)}{6}$$

Podemos obter o polinômio interpolador de outra forma. Sabemos que o polinômio interpolador existe, é único e tem grau menor ou igual a  $n$ , logo podemos escrevê-lo como

$$p(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

mas temos que

$$p(x_i) = f(x_i), \quad x_i \neq x_k \text{ para } i \neq k \quad i = 0, 1, 2, \dots, n$$

assim

$$\text{para } i = 0 \implies a_0 + a_1x_0 + a_2x_0^2 + \dots + a_nx_0^n = f(x_0)$$

$$\text{para } i = 1 \implies a_0 + a_1x_1 + a_2x_1^2 + \dots + a_nx_1^n = f(x_1)$$

$$\text{para } i = 2 \implies a_0 + a_1x_2 + a_2x_2^2 + \dots + a_nx_2^n = f(x_2)$$

$$\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots$$

$$\text{para } i = n \implies a_0 + a_1x_n + a_2x_n^2 + \dots + a_nx_n^n = f(x_n)$$

logo temos um sistema de equações algébricas lineares para determinação dos coeficientes do polinômio interpolador. Na forma matricial,

$$\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^n \\ 1 & x_3 & x_3^2 & x_3^3 & \cdots & x_3^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ f(x_3) \\ \vdots \\ f(x_n) \end{bmatrix}$$

A matriz de coeficientes do sistema é uma matriz de Vandermonde e como as abcissas  $x_i$  são distintas seu determinante é diferente de zero. Logo este sistema possui solução e ela é única. Esta forma de obtenção do polinômio interpolador é pouco usada, visto que envolve a resolução de um sistema linear. Ademais esta matriz de coeficientes tende a ser mal condicionada se duas abcissas estão próximas.

**Exemplo:** Dados os seguintes pontos de suporte encontrar o polinômio interpolador.

$$\begin{array}{c|ccc} x & -1 & 0 & 2 \\ \hline f(x) & 4 & 1 & -1 \end{array}$$

A princípio o polinômio interpolador tem grau igual a 2, logo,

$$p(x) = a_0 + a_1x + a_2x^2$$

para os pontos de suporte

$$\begin{aligned} p(x_0) = p(-1) &= a_0 + a_1x_0 + a_2x_0^2 = a_0 + a_1(-1) + a_2(-1)^2 = a_0 - a_1 + a_2 = f(x_0) = 4 \\ p(x_1) = p(0) &= a_0 + a_1x_1 + a_2x_1^2 = a_0 + a_1(0) + a_2(0)^2 = a_0 = f(x_1) = 1 \\ p(x_2) = p(2) &= a_0 + a_1x_2 + a_2x_2^2 = a_0 + a_1(2) + a_2(2)^2 = a_0 + 2a_1 + 4a_2 = f(x_2) = -1 \end{aligned}$$

resumindo temos o seguinte sistema linear

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ -1 \end{bmatrix}$$

resolvendo o sistema obtemos  $a_0 = 1$ ,  $a_1 = -7/3$  e  $a_2 = 2/3$ . Portanto

$$p(x) = 1 - \frac{7}{3}x + \frac{2}{3}x^2$$

A seguir, estudaremos o erro cometido por aproximar uma função  $f(x)$  por um polinômio interpolador.

## Erro na Interpolação Polinomial

Sendo conhecidos os pontos de suporte  $(x_i, f(x_i))$ ,  $x_i \neq x_k$ ,  $i \neq k$ ,  $i = 0, 1, 2, \dots, n$ . Desejamos agora saber como o polinômio interpolador  $p(x)$  aproxima  $f(x)$  para argumentos diferentes de  $x_i$ . Ou seja, estamos interessados em conhecer o erro  $f(x) - p(x)$ .

**Teorema 4.1.2** *Se  $f(x)$  é  $(n+1)$  vezes diferenciável, então para todo argumento  $\bar{x}$  existe um número  $\alpha$  no intervalo  $[x_0, x_n]$ , que contém  $\bar{x}$  e todas as abcissas de suporte, satisfazendo*

$$f(\bar{x}) - p(\bar{x}) = \frac{\omega(\bar{x})f^{(n+1)}(\alpha)}{(n+1)!}$$

onde  $\omega(x) = (x - x_0)(x - x_1)(x - x_2) \cdots (x - x_n)$ .

**Demonstração:** Seja  $p(x)$  o polinômio interpolador da função nos pontos  $x_i$ , e seja  $\bar{x} \neq x_i$ . Pode-se encontrar uma constante  $k$  tal que a função

$$F(x) = f(x) - p(x) - k \cdot \omega(x)$$

se anule para  $x = \bar{x}$ , isto é,

$$F(\bar{x}) = 0$$

logo  $F(x)$  tem  $(n + 2)$  zeros, a saber,  $x_i$ ,  $i = 0, 1, 2, \dots, n$  e  $\bar{x}$ . Pelo teorema de Rolle, aplicado sucessivamente,  $F'(x)$  tem no máximo  $(n + 1)$  zeros,  $F''(x)$  tem no máximo  $(n)$  zeros,  $F'''(x)$  tem no máximo  $(n - 1)$  zeros e finalmente  $F^{(n+1)}(x)$  tem um zero  $\alpha \in [x_0, x_n]$ .

Como  $p(x)$  é um polinômio de grau  $n$ , logo  $p^{(n+1)}(x) = 0$  e então podemos calcular  $F^{(n+1)}(x)$ ,

$$F^{(n+1)}(x) = f^{(n+1)}(x) - 0 - k(n + 1)!$$

avaliando a função acima em  $x = \alpha$

$$F^{(n+1)}(\alpha) = 0 = f^{(n+1)}(\alpha) - 0 - k(n + 1)!$$

explicitando  $k$ ,

$$k = \frac{f^{(n+1)}(\alpha)}{(n + 1)!}$$

logo em  $x = \bar{x}$

$$f(\bar{x}) - p(\bar{x}) = k\omega(\bar{x}) = \frac{\omega(\bar{x})}{(n + 1)!} f^{(n+1)}(\alpha).$$

## Forma de Newton

Já vimos que podemos obter o polinômio interpolador através dos polinômios de Lagrange ou via sistema linear. Embora a forma de Lagrange seja interessante teoricamente, ela não é, em geral, adequada para cálculos reais, particularmente para um grande número de pontos de suporte.

Desenvolveremos a seguir, outra forma de obtenção do polinômio interpolador, chamada *Forma de Newton*.

### Definição: (Diferenças Divididas)

As diferenças divididas são definidas pelas seguintes relações. Dados os pontos de suporte  $(x_i, f(x_i))$ ,  $i = 0, 1, 2, \dots, n$ ,

$$\left\{ \begin{array}{l} f[x_j] = f(x_j), \quad j = 0, 1, 2, \dots, n \\ \text{Para } r = 1, 2, \dots, n \\ f[x_j, x_{j+1}, \dots, x_{j+r-1}, x_{j+r}] = \frac{f[x_{j+1}, \dots, x_{j+r-1}, x_{j+r}] - f[x_j, x_{j+1}, \dots, x_{j+r-1}]}{x_{j+r} - x_j} \end{array} \right.$$

Podemos então montar uma tabela com as diferenças divididas, como abaixo



	Dif. Ordem 0	Dif. Ordem 1	Dif. Ordem 2	...	Dif. Ordem $n$
$x$	$f[x_j]$	$f[x_j, x_{j+1}]$	$f[x_j, x_{j+1}, x_{j+2}]$	...	...
$x_0$	$f[x_0]$				
$x_1$	$f[x_1]$	$f[x_0, x_1]$			
$x_2$	$f[x_2]$	$f[x_1, x_2]$	$f[x_0, x_1, x_2]$		
$\vdots$	$\vdots$	$\vdots$	$\vdots$		
$x_n$	$f[x_n]$	$f[x_{n-1}, x_n]$	$f[x_{n-2}, x_{n-1}, x_n]$	...	$f[x_0, x_1, x_2, \dots, x_{n-1}, x_n]$

Para o caso em que as abscissas de suporte são uniformemente distribuídas poderíamos definir as chamadas diferenças regressivas:

$$\begin{cases} \nabla^0 f_j = f(x_j), & j = 0, 1, 2, 3, \dots, n \\ \text{Para } r = 1, 2, \dots, n \\ \nabla^r f_j = \nabla^{r-1} f_j - \nabla^{r-1} f_{j-1}, & j = 0, 1, 2, 3, \dots, n \end{cases}$$

Da mesma forma podemos montar uma tabela com as diferenças regressivas.

$x$	$\nabla^0 f$	$\nabla^1 f$	$\nabla^2 f$	...	...
$x_0$	$\nabla^0 f_1$				
$x_1$	$\nabla^0 f_2$	$\nabla^1 f_1$			
$x_2$	$\nabla^0 f_3$	$\nabla^1 f_2$	$\nabla^2 f_2$		
$\vdots$	$\vdots$	$\vdots$	$\vdots$		
$x_n$	$\nabla^0 f_n$	$\nabla^1 f_n$	$\nabla^2 f_n$	...	$\nabla^n f_n$

Agora podemos usar as diferenças divididas para obter a forma de Newton do polinômio interpolador.

Uma importante observação deve ser feita sobre as diferenças divididas (ou regressivas). As diferenças divididas são invariantes a permutações de seus índices, ou seja, a ordem dos pares  $(x_i, f(x_i))$ , no esquema de cálculo das diferenças divididas é arbitrária, não sendo portanto obrigatório que  $x_0 < x_1 < x_2 < \dots < x_n$ . Por exemplo,

$$f[x_0, x_1, x_2] = f[x_0, x_2, x_1] = f[x_1, x_0, x_2] = f[x_1, x_2, x_0] = f[x_2, x_1, x_0] = f[x_2, x_0, x_1]$$

## Forma de Newton

O princípio da forma de Newton é escrever o polinômio interpolador (que é único), da seguinte maneira,

$$p(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1})$$

resta então, determinar os coeficientes  $a_0, a_1, \dots, a_n$ .

Faremos a construção do polinômio interpolador da seguinte forma; inicialmente construíremos  $p_0(x)$  que interpola  $f(x)$  em  $x = x_0$ , a seguir determinaremos  $p_1(x)$  que interpola  $f(x)$  em  $x = x_0$  e  $x = x_1$ , depois encontraremos  $p_2(x)$  que interpola  $f(x)$  em  $x = x_0, x = x_1$  e  $x = x_2$ , e assim sucessivamente até encontrarmos  $p_n(x) = p(x)$  que interpola  $f(x)$  em  $x = x_0, x = x_1, \dots, x = x_n$ .

Seja  $p_0(x)$  o polinômio de grau 0 que interpola  $f(x)$  em  $x = x_0$ . Logo  $p_0(x) = f(x_0) = f[x_0]$ .

Mas

$$f[x_0, x] = \frac{f[x] - f[x_0]}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0}$$

$$f(x) = \underbrace{f(x_0)}_{p_0(x)} + \underbrace{f[x_0, x](x - x_0)}_{E_0(x)}$$

Onde  $E_0(x)$  é o erro cometido quando interpolamos  $f(x)$  por  $p_0(x)$ .

Seja agora  $p_1(x)$ , o polinômio de primeiro grau que interpola  $f(x)$  em  $x = x_0$  e  $x = x_1$ . Logo,

$$\begin{aligned} f[x_0, x_1, x] &= f[x_1, x_0, x] = \frac{f[x_0, x] - f[x_1, x_0]}{x - x_1} = \frac{\frac{f[x] - f[x_0]}{x - x_0} - f[x_1, x_0]}{x - x_1} = \\ &= \frac{f[x] - f[x_0] - f[x_1, x_0](x - x_0)}{(x - x_0)(x - x_1)} \end{aligned}$$

logo

$$f(x) = \underbrace{f(x_0) + f[x_1, x_0](x - x_0)}_{p_1(x)} + \underbrace{f[x_0, x_1, x](x - x_0)(x - x_1)}_{E_1(x)}$$

logo,

$$p_1(x) = f(x_0) + f[x_0, x_1](x - x_0)$$

ou e

$$E_1(x) = f[x_0, x_1, x](x - x_0)(x - x_1)$$

Obviamente, podemos ver que  $p_1(x)$  interpola  $f(x)$  em  $x_0$  e  $x_1$ .

Já construímos  $p_0(x)$  e  $p_1(x)$ , vamos agora construir  $p_2(x)$ , o polinômio que interpola  $f(x)$  em  $x_0, x_1$  e  $x_2$ .

Analogamente,

$$\begin{aligned} f[x_0, x_1, x_2, x] &= f[x_2, x_1, x_0, x] = \frac{f[x_1, x_0, x] - f[x_2, x_1, x_0]}{x - x_1} = \frac{\frac{f[x_0, x] - f[x_1, x_0]}{x - x_1} - f[x_2, x_1, x_0]}{x - x_2} = \\ &= \frac{\frac{f[x] - f[x_0]}{x - x_0} - f[x_1, x_0]}{x - x_1} - f[x_2, x_1, x_0] = \\ &= \frac{f[x] - f[x_0] - f[x_1, x_0](x - x_0) - f[x_2, x_1, x_0](x - x_0)(x - x_1)}{(x - x_0)(x - x_1)(x - x_2)} \end{aligned}$$

logo

$$f(x) = \underbrace{f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1)}_{p_2(x)} + \underbrace{f[x_0, x_1, x_2, x](x - x_0)(x - x_1)(x - x_2)}_{E_2(x)}$$

Então

$$p_2(x) = \underbrace{f[x_0] + f[x_0, x_1](x - x_0)}_{p_1(x)} + f[x_0, x_1, x_2](x - x_0)(x - x_1)$$

e

$$E_2(x) = f[x_0, x_1, x_2, x](x - x_0)(x - x_1)(x - x_2)$$

Pode-se verificar que  $p_2(x)$  interpola  $f(x)$  em  $x_0, x_1$  e  $x_2$ .

Aplicando o mesmo procedimento sucessivamente para

$$x_0, x_1, x_2, x_3$$

$$x_0, x_1, x_2, x_3, x_4$$

$$x_0, x_1, x_2, x_3, x_4, x_5$$

.....

$$x_0, x_1, x_2, x_3, x_4, x_5, \dots, x_n$$

obtemos a forma de Newton para o polinômio interpolador que interpola  $f(x)$  em  $x_0, x_1, x_2, \dots, x_n$ .

$$p_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots \\ + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1) \dots (x - x_{n-1})$$

ou ainda

$$p_{k+1}(x) = p_k(x) + f[x_0, x_1, \dots, x_{k+1}](x - x_0)(x - x_1) \dots (x - x_k)$$

logo os coeficientes  $a_k$ ,  $k = 0, 1, \dots, n$  são dados por

$$a_k = f[x_0, x_1, \dots, x_k]$$

Neste caso temos a seguinte expressão para o erro

$$E_n(x) = f[x_0, x_1, \dots, x_n, x](x - x_0)(x - x_1) \dots (x - x_n)$$

Já tínhamos deduzido uma fórmula geral do erro para a interpolação polinomial,

$$f(\bar{x}) - p(\bar{x}) = \frac{\omega(\bar{x})f^{(n+1)}(\alpha)}{(n+1)!} \quad \alpha \in [x_0, x_n]$$

onde  $\omega(x) = (x - x_0)(x - x_1)(x - x_2) \dots (x - x_n)$ , como  $\omega(\bar{x}) \neq 0$ , temos que igualando as duas expressões para o erro

$$f[x_0, x_1, \dots, x_n, x] = \frac{f^{(n+1)}(\alpha)}{(n+1)!} \quad \alpha \in [x_0, x_n]$$

que relaciona derivadas e diferenças divididas.

### Vantagens da Forma de Newton

Como as diferenças divididas são invariantes à permutações dos pontos de suporte, podemos avaliar o polinômio interpolador em qualquer ordem, assim como adicionar pontos de suporte novos sem que haja necessidade de iniciar os cálculos novamente, como ocorre com outras formas do polinômio interpolador. Além disso, podemos avaliar o polinômio de uma maneira mais econômica através da forma dos parênteses encaixados ou esquema de Horner, que é,

$$p(z) = (\dots(a_n(z - x_{n-1}) + a_{n-1})(z - x_{n-2}) + \dots + a_1)(z - x_0) + a_0$$

**Exemplo:** Dados os seguintes pontos de suporte encontrar o polinômio interpolador na forma de Newton.

$x$	-1	0	2
$f(x)$	4	1	-1

Montando a tabela de diferenças divididas.

	Dif. Ordem 0	Dif. Ordem 1	Dif. Ordem 2
$x$	$f[x_j]$	$f[x_j, x_{j+1}]$	$f[x_j, x_{j+1}, x_{j+2}]$
$x_0 = -1$	$f[x_0] = 4$		
$x_1 = 0$	$f[x_1] = 1$	$f[x_0, x_1] = -3$	
$x_2 = 2$	$f[x_2] = -1$	$f[x_1, x_2] = -1$	$f[x_0, x_1, x_2] = \frac{2}{3}$

onde

$$f[x_0, x_1] = \frac{f[x_1] - f[x_0]}{x_1 - x_0} = \frac{1 - 4}{0 - (-1)} = -3$$

$$f[x_1, x_2] = \frac{f[x_2] - f[x_1]}{x_2 - x_1} = \frac{-1 - 1}{2 - 0} = -1$$

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{-1 - (-3)}{2 - (-1)} = \frac{2}{3}$$

logo

$$p(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1)$$

$$p(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1)$$

$$p(x) = 4 - 3(x + 1) + \frac{2}{3}(x + 1)x$$

## Fenômeno de Runge

### Interpolação por Splines

Referências:

Stoer e Burlisch - Albrecht - Ruggiero e Lopes

**Exemplo:**

Considere o reservatório abaixo para o qual temos a seguinte tabulação de recalque e rendimento.

$H/h$	6,0	7,0	8,0	9,0
%	0,70	0,63	0,57	0,50

Ache uma interpolante para esta tabela.

$$p_3(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2)$$

onde os coeficientes são dados por

$$a_0 = f[x_0]$$

$$a_1 = f[x_0, x_1]$$

$$a_2 = f[x_0, x_1, x_2]$$

$$a_3 = f[x_0, x_1, x_2, x_3]$$

Montando uma tabella de diferenças divididas obtemos

$$a_0 = 0.7000$$

$$a_1 = -0.0700$$

$$a_2 = 0.0050$$

$$a_3 = -0.0033$$

Portanto

$$p_3(x) = 0.7 - 0.07(x - 6.0) + 0.005(x - 6.0)(x - 7.0) - 0.0033(x - 6.0)(x - 7.0)(x - 8.0)$$

## Chapter 5

# Método dos Mínimos Quadrados

Como na interpolação temos uma função  $f(x)$  conhecida somente em alguns pontos  $(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))$  e desejamos conhecer o valor da função num ponto  $\bar{x}$ , ou seja, queremos calcular  $f(\bar{x})$ .

Vimos que a interpolação polinomial realiza esta tarefa. Entretanto, devido as características dos polinômios de alto grau geramos uma curva não suave. Em vários problemas estamos interessados em aproximar a função  $f(x)$  tabelada por uma curva suave.

Na interpolação a curva gerada contém todos os pontos de suporte. Agora queremos buscar uma curva que aproxime  $f(x)$  sem que necessariamente contenha estes pontos. Basta que a curva *passee*

*perto* dos pontos de suporte.

De uma maneira geral queremos ajustar uma curva com a seguinte forma

$$p_m(x) = a_0g_0(x) + a_1g_1(x) + a_2g_2(x) + \dots + a_mg_m(x)$$

onde as  $g_0(x), g_1(x), g_2(x), \dots, g_m(x)$  formam um conjunto de funções linearmente independentes. Logo, para cada um dos pontos temos um desvio

$$\delta_i = p_m(x_i) - f(x_i); \quad i = 0, 1, 2, \dots, n$$

No caso de polinômios teríamos três casos:

1. Se  $m = n$   $\longrightarrow$  temos o caso da interpolação polinomial que pode gerar uma curva não suave;
2. Se  $m > n$   $\longrightarrow$  teríamos que adicionar  $m - n$  pontos para que tenhamos o número de pontos igual ao número de incógnitas e recairíamos no caso anterior;
3. Se  $m < n$   $\longrightarrow$  neste caso temos que a curva não passaria por todos os pontos, porém seria uma curva suave.

Logo o caso desejado é o terceiro.

Queremos que os desvios sejam mínimos, uma maneira de fazer isto é minimizar a soma

$$\sum_{i=0}^n \delta_i = \delta_0 + \delta_1 + \delta_2 + \dots + \delta_n$$

Entretanto, os desvios podem se cancelar, a solução é minimizar a soma dos valores absolutos dos desvios

$$\sum_{i=0}^n |\delta_i| = |\delta_0| + |\delta_1| + |\delta_2| + \dots + |\delta_n|$$

que no entanto não é derivável no ponto 0.

Enfim, a melhor escolha é minimizar

$$\sum_{i=0}^n \delta_i^2 = \delta_0^2 + \delta_1^2 + \delta_2^2 + \dots + \delta_n^2$$

Daí o nome Método dos Mínimos Quadrados.

Temos que minimizar a função

$$\sum_{i=0}^n [\delta_i]^2 = \sum_{i=0}^n [p_m(x_i) - f(x_i)]^2$$

ou

$$\sum_{i=0}^n [\delta_i]^2 = \sum_{i=0}^n [a_0g_0(x) + a_1g_1(x) + \dots + a_mg_m(x) - f(x_i)]^2$$

Como são dados  $x_i$ ,  $f(x_i)$  e os  $g_i(x)$ , as incógnitas do problema são os coeficientes  $a_i$ . Visto que temos  $(m + 1)$  incógnitas, para minimizar a função acima, precisamos calcular as derivadas da função em relação a cada uma das variáveis  $a_0, a_1, a_2, \dots, a_m$  e igualá-las a 0. Isto é,

$$\left. \begin{aligned} \frac{\partial}{\partial a_0} \left[ \sum_{i=0}^n (\delta_i)^2 \right] &= 0 \\ \frac{\partial}{\partial a_1} \left[ \sum_{i=0}^n (\delta_i)^2 \right] &= 0 \\ \frac{\partial}{\partial a_2} \left[ \sum_{i=0}^n (\delta_i)^2 \right] &= 0 \\ &\vdots \\ \frac{\partial}{\partial a_m} \left[ \sum_{i=0}^n (\delta_i)^2 \right] &= 0 \end{aligned} \right\} (m + 1) \text{ equações}$$

As equações acima podem ser escritas compactamente na seguinte forma.

$$\frac{\partial}{\partial a_j} \left[ \sum_{i=0}^n (\delta_i)^2 \right] = 0 \quad j = 0, 1, 2, \dots, m$$

este sistema de equações é chamado **Equações Normais**.

Vamos expandir este sistema. A derivada da soma é a soma das derivadas, logo

$$\frac{\partial}{\partial a_j} \left[ \sum_{i=0}^n (\delta_i)^2 \right] = \sum_{i=0}^n \frac{\partial}{\partial a_j} (\delta_i)^2 = \sum_{i=0}^n 2\delta_i \frac{\partial \delta_i}{\partial a_j} = 0 \quad j = 0, 1, 2, \dots, m$$

então

$$\sum_{i=0}^n \delta_i \frac{\partial \delta_i}{\partial a_j} = 0 \quad j = 0, 1, 2, \dots, m$$

Vamos agora calcular o termo  $\frac{\partial \delta_i}{\partial a_j}$ .

O desvio  $\delta_i$  é dado por

$$\delta_i = p_m(x_i) - f(x_i)$$



portanto

$$\begin{aligned}\frac{\partial \delta_i}{\partial a_j} &= \frac{\partial}{\partial a_j} [a_0 g_0(x_i) + a_1 g_1(x_i) + \dots + a_j g_j(x_i) + \dots + a_m g_m(x_i) - f(x_i)] \\ &= 0 + 0 + \dots + g_j(x_i) + \dots + 0 - 0 \\ &= g_j(x_i)\end{aligned}$$

Como

$$\sum_{i=0}^n \delta_i \frac{\partial \delta_i}{\partial a_j} = 0 \quad j = 0, 1, 2, \dots, m$$

substituindo a expressão calculada para  $\frac{\partial \delta_i}{\partial a_j}$

$$\sum_{i=0}^n \delta_i g_j(x_i) = 0 \quad j = 0, 1, 2, \dots, m$$

mas novamente  $\delta_i = p_m(x_i) - f(x_i)$ , logo

$$\sum_{i=0}^n [a_0 g_0(x_i) + a_1 g_1(x_i) + \dots + a_m g_m(x_i) - f(x_i)] g_j(x_i) = 0; \quad j = 0, 1, 2, \dots, m$$

$$\sum_{i=0}^n [a_0 g_0(x_i) g_j(x_i) + a_1 g_1(x_i) g_j(x_i) + \dots + a_m g_m(x_i) g_j(x_i) - f(x_i) g_j(x_i)] = 0; \quad j = 0, 1, 2, \dots, m$$

ou

$$\sum_{i=0}^n a_0 g_0(x_i) g_j(x_i) + \sum_{i=0}^n a_1 g_1(x_i) g_j(x_i) + \dots + \sum_{i=0}^n a_m g_m(x_i) g_j(x_i) = \sum_{i=0}^n f(x_i) g_j(x_i); \quad j = 0, 1, 2, \dots, m$$

colocando em evidência os coeficientes  $a_i$ .

$$a_0 \sum_{i=0}^n g_0(x_i) g_j(x_i) + a_1 \sum_{i=0}^n g_1(x_i) g_j(x_i) + \dots + a_m \sum_{i=0}^n g_m(x_i) g_j(x_i) = \sum_{i=0}^n f(x_i) g_j(x_i); \quad j = 0, 1, 2, \dots, m$$

para cada valor de  $j$  temos uma equação

$$j = 0 \rightarrow a_0 \sum_{i=0}^n g_0(x_i) g_0(x_i) + a_1 \sum_{i=0}^n g_1(x_i) g_0(x_i) + \dots + a_m \sum_{i=0}^n g_m(x_i) g_0(x_i) = \sum_{i=0}^n f(x_i) g_0(x_i)$$

$$j = 1 \rightarrow a_0 \sum_{i=0}^n g_0(x_i) g_1(x_i) + a_1 \sum_{i=0}^n g_1(x_i) g_1(x_i) + \dots + a_m \sum_{i=0}^n g_m(x_i) g_1(x_i) = \sum_{i=0}^n f(x_i) g_1(x_i)$$

$$j = 2 \rightarrow a_0 \sum_{i=0}^n g_0(x_i) g_2(x_i) + a_1 \sum_{i=0}^n g_1(x_i) g_2(x_i) + \dots + a_m \sum_{i=0}^n g_m(x_i) g_2(x_i) = \sum_{i=0}^n f(x_i) g_2(x_i)$$

⋮

$$j = m \rightarrow a_0 \sum_{i=0}^n g_0(x_i) g_m(x_i) + a_1 \sum_{i=0}^n g_1(x_i) g_m(x_i) + \dots + a_m \sum_{i=0}^n g_m(x_i) g_m(x_i) = \sum_{i=0}^n f(x_i) g_m(x_i)$$

Que é um sistema linear simétrico com os coeficientes  $a_0, a_1, a_2, \dots, a_m$  como incógnitas. Genericamente os coeficientes da matriz têm a seguinte forma

$$a_{kj} = \sum_{i=0}^n g_k(x_i)g_j(x_i); \quad k = 0, 1, 2, \dots, m; \quad j = 0, 1, 2, \dots, m$$

que reescrito na forma matricial torna-se

$$\begin{bmatrix} \sum_{i=0}^n g_0(x_i)g_0(x_i) & \sum_{i=0}^n g_1(x_i)g_0(x_i) & \sum_{i=0}^n g_2(x_i)g_0(x_i) & \cdots & \sum_{i=0}^n g_m(x_i)g_0(x_i) \\ \sum_{i=0}^n g_0(x_i)g_1(x_i) & \sum_{i=0}^n g_1(x_i)g_1(x_i) & \sum_{i=0}^n g_2(x_i)g_1(x_i) & \cdots & \sum_{i=0}^n g_m(x_i)g_1(x_i) \\ \sum_{i=0}^n g_0(x_i)g_2(x_i) & \sum_{i=0}^n g_1(x_i)g_2(x_i) & \sum_{i=0}^n g_2(x_i)g_2(x_i) & \cdots & \sum_{i=0}^n g_m(x_i)g_2(x_i) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum_{i=0}^n g_0(x_i)g_m(x_i) & \sum_{i=0}^n g_1(x_i)g_m(x_i) & \sum_{i=0}^n g_2(x_i)g_m(x_i) & \cdots & \sum_{i=0}^n g_m(x_i)g_m(x_i) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n f(x_i)g_0(x_i) \\ \sum_{i=0}^n f(x_i)g_1(x_i) \\ \sum_{i=0}^n f(x_i)g_2(x_i) \\ \vdots \\ \sum_{i=0}^n f(x_i)g_m(x_i) \end{bmatrix}$$

O caso mais freqüentemente usado em ajuste de curvas é o das funções polinomiais, isto é,  $p_m(x)$  tem a forma geral:

$$p_m(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_mx^m$$

ou

$$g_0(x) = 1; \quad g_1(x) = x; \quad g_2(x) = x^2; \quad \dots \quad ; g_m(x) = x^m.$$

Neste caso o sistema linear se torna

$$\begin{bmatrix} \sum_{i=0}^n 1 & \sum_{i=0}^n x & \sum_{i=0}^n x^2 & \cdots & \sum_{i=0}^n x^m \\ \sum_{i=0}^n x & \sum_{i=0}^n x^2 & \sum_{i=0}^n x^3 & \cdots & \sum_{i=0}^n x^{m+1} \\ \sum_{i=0}^n x^2 & \sum_{i=0}^n x^3 & \sum_{i=0}^n x^4 & \cdots & \sum_{i=0}^n x^{m+2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum_{i=0}^n x^m & \sum_{i=0}^n x^{m+1} & \sum_{i=0}^n x^{m+2} & \cdots & \sum_{i=0}^n x^{m+m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n f(x_i) \\ \sum_{i=0}^n f(x_i)x \\ \sum_{i=0}^n f(x_i)x^2 \\ \vdots \\ \sum_{i=0}^n f(x_i)x^m \end{bmatrix}$$

que envolve  $n + 1$  equações e  $n + 1$  incógnitas.

**Exemplo:** Seja a função tabelada abaixo:

$x_i$	0	1	2	3	4
$f(x_i)$	1100	1080	1040	960	840

Vamos ajustar uma parábola aos pontos dados, ou seja, queremos achar

$$p_m(x) = a_2x^2 + a_1x + a_0 = a_2g_2(x) + a_1g_1(x) + a_0g_0(x)$$

O sistema linear para este ajuste é,

$$\begin{bmatrix} \sum_{i=0}^n 1 & \sum_{i=0}^n x & \sum_{i=0}^n x^2 \\ \sum_{i=0}^n x & \sum_{i=0}^n x^2 & \sum_{i=0}^n x^3 \\ \sum_{i=0}^n x^2 & \sum_{i=0}^n x^3 & \sum_{i=0}^n x^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n f(x_i) \\ \sum_{i=0}^n f(x_i)x \\ \sum_{i=0}^n f(x_i)x^2 \end{bmatrix}$$

temos que calcular os seguintes coeficientes

$$\sum_{i=0}^4 1 = 1 + 1 + 1 + 1 + 1 = 5$$

$$\sum_{i=0}^4 x_i = x_0 + x_1 + x_2 + x_3 + x_4 = 0 + 1 + 2 + 3 + 4 = 10$$

$$\begin{aligned} \sum_{i=0}^4 x_i^2 &= x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 = 0^2 + 1^2 + 2^2 + 3^2 + 4^2 = \\ &= 0 + 1 + 4 + 9 + 16 = 30 \end{aligned}$$

$$\begin{aligned} \sum_{i=0}^4 x_i^3 &= x_0^3 + x_1^3 + x_2^3 + x_3^3 + x_4^3 = 0^3 + 1^3 + 2^3 + 3^3 + 4^3 = \\ &= 0 + 1 + 8 + 27 + 64 = 100 \end{aligned}$$

$$\begin{aligned} \sum_{i=0}^4 x_i^4 &= x_0^4 + x_1^4 + x_2^4 + x_3^4 + x_4^4 = 0^4 + 1^4 + 2^4 + 3^4 + 4^4 = \\ &= 0 + 1 + 16 + 81 + 256 = 354 \end{aligned}$$

$$\begin{aligned} \sum_{i=0}^4 f(x_i) &= f(x_0) + f(x_1) + f(x_2) + f(x_3) + f(x_4) = \\ &= 1100 + 1080 + 1040 + 960 + 840 = 5020 \end{aligned}$$

$$\sum_{i=0}^4 f(x_i)x_i = f(x_0)x_0 + f(x_1)x_1 + f(x_2)x_2 + f(x_3)x_3 + f(x_4)x_4 =$$

$$= 1100 \cdot 0 + 1080 \cdot 1 + 1040 \cdot 2 + 960 \cdot 3 + 840 \cdot 4 = 9400$$

$$\sum_{i=0}^4 f(x_i)x_i^2 = f(x_0)x_0^2 + f(x_1)x_1^2 + f(x_2)x_2^2 + f(x_3)x_3^2 + f(x_4)x_4^2 =$$

$$= 1100 \cdot 0 + 1080 \cdot 1 + 1040 \cdot 4 + 960 \cdot 9 + 840 \cdot 16 = 27320$$

resultando no seguinte sistema linear

$$\begin{bmatrix} 5 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 5020 \\ 9400 \\ 27320 \end{bmatrix}$$

resolvendo, obtemos

$$a_0 = 1097,714286$$

$$a_1 = 4,571428186$$

$$a_2 = -17,14285706$$

$$p(x) = 1097,714286 + 4,571428186x - 17,14285706x^2$$

$$p(0) = 1097,714286$$

$$p(1) = 1085,142857$$

$$p(2) = 1038,285714$$

$$p(3) = 957,1434311$$

$$p(4) = 841,7142857$$

Resumindo

$i$	$x_i$	$f(x_i)$	$p(x_i)$	$\delta_i = p(x_i) - f(x_i)$
0	0	1100	1097,714286	-2,285714
1	1	1080	1085,142857	5,142857
2	2	1040	1038,285714	-1,714286
3	3	960	957,1434311	-2,8565689
4	4	840	841,7142857	1,7142857

## Ajuste de Curvas – Método dos Mínimos Quadrados – Exercícios

1. Ajuste uma reta pelos pontos abaixo tabelados usando o método dos Mínimos Quadrados.

$x$	1	2	3	4	5	6
$f(x)$	0,5	0,6	0,9	0,8	1,2	1,5

2. Dada a função tabelada abaixo ajuste um polinômio de grau 2 (uma parábola) a estes pontos, usando o Método dos Mínimos Quadrados.

$x$	$f(x)$
0,0	10,0
5,0	35,0
10,0	110,0
15,0	235,0
20,0	410,0

3. A tabela abaixo fornece o número de habitantes de uma cidade para os últimos anos, em milhares de pessoas. Ajuste uma reta (polinômio de grau 1) a estes pontos com o Método dos Mínimos Quadrados. A seguir estime a população desta cidade no ano de 1994.

Ano	1989	1990	1991	1992	1993
Habitantes	0,48	0,50	0,52	0,54	0,55

4. Dada a tabela abaixo do calor específico da água em função da temperatura, achar  $c = c(T) = a + bT + dT^2 + eT^3$ .

$T(^{\circ}\text{C})$	$c(T)$
0	1.00762
5	1.00392
10	1.00153
15	1.00000
20	0.99907
25	0.99852
30	0.98765
35	0.98000
40	0.97666
45	0.97000

5. Uma função  $f(x)$  que é integrável em  $[0,5]$  é conhecida somente em alguns pontos como mostra a tabela abaixo. Pelo método dos mínimos quadrados ajuste uma linha reta a estes pontos e aproxime  $\int_0^5 f(x) dx$  pela integração da reta encontrada.

$x$	$f(x)$
0	4,1
1	6,5
2	9,0
7	13,2
8	17,1
10	20,1

6. O número de bactérias por unidade de volume existente em uma cultura após  $x$  horas é apresentado na tabela:

nº horas ( $x$ )	0	1	2	3	4	5	6
nº bactérias/vol	42	47	65	92	132	190	275

Se plotarmos a curva acima veremos que é do tipo exponencial, e devemos portanto ajustar uma curva do tipo  $y = ab^x$ .

Utilizando o Método dos Mínimos Quadrados, obtenha os valores de  $a$  e  $b$ .

*Sugestão:*

$$y = ab^x$$

$$\log y = \log ab^x$$

$$\log y = \log a + \log b^x$$

$$\log y = \log a + x \log b$$

$$Y = A + BX$$

**Solução:**

Podemos resolver o problema como sendo o ajuste de uma reta aos pontos  $(x, \log y)$ . Isto é,

$X$	0	1	2	3	4	5	6
$Y$	1,623249290	1,672097858	1,81291357	1,963787827	2,120573931	2,27875360	2,439332694

Para ajustar uma reta temos o seguinte sistema a resolver

$$\begin{cases} A \sum_{i=0}^n 1 + B \sum_{i=0}^n X_i = \sum_{i=0}^n Y_i \\ A \sum_{i=0}^n X_i + B \sum_{i=0}^n X_i^2 = \sum_{i=0}^n Y_i X_i \end{cases}$$

$$\sum_{i=0}^6 1 = 1 + 1 + 1 + 1 + 1 + 1 + 1 = 7$$

$$\sum_{i=0}^6 X_i = X_0 + X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 0 + 1 + 2 + 3 + 4 + 5 + 6 = 21$$

$$\begin{aligned}\sum_{i=0}^6 X_i^2 &= X_0^2 + X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2 + X_6^2 = 0^2 + 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = \\ &= 0 + 1 + 4 + 9 + 16 + 25 + 36 = 91\end{aligned}$$

$$\begin{aligned}\sum_{i=0}^6 Y_i &= Y_0 + Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6 = \\ &= 1,623249290 + 1,672097858 + 1,81291357 + 1,963787827 + \\ &+ 2,120573931 + 2,27875360 + 2,439332694 = 13,91070877\end{aligned}$$

$$\begin{aligned}\sum_{i=0}^6 Y_i X_i &= Y_0 X_0 + Y_1 X_1 + Y_2 X_2 + Y_3 X_3 + Y_4 X_4 + Y_5 X_5 + Y_6 X_6 = \\ &= 1,623249290 \cdot 0 + 1,672097858 \cdot 1 + 1,81291357 \cdot 2 + 1,963787827 \cdot 3 + \\ &+ 2,120573931 \cdot 4 + 2,27875360 \cdot 5 + 2,439332694 \cdot 6 = 45,70134837\end{aligned}$$

Substituindo-se no sistema temos

$$\begin{cases} 7A + 21B = 13,91070877 \\ 21A + 91B = 45,70134837 \end{cases}$$

resolvendo, obtemos  $A = 1,561970318$  e  $B = 0,141757931$ . Mas,

$$A = \log a = 1,561970318 \longrightarrow a = 36,47290185$$

e

$$B = \log b = 0,141757931 \longrightarrow b = 1,385983087$$

Assim,

$$y = 36,47290185 \cdot 1,385983087^x$$

7. Mostre que a reta da regressão linear de  $x$  em  $y$  passa através do ponto  $(\bar{x}, \bar{y})$  que é a média dos valores de  $x$  e  $y$ , isto é,

$$\bar{x} = \frac{\sum_{i=0}^n x_i}{(n+1)} \quad \text{e} \quad \bar{y} = \frac{\sum_{i=0}^n y_i}{(n+1)}$$

**Solução:**

Por simplicidade de notação omitiremos os sub-índices e super-índice dos somatórios. Pelo Método dos Mínimos Quadrados para a regressão linear podemos escrever:

$$\begin{cases} (\sum 1) a_0 + (\sum x_i) a_1 = \sum y_i \\ (\sum x_i) a_0 + (\sum x_i^2) a_1 = \sum y_i x_i \end{cases}$$

Substituindo a segunda equação por ela menos o múltiplo

$$\frac{\sum x_i}{\sum 1}$$

da primeira equação obtemos

$$\begin{cases} (\sum 1) a_0 + (\sum x_i) a_1 = \sum y_i \\ + \left[ (\sum x_i^2) - \frac{(\sum x_i)^2}{(\sum 1)} \right] a_1 = \sum y_i x_i - \frac{\sum y_i \sum x_i}{\sum 1} \end{cases}$$

logo

$$a_1 = \frac{\sum y_i x_i \sum 1 - \sum y_i \sum x_i}{\sum 1 \sum x_i^2 - (\sum x_i)^2}$$

e

$$a_0 = \frac{1}{\sum 1} \left\{ \sum y_i - \sum x_i \left[ \frac{\sum y_i x_i \sum 1 - \sum y_i \sum x_i}{\sum 1 \sum x_i^2 - (\sum x_i)^2} \right] \right\}$$

ou

$$a_0 = \frac{\sum y_i}{\sum 1} - \frac{\sum x_i}{\sum 1} \left[ \frac{\sum y_i x_i \sum 1 - \sum y_i \sum x_i}{\sum 1 \sum x_i^2 - (\sum x_i)^2} \right]$$

Sabemos que a reta da regressão linear é dada por

$$y = a_1 x + a_0$$

onde os coeficientes  $a_0$  e  $a_1$  são dados acima.

Vamos agora substituir o ponto médio na reta acima

$$a_1 \bar{x} + a_0 = a_1 \frac{\sum x_i}{(n+1)} + a_0 =$$



$$= \left[ \frac{\sum y_i x_i \sum 1 - \sum y_i \sum x_i}{\sum 1 \sum x_i^2 - (\sum x_i)^2} \right] \frac{\sum x_i}{(n+1)} + \frac{\sum y_i}{\sum 1} - \frac{\sum x_i}{\sum 1} \left[ \frac{\sum y_i x_i \sum 1 - \sum y_i \sum x_i}{\sum 1 \sum x_i^2 - (\sum x_i)^2} \right]$$

$$= \frac{\sum y_i}{\sum 1} = \bar{y}$$

■

8. A intensidade de radiação de uma fonte radioativa é dada por

$$I = I_0 e^{-\alpha t}$$

Determinar  $I_0$  e  $\alpha$  para os seguintes resultados experimentais.

$t$	0,2	0,3	0,4	0,5	0,6	0,7	0,8
$I$	3,16	2,38	1,75	1,34	1,00	0,74	0,56

**Solução:**

$$I = I_0 e^{-\alpha t}$$

Aplicando logaritmo aos dois lados da equação temos:

$$\ln I = \ln I_0 - \alpha t$$

logo, podemos resolver o problema como sendo o ajuste de uma reta aos pontos  $(t, \ln I)$ . Isto é,

$$y = a_0 + a_1 x = \ln I_0 - \alpha t = \ln I$$

assim,  $a_0 = \ln I_0$  e  $a_1 = -\alpha$ .

$t$	0,2	0,3	0,4	0,5	0,6	0,7	0,8
$\ln I$	1,150572	0,867100	0,559616	0,292670	0,000000	-0,301105	-0,579818

Como queremos ajustar uma reta temos o seguinte sistema a resolver

$$\begin{cases} a_0 \sum_{i=0}^n 1 + a_1 \sum_{i=0}^n x_i = \sum_{i=0}^n f(x_i) \\ a_0 \sum_{i=0}^n x_i + a_1 \sum_{i=0}^n x_i^2 = \sum_{i=0}^n f(x_i) x_i \end{cases}$$

Onde  $x = t$  e  $y = \ln I$ .

$$\sum_{i=0}^6 1 = 1 + 1 + 1 + 1 + 1 + 1 + 1 = 7$$

$$\sum_{i=0}^6 x_i = x_0 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

$$= 0,2 + 0,3 + 0,4 + 0,5 + 0,6 + 0,7 + 0,8 = 3,5$$

$$\sum_{i=0}^6 x_i^2 = x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2$$

$$= 0,2^2 + 0,3^2 + 0,4^2 + 0,5^2 + 0,6^2 + 0,7^2 + 0,8^2 =$$

$$= 0,04 + 0,09 + 0,16 + 0,25 + 0,36 + 0,49 + 0,64 = 2,03$$

$$\sum_{i=0}^6 f(x_i) = f(x_0) + f(x_1) + f(x_2) + f(x_3) + f(x_4) + f(x_5) + f(x_6) =$$

$$= 1,150572 + 0,867100 + 0,559616 + 0,292670 +$$

$$+ 0,000000 - 0,301105 - 0,579818 = 1,989035$$

$$\sum_{i=0}^6 f(x_i)x_i = f(x_0)x_0 + f(x_1)x_1 + f(x_2)x_2 + f(x_3)x_3$$

$$+ f(x_4)x_4 + f(x_5)x_5 + f(x_6)x_6 =$$

$$= 1,150572 \cdot 0,2 + 0,867100 \cdot 0,3 + 0,559616 \cdot 0,4 + 0,292670 \cdot 0,5 +$$

$$+ 0,000000 \cdot 0,6 - 0,301105 \cdot 0,7 - 0,579818 \cdot 0,8 = 0,185798$$

Substituindo-se no sistema temos

$$\begin{cases} 7a_0 + 3,5a_1 = 1,989035 \\ 3,5a_0 + 2,03a_1 = 0,185798 \end{cases}$$

resolvendo, obtemos  $a_0 = 1,728292$  e  $a_1 = -2,888284$ . Mas,

$$a_0 = \ln I_0 = 1,728292 \longrightarrow I_0 = 5,631027$$

e

$$a_1 = -\alpha = -2,888284 \longrightarrow \alpha = 2,888284$$

Assim,

$$I = 5,631027e^{-2,888284t}$$

9. O calor específico da água, como uma função da temperatura, é,

Temperatura (° C)	Calor Específico
20	0,99907
25	0,99852
30	0,99826
35	0,99818
40	0,99828
45	0,99849
50	0,99878

- (a) Use interpolação linear para estimar o calor específico em 37° C.
- (b) Ajuste uma reta pelo Método dos Mínimos Quadrados aos pontos tabelados. A seguir, calcule o calor específico em 37° C.

**Solução:**

(a)

$$\frac{x - 0,99818}{0,99828 - 0,99818} = \frac{37 - 35}{40 - 35}$$

logo

$$x = 0,0001\frac{2}{5} + 0,99818 = 0,99822$$

(b)

$$\begin{bmatrix} \sum_{i=0}^6 1 & \sum_{i=0}^6 x_i \\ \sum_{i=0}^6 x_i & \sum_{i=0}^6 x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^6 f(x_i) \\ \sum_{i=0}^6 f(x_i)x_i \end{bmatrix}$$

mas

$$\sum_{i=0}^6 1 = 7$$

$$\sum_{i=0}^6 x_i = 245$$

$$\sum_{i=0}^6 x_i^2 = 9275$$

$$\sum_{i=0}^6 f(x_i) = 6,98958$$

$$\sum_{i=0}^6 f(x_i)x_i = 244,63075$$

temos então,

$$\begin{bmatrix} 7 & 245 \\ 245 & 9275 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 6,98958 \\ 244,63075 \end{bmatrix}$$

resolvendo o sistema temos  $a_0 = 0,998738929$  e  $a_1 = -0,0000065$ . Assim para  $37^\circ\text{C}$  temos

$$CE = a_0 + a_1x = 0,998738929 - 0,0000065 \cdot 37 = 0,998498429$$

10. Mostre que se fossemos ajustar um polinômio de segundo grau, pelo Método dos Mínimos Quadrados, a dois pontos somente, o sistema de equações lineares resultante terá um número infinito de soluções, correspondente ao fato de que por dois pontos podemos passar um número infinito de parábolas.

**Solução:**

Polinômio de Segundo grau

$$p(x) = a_0 + a_1x + a_2x^2$$

logo

$$\begin{bmatrix} \sum_{i=0}^n 1 & \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 & \sum_{i=0}^n x_i^3 \\ \sum_{i=0}^n x_i^2 & \sum_{i=0}^n x_i^3 & \sum_{i=0}^n x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n f(x_i) \\ \sum_{i=0}^n f(x_i)x_i \\ \sum_{i=0}^n f(x_i)x_i^2 \end{bmatrix}$$

com somente dois pontos  $(x_0, f(x_0))$  e  $(x_1, f(x_1))$  temos

$$\begin{bmatrix} 2 & (x_0 + x_1) & (x_0^2 + x_1^2) \\ (x_0 + x_1) & (x_0^2 + x_1^2) & (x_0^3 + x_1^3) \\ (x_0^2 + x_1^2) & (x_0^3 + x_1^3) & (x_0^4 + x_1^4) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} (f(x_0) + f(x_1)) \\ (f(x_0)x_0 + f(x_1)x_1) \\ (f(x_0)x_0^2 + f(x_1)x_1^2) \end{bmatrix}$$

Vamos tentar resolver o sistema acima por Eliminação de Gauss. Gruparemos o segundo membro com a matriz de coeficientes para simplificar a notação.

$$\left[ \begin{array}{ccc|c} 2 & (x_0 + x_1) & (x_0^2 + x_1^2) & (f(x_0) + f(x_1)) \\ (x_0 + x_1) & (x_0^2 + x_1^2) & (x_0^3 + x_1^3) & (f(x_0)x_0 + f(x_1)x_1) \\ (x_0^2 + x_1^2) & (x_0^3 + x_1^3) & (x_0^4 + x_1^4) & (f(x_0)x_0^2 + f(x_1)x_1^2) \end{array} \right] \begin{array}{l} \\ \text{(ii)} \leftarrow \text{(ii)} - (x_0 + x_1)/2\text{(i)} \\ \text{(iii)} \leftarrow \text{(iii)} - (x_0^2 + x_1^2)/2\text{(i)} \end{array}$$

logo

$$\left[ \begin{array}{ccc|c} 2 & (x_0 + x_1) & (x_0^2 + x_1^2) & (f(x_0) + f(x_1)) \\ 0 & (x_0^2 + x_1^2) - \frac{(x_0+x_1)^2}{2} & (x_0^3 + x_1^3) - \frac{(x_0+x_1)(x_0^2+x_1^2)}{2} & (f(x_0)x_0 + f(x_1)x_1) - \frac{(x_0+x_1)(f(x_0)+f(x_1))}{2} \\ 0 & (x_0^3 + x_1^3) - \frac{(x_0+x_1)(x_0^2+x_1^2)}{2} & (x_0^4 + x_1^4) - \frac{(x_0^2+x_1^2)^2}{2} & (f(x_0)x_0^2 + f(x_1)x_1^2) - \frac{(x_0^2+x_1^2)(f(x_0)+f(x_1))}{2} \end{array} \right]$$

$$\left[ \begin{array}{ccc|c} 2 & (x_0 + x_1) & (x_0^2 + x_1^2) & (f(x_0) + f(x_1)) \\ 0 & \frac{2x_0^2+2x_1^2-x_0^2-2x_0x_1-x_1^2}{2} & \frac{2x_0^3+2x_1^3-x_0^3-x_1x_0^2-x_0x_1^2-x_1^3}{2} & \frac{2f(x_0)x_0+2f(x_1)x_1-f(x_0)x_0-f(x_0)x_1-f(x_1)x_0-f(x_1)x_1}{2} \\ 0 & \frac{2x_0^3+2x_1^3-x_0^3-x_1x_0^2-x_0x_1^2-x_1^3}{2} & \frac{2x_0^4+2x_1^4-x_0^4-2x_0^2x_1^2-x_1^4}{2} & \frac{2f(x_0)x_0^2+2f(x_1)x_1^2-f(x_0)x_0^2-f(x_0)x_1^2-f(x_1)x_0^2-f(x_1)x_1^2}{2} \end{array} \right]$$

$$\left[ \begin{array}{ccc|c} 2 & (x_0 + x_1) & (x_0^2 + x_1^2) & (f(x_0) + f(x_1)) \\ 0 & \frac{x_0^2 - 2x_0x_1 + x_1^2}{2} & \frac{x_0^3 - x_1x_0^2 - x_0x_1^2 + x_1^3}{2} & \frac{f(x_0)x_0 - f(x_0)x_1 - f(x_1)x_0 + f(x_1)x_1}{2} \\ 0 & \frac{x_0^3 - x_1x_0^2 - x_0x_1^2 + x_1^3}{2} & \frac{x_0^4 - 2x_0^2x_1^2 + x_1^4}{2} & \frac{f(x_0)x_0^2 - f(x_0)x_1^2 - f(x_1)x_0^2 + f(x_1)x_1^2}{2} \end{array} \right]$$

$$\left[ \begin{array}{ccc|c} 2 & (x_0 + x_1) & (x_0^2 + x_1^2) & (f(x_0) + f(x_1)) \\ 0 & \frac{(x_0 - x_1)^2}{2} & \frac{(x_0 - x_1)x_0^2 - (x_0 - x_1)x_1^2}{2} & \frac{f(x_0)(x_0 - x_1) - f(x_1)(x_0 - x_1)}{2} \\ 0 & \frac{(x_0 - x_1)x_0^2 - (x_0 - x_1)x_1^2}{2} & \frac{(x_0^2 - x_1^2)^2}{2} & \frac{f(x_0)(x_0^2 - x_1^2) - f(x_1)(x_0^2 - x_1^2)}{2} \end{array} \right]$$

$$\left[ \begin{array}{ccc|c} 2 & (x_0 + x_1) & (x_0^2 + x_1^2) & (f(x_0) + f(x_1)) \\ 0 & \frac{(x_0 - x_1)^2}{2} & \frac{(x_0 - x_1)(x_0^2 - x_1^2)}{2} & \frac{(f(x_0) - f(x_1))(x_0 - x_1)}{2} \\ 0 & \frac{(x_0 - x_1)(x_0^2 - x_1^2)}{2} & \frac{(x_0^2 - x_1^2)^2}{2} & \frac{(f(x_0) - f(x_1))(x_0^2 - x_1^2)}{2} \end{array} \right]$$

Dividindo a segunda linha por  $(x_0 - x_1)$  e a terceira linha por  $(x_0^2 - x_1^2)$

$$\left[ \begin{array}{ccc|c} 2 & (x_0 + x_1) & (x_0^2 + x_1^2) & (f(x_0) + f(x_1)) \\ 0 & \frac{(x_0 - x_1)}{2} & \frac{(x_0^2 - x_1^2)}{2} & \frac{(f(x_0) - f(x_1))}{2} \\ 0 & \frac{(x_0 - x_1)}{2} & \frac{(x_0^2 - x_1^2)}{2} & \frac{(f(x_0) - f(x_1))}{2} \end{array} \right]$$

Logo o sistema acima é indeterminado, visto que possui duas linhas iguais.

11. Escreva um procedimento computacional, em pseudocódigo ou usando alguma linguagem de programação, para ler  $(n + 1)$  pontos conhecidos  $(x_i, f(x_i))$  de uma função  $f(x)$  e um número inteiro  $m$ ,  $m < n$ , e a seguir, monte o sistema de equações lineares para a obtenção do polinômio de grau  $m$  ajustado pelo Método dos Mínimos Quadrados.

12. Ajuste aos seguintes pontos dados

$x_i$	$f(x_i)$
-6	10
-4	9
-2	6
0	5
2	4
4	4

uma curva do tipo:

$$y = \frac{1}{a_0 + a_1x}$$

*Sugestão:*

$$z = 1/f(x)$$

$$z = a_0 + a_1x$$

**Solução:**

Logo, podemos resolver o problema como sendo o ajuste de uma reta aos pontos  $(x, 1/f(x))$ . Isto é,

$x_i$	$z_i = 1/f(x_i)$
-6	1/10=0.100000000
-4	1/9=0.111111111
-2	1/6=0.166666667
0	1/5=0.200000000
2	1/4=0.250000000
4	1/4=0.250000000

Como queremos ajustar uma reta temos o seguinte sistema a resolver

$$\begin{cases} a_0 \sum_{i=0}^n 1 + a_1 \sum_{i=0}^n x_i = \sum_{i=0}^n z_i \\ a_0 \sum_{i=0}^n x_i + a_1 \sum_{i=0}^n x_i^2 = \sum_{i=0}^n z_i x_i \end{cases}$$

Onde  $z = 1/f(x)$ .

$$\sum_{i=0}^5 1 = 1 + 1 + 1 + 1 + 1 + 1 = 6$$

$$\sum_{i=0}^5 x_i = x_0 + x_1 + x_2 + x_3 + x_4 + x_5 = (-6) + (-4) + (-2) + 0 + 2 + 4 = -6$$

$$\begin{aligned} \sum_{i=0}^5 x_i^2 &= x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = (-6)^2 + (-4)^2 + (-2)^2 + (0)^2 + (2)^2 + (4)^2 = \\ &= 36 + 16 + 4 + 0 + 4 + 16 = 76 \end{aligned}$$

$$\sum_{i=0}^5 z_i = z_0 + z_1 + z_2 + z_3 + z_4 + z_5 = \frac{1}{10} + \frac{1}{9} + \frac{1}{6} + \frac{1}{5} + \frac{1}{4} + \frac{1}{4} = 1,077777778$$

$$\sum_{i=0}^5 z_i x_i = z_0 x_0 + z_1 x_1 + z_2 x_2 + z_3 x_3 + z_4 x_4 + z_5 x_5 =$$

$$= \frac{1}{10} \cdot (-6) + \frac{1}{9} \cdot (-4) + \frac{1}{6} \cdot (-2) + \frac{1}{5} \cdot (0) + \frac{1}{4} \cdot (2) + \frac{1}{4} \cdot (4) = 0,122222222$$

Substituindo-se no sistema temos

$$\begin{cases} 6a_0 - 6a_1 = 1,077777778 \\ -6a_0 + 76a_1 = 0,122222222 \end{cases}$$

resolvendo, obtemos  $a_0 = 0.196772487$  e  $a_1 = 0,017142857$ . Logo,

$$y = \frac{1}{0.196772487 + 0,017142857x}$$

Verificando

$$y_0 = \frac{1}{0.196772487 + 0,017142857x_0} = \frac{1}{0.196772487 + 0,017142857 \cdot (-6)} = 10.64788720$$

$$y_1 = \frac{1}{0.196772487 + 0,017142857x_1} = \frac{1}{0.196772487 + 0,017142857 \cdot (-4)} = 7.800247578$$

$$y_2 = \frac{1}{0.196772487 + 0,017142857x_2} = \frac{1}{0.196772487 + 0,017142857 \cdot (-2)} = 6.154347199$$

$$y_3 = \frac{1}{0.196772487 + 0,017142857x_3} = \frac{1}{0.196772487 + 0,017142857 \cdot (0)} = 5.082011287$$

$$y_4 = \frac{1}{0.196772487 + 0,017142857x_4} = \frac{1}{0.196772487 + 0,017142857 \cdot (2)} = 4.327913901$$

$$y_5 = \frac{1}{0.196772487 + 0,017142857x_5} = \frac{1}{0.196772487 + 0,017142857 \cdot (4)} = 3.768693923$$

13. Sabendo-se que a dependência funcional entre a carga  $Q$  de um condensador e o tempo  $t$  é do tipo  $Q = a10^{kt}$ , determinar  $a$  e  $k$  a partir da seguinte tabela:

$t$ (segundos)	0,5	0,6	0,7	0,8	0,9
$Q$ (Coulomb)	4,78	3,97	3,30	2,75	1,9

### Solução:

Podemos aplicar a função logaritmo de base decimal a ambos os lados da equação sugerida, e obteríamos:

$$Q = a10^{kt}$$

$$\log Q = \log(a10^{kt})$$

$$= \log a + \log(10^{kt})$$



$$= \log a + kt \log(10)$$

$$= \log a + kt$$

que possui a forma geral

$$y = a_0 + a_1 x$$

onde

$$y = \log Q$$

$$x = t$$

$$a_0 = \log a$$

e

$$a_1 = k$$

Devemos realizar um ajuste de uma reta pelo Método dos Mínimos Quadrados aos pontos  $(x, y) = (t, \log Q)$ . Assim

$x = t$	0,5	0,6	0,7	0,8	0,9
$y = \log Q$	0,679427897	0,598790507	0,518513940	0,439332694	0,278753601

$$\begin{cases} a_0 \sum_{i=0}^n 1 + a_1 \sum_{i=0}^n x_i = \sum_{i=0}^n y_i \\ a_0 \sum_{i=0}^n x_i + a_1 \sum_{i=0}^n x_i^2 = \sum_{i=0}^n y_i x_i \end{cases}$$

$$\sum_{i=0}^4 1 = 1 + 1 + 1 + 1 + 1 = 5$$

$$\sum_{i=0}^4 x_i = x_0 + x_1 + x_2 + x_3 + x_4 = 0.5 + 0.6 + 0.7 + 0.8 + 0.9 = 3,5$$

$$\sum_{i=0}^4 x_i^2 = x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 = 0.5^2 + 0.6^2 + 0.7^2 + 0.8^2 + 0.9^2 =$$

$$= 0.25 + 0.36 + 0.49 + 0.64 + 0.81 = 2,55$$

$$\sum_{i=0}^4 y_i = y_0 + y_1 + y_2 + y_3 + y_4 =$$

$$= 0,679427897 + 0,598790507 + 0,518513940 + 0,439332694 + 0,278753601$$

$$= 2,514818638$$

$$\sum_{i=0}^4 y_i x_i = y_0 x_0 + y_1 x_1 + y_2 x_2 + y_3 x_3 + y_4 x_4 =$$

$$= 0,679427897 \cdot 0,5 + 0,598790507 \cdot 0,6 + 0,518513940 \cdot 0,7 +$$

$$+ 0,439332694 \cdot 0,8 + 0,278753601 \cdot 0,9 = 1,664292406$$

Substituindo-se no sistema temos

$$\begin{cases} 5a_0 + 3,5a_1 = 2,514818638 \\ 3,5a_0 + 2,55a_1 = 1,664292406 \end{cases}$$

Resolvendo, obtemos

$$a_0 = 1,175528212 \quad \text{e} \quad a_1 = -0,960806406$$

Mas

$$a_0 = \log a = 1,175528212 \implies a = 10^{a_0} = 10^{1,175528212} = 14,98056564$$

e

$$a_1 = k = -0,960806406$$

Enfim a curva ajustada é

$$Q = a10^{kt} = 14,98056564 \cdot 10^{-0,960806406t}$$

# Chapter 6

## Integração Numérica

### 6.1 Introdução

Calcular integrais definidas de uma dada função real  $f(x)$ ,

$$\int_a^b f(x) dx,$$

é um problema clássico. Para alguns integrandos  $f(x)$  simples, a integral indefinida

$$\int f(x) dx = F(x) + C, \quad F'(x) = f(x)$$

pode ser obtida como uma expressão em  $x$ . Então,

$$\int_a^b f(x) dx = F(b) - F(a).$$

No caso geral, entretanto, integrais definidas são calculadas com métodos numéricos que aproximam a integral por somas finitas associadas a alguma partição do intervalo de integração  $[a, b]$ , esses métodos são chamados de integração numérica ou quadratura numérica.

A forma geral das regras de integração numérica é:

$$\int_a^b f(x) dx \approx \sum_{i=1}^p \omega_i f(x_i)$$

onde

- $x_i$  são chamados pontos de integração;
- $\omega_i$  são chamados pesos dos pontos de integração;
- $p$  é o número de pontos de integração;

#### Ordem de uma regra de integração

Chamamos de ordem de uma regra o grau do polinômio de mais alto grau integrado exatamente pela regra de integração. Por exemplo, se uma regra integra sem erros qualquer polinômio de quinto grau, esta regra tem ordem igual a 5.

## 6.2 Regras de Newton-Cotes

A aproximação das Regras de Newton-Cotes consiste em substituir o integrando  $f(x)$  na integral por um polinômio interpolador, visto que integrar polinômios é uma tarefa fácil. Ou seja, dada uma partição uniforme do intervalo de integração se conhecemos  $(n + 1)$  pontos de uma função  $f(x)$ , a saber,

$$(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))$$

ou

$$(x_i, f(x_i)) \quad i = 0, 1, 2, \dots, n$$

que definem um polinômio interpolador  $p(x)$ , assim temos

$$\int_a^b f(x) dx \approx \int_a^b p(x) dx.$$

Usando polinômios de diferentes graus obteremos diferentes regras de integração.

Na forma de Lagrange o polinômio interpolador é escrito como

$$p(x) = \sum_{i=0}^n f(x_i) L_i(x) = \sum_{i=0}^n f(x_i) \prod_{j=0, j \neq i}^n \frac{(x - x_j)}{(x_i - x_j)}$$

Substituindo na integral do polinômio

$$\int_a^b f(x) dx \approx \int_a^b \left[ \sum_{i=0}^n f(x_i) L_i(x) \right] dx = \sum_{i=0}^n f(x_i) \int_a^b L_i(x) dx$$

Para simplificar a integração do polinômio interpolador podemos utilizar a seguinte mudança de variável  $x = a + ht$ , onde  $h = (b - a)/n$ , temos então que:

$$x = a + ht$$

$$dx = a + h dt$$

$$x_j = a + hj$$

$$x_i = a + hi$$

$$x - x_j = a + ht - a - hj = h(t - j)$$

$$x_i - x_j = a + hi - a - hj = h(i - j)$$

$$a = a + ht \implies t = 0$$

$$b = a + ht \implies t = \frac{b-a}{h} = n$$

logo,

$$\frac{x - x_j}{x_i - x_j} = \frac{h(t - j)}{h(i - j)} = \frac{(t - j)}{(i - j)}$$

$$L_i(x) = \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j} = \varphi_i(t) = \prod_{j=0, j \neq i}^n \frac{t - j}{i - j}$$

A integração do polinômio dará:

$$\begin{aligned} \int_a^b p(x) dx &= \int_{x_0}^{x_n} p(x) dx = \sum_{i=0}^n f(x_i) \int_{x_0}^{x_n} L_i(x) dx = \sum_{i=0}^n f(x_i) \int_0^n \varphi_i(t) h dt = \\ &= h \sum_{i=0}^n f(x_i) \int_0^n \varphi_i(t) dt = h \sum_{i=0}^n f(x_i) \alpha_i \end{aligned}$$

Observe que os coeficientes (ou pesos)

$$\alpha_i = \int_0^n \varphi_i(t) dt$$

dependem somente de  $n$ ; eles não dependem da função  $f(x)$  a ser integrada, nem dos limites de integração  $a$  e  $b$ .

### 6.2.1 Regra dos Trapézios

Esta regra é obtida quando o grau do polinômio interpolador é igual a um ( $n = 1$ ). Logo temos

$$\int_a^b f(x) dx \approx \int_a^b p_1(x) dx = h \sum_{i=0}^1 f(x_i) \alpha_i = h \{f(x_0) \alpha_0 + f(x_1) \alpha_1\}$$

onde  $p_1(x)$  é o polinômio de grau 1 que interpola  $f(x)$  no intervalo  $[a, b]$ . Graficamente,

Vimos que

$$\varphi_i(t) = \prod_{j=0, j \neq i}^n \frac{t - j}{i - j}$$

logo

$$\varphi_0(t) = \prod_{j=0, j \neq 0}^1 \frac{t - j}{0 - j} = \frac{t - 1}{0 - 1} = \frac{t - 1}{-1} = -(t - 1) = 1 - t$$

e

$$\varphi_1(t) = \prod_{j=0, j \neq 1}^1 \frac{t-j}{1-j} = \frac{t-0}{1-0} = \frac{t-0}{1} = t$$

dessa forma

$$\alpha_0 = \int_0^1 (1-t) dt = \left[ t - \frac{t^2}{2} \right]_0^1 = \frac{1}{2}$$

e

$$\alpha_1 = \int_0^1 t dt = \left[ \frac{t^2}{2} \right]_0^1 = \frac{1}{2}$$

Enfim

$$\int_a^b f(x) dx \approx h \{f(x_0)\alpha_0 + f(x_1)\alpha_1\} = h \left\{ f(x_0)\frac{1}{2} + f(x_1)\frac{1}{2} \right\} = \frac{h}{2} \{f(x_0) + f(x_1)\}$$

que a expressão da área de um trapézio com bases iguais a  $f(x_0)$ ,  $f(x_1)$  e altura igual a  $h$ .

Podemos também usar a expressão do polinômio interpolador sem mudança de variável, logo, usando a forma de Lagrange para o polinômio interpolador temos

$$p_1(x) = f(x_0)L_0(x) + f(x_1)L_1(x) = f(x_0)\frac{(x-x_1)}{(x_0-x_1)} + f(x_1)\frac{(x-x_0)}{(x_1-x_0)}$$

onde  $x_0 = a$  e  $x_1 = b$ .

Substituindo  $p_1(x)$  na integral temos

$$\begin{aligned} \int_a^b f(x) dx &\approx \int_{x_0}^{x_1} \left[ f(x_0)\frac{(x-x_1)}{(x_0-x_1)} + f(x_1)\frac{(x-x_0)}{(x_1-x_0)} \right] dx \\ &= \int_{x_0}^{x_1} \left[ f(x_0)\frac{(x-x_1)}{(x_0-x_1)} \right] dx + \int_{x_0}^{x_1} \left[ f(x_1)\frac{(x-x_0)}{(x_1-x_0)} \right] dx \\ &= \frac{f(x_0)}{(x_0-x_1)} \int_{x_0}^{x_1} (x-x_1) dx + \frac{f(x_1)}{(x_1-x_0)} \int_{x_0}^{x_1} (x-x_0) dx \\ &= \frac{f(x_0)}{(x_0-x_1)} \left[ \frac{x^2}{2} - x_1x \right]_{x_0}^{x_1} + \frac{f(x_1)}{(x_1-x_0)} \left[ \frac{x^2}{2} - x_0x \right]_{x_0}^{x_1} \\ &= \frac{f(x_0)}{(x_0-x_1)} \left[ \frac{x_1^2}{2} - x_1x_1 - \frac{x_0^2}{2} + x_1x_0 \right] + \frac{f(x_1)}{(x_1-x_0)} \left[ \frac{x_1^2}{2} - x_0x_1 - \frac{x_0^2}{2} + x_0^2 \right] \\ &= \frac{f(x_0)}{(x_0-x_1)} \left[ -\frac{x_1^2}{2} + x_1x_0 - \frac{x_0^2}{2} \right] + \frac{f(x_1)}{(x_1-x_0)} \left[ \frac{x_1^2}{2} - x_0x_1 + \frac{x_0^2}{2} \right] \\ &= \frac{f(x_0)}{(x_1-x_0)} \left[ \frac{x_1^2}{2} - x_1x_0 + \frac{x_0^2}{2} \right] + \frac{f(x_1)}{(x_1-x_0)} \left[ \frac{x_1^2}{2} - x_0x_1 + \frac{x_0^2}{2} \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{f(x_0)}{2(x_1 - x_0)} [x_1^2 - 2x_1x_0 + x_0^2] + \frac{f(x_1)}{2(x_1 - x_0)} [x_1^2 - 2x_0x_1 + x_0^2] \\
&= \frac{f(x_0)}{2(x_1 - x_0)} [x_1 - x_0]^2 + \frac{f(x_1)}{2(x_1 - x_0)} [x_1 - x_0]^2 \\
&= \frac{f(x_0)}{2} [x_1 - x_0] + \frac{f(x_1)}{2} [x_1 - x_0] = \frac{h}{2} [f(x_0) + f(x_1)]
\end{aligned}$$

Que é a área de um trapézio delimitado por  $p_1(x)$  e pelo intervalo  $[a, b]$ .

Entretanto, se utilizamos somente um trapézio (um único polinômio de primeiro grau) para aproximar a integral podemos cometer erros significativos, como no exemplo abaixo,

Uma maneira de contornar este inconveniente sem ter que aumentar o grau do polinômio interpolador é utilizar a regra dos trapézios em sub-intervalos definidos a partir de uma partição regular do intervalo de integração  $[a, b]$ .

Desta forma dado o número de sub-intervalos  $N$ , cada um teria um comprimento igual a  $h = \frac{b-a}{N}$ . Assim para um sub-intervalo genérico  $[x_i, x_{i+1}]$

$$\int_{x_i}^{x_{i+1}} f(x) dx \approx \int_{x_i}^{x_{i+1}} p_1(x) = \frac{h}{2} [f(x_i) + f(x_{i+1})]$$

Como temos  $N$  sub-intervalos a aproximação para a integral será igual a soma das contribuições elementares de cada trapézio. Isto é,

$$\int_a^b f(x) dx \approx \sum_{i=0}^{N-1} \frac{h}{2} [f(x_i) + f(x_{i+1})] = \frac{h}{2} \sum_{i=0}^{N-1} [f(x_i) + f(x_{i+1})]$$

### 6.2.2 Regra de Simpson

A regra de Simpson é obtida quando o polinômio interpolador tem grau igual a 2. Então,

$$\int_a^b f(x) dx \approx \int_a^b p_2(x) dx$$

onde  $p_2(x)$  é o polinômio de grau 2 que interpola  $f(x)$  no intervalo  $[a, b]$ . Graficamente,

Usando a forma de Lagrange para o polinômio interpolador

$$p(x) = \sum_{i=0}^n f(x_i) \prod_{j=0, j \neq i}^n \frac{(x - x_j)}{(x_i - x_j)}$$

$$p_2(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x)$$

$$p_2(x) = f(x_0) \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f(x_1) \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + f(x_2) \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

onde  $x_0 = a$ ,  $x_1 = (a + b)/2$  e  $x_2 = b$ .

Substituindo na integral do polinômio

$$\int_a^b f(x) dx \approx \int_{x_0}^{x_2} \left[ \sum_{i=0}^2 \prod_{j=0, j \neq i}^2 \frac{x - x_j}{x_i - x_j} \right] dx = \sum_{i=0}^2 \int_{x_0}^{x_2} \left[ \prod_{j=0, j \neq i}^2 \frac{x - x_j}{x_i - x_j} \right] dx$$

Como fizemos na regra dos trapézios podemos utilizar a mudança de variável sugerida.

$$\int_a^b p(x) dx = h \sum_{i=0}^n f(x_i) \int_0^n \varphi_i(t) dt = h \sum_{i=0}^n f(x_i) \alpha_i$$

onde

$$\alpha_i = \int_0^n \varphi_i(t) dt \quad \text{e} \quad \varphi_i(t) = \prod_{j=0, j \neq i}^n \frac{t - j}{i - j}$$

para  $n = 2$

$$\int_a^b p_2(x) dx = h \sum_{i=0}^2 f(x_i) \int_0^2 \varphi_i(t) dt = h \sum_{i=0}^2 f(x_i) \alpha_i$$



e

$$\alpha_i = \int_0^2 \varphi_i(t) dt \quad \text{e} \quad \varphi_i(t) = \prod_{j=0, j \neq i}^2 \frac{t-j}{i-j}$$

Portanto

$$\alpha_0 = \int_0^2 \frac{t-1}{0-1} \frac{t-2}{0-2} dt = \frac{1}{2} \int_0^2 (t^2 - 3t + 2) dt = \frac{1}{2} \left( \frac{8}{3} - \frac{12}{2} + 4 \right) = \frac{1}{3}$$

$$\alpha_1 = \int_0^2 \frac{t-0}{1-0} \frac{t-2}{1-2} dt = - \int_0^2 (t^2 - 2t) dt = - \left( \frac{8}{3} - 4 \right) = \frac{4}{3}$$

$$\alpha_2 = \int_0^2 \frac{t-0}{2-0} \frac{t-1}{2-1} dt = \frac{1}{2} \int_0^2 (t^2 - t) dt = \frac{1}{2} \left( \frac{8}{3} - \frac{4}{2} \right) = \frac{1}{3}$$

que fornece o seguinte valor aproximado para a integral

$$\int_a^b f(x) dx \approx \int_a^b p_2(x) dx = \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)]$$

Esta é a chamada *Regra de Simpson*.

Como na regra dos trapézios se utilizarmos somente um polinômio de segundo grau podemos cometer grandes erros, como mostra a figura abaixo.

Logo, para diminuir este tipo de erro devemos usar esta regra de forma repetida, semelhantemente ao proposto na regra dos trapézios. Aplicando a regra de Simpson a pares de sub-intervalos

definidos a partir de uma partição regular do intervalo de integração  $[a, b]$ .

Em um par de sub-intervalos genérico  $[x_{2i}, x_{2i+1}, x_{2i+2}]$  a contribuição elementar é

$$I_i = \frac{h}{3} [f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})]$$

A aproximação para o valor da integral é igual a soma das aproximações elementares obtidas para cada par de sub-intervalos. Isto é,

$$\int_a^b f(x) dx \approx \frac{h}{3} \sum_{i=0}^{\frac{N}{2}-1} [f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})]$$

**Exemplos:**

## Erro nas Fórmulas de Newton-Cotes

### Trapézios

Da interpolação polinomial, no caso linear,

$$f(x) - p(x) = (x - x_0)(x - x_1) \frac{f''(\alpha)}{2!} \quad \alpha \in (x_0, x_1)$$

ou

$$f(x) = p(x) + (x - x_0)(x - x_1) \frac{f''(\alpha)}{2!} \quad \alpha \in (x_0, x_1)$$

Integrando de  $x_0$  a  $x_1$  temos

$$\int_a^b f(x) dx = \int_{x_0}^{x_1} f(x) dx = \int_{x_0}^{x_1} p(x) dx + \int_{x_0}^{x_1} (x - x_0)(x - x_1) \frac{f''(\alpha)}{2!} dx$$

ou

$$\int_{x_0}^{x_1} f(x) dx - \int_{x_0}^{x_1} p(x) dx = \int_{x_0}^{x_1} (x - x_0)(x - x_1) \frac{f''(\alpha)}{2!} dx$$

ou ainda

$$E_T = \int_{x_0}^{x_1} (x - x_0)(x - x_1) \frac{f''(\alpha)}{2!} dx = \int_{x_0}^{x_1} \frac{\omega(x)}{2} f''(\alpha) dx = \frac{1}{2} \int_{x_0}^{x_1} \omega(x) f''(\alpha) dx$$

Note que  $\forall x \in (x_0, x_1)$ ,  $\omega(x) < 0$ , e se  $f''(x)$  é contínua em  $[x_0, x_1]$  existem  $m$  e  $M$ , tais que

$$m \leq f''(x) \leq M$$

e então

$$m \leq f''(\alpha) \leq M$$

como  $\omega(x) < 0$  então:

$$m\omega(x) \geq \omega(x)f''(\alpha) \geq M\omega(x)$$

Integrando-se

$$\int_{x_0}^{x_1} m\omega(x) dx \geq \int_{x_0}^{x_1} \omega(x)f''(\alpha) dx \geq \int_{x_0}^{x_1} M\omega(x) dx$$

ou

$$M \int_{x_0}^{x_1} \omega(x) dx \leq \int_{x_0}^{x_1} \omega(x)f''(\alpha) dx \leq m \int_{x_0}^{x_1} \omega(x) dx \quad (*)$$

Como  $\omega(x) < 0$ , então:

$$\int_{x_0}^{x_1} \omega(x) dx < 0$$

Dividindo-se as desigualdades (\*) por

$$\int_{x_0}^{x_1} \omega(x) dx$$

Resulta

$$m \leq \left[ \frac{\int_{x_0}^{x_1} \omega(x) f''(\alpha) dx}{\int_{x_0}^{x_1} \omega(x) dx} \right] \leq M$$

Do Teorema do Valor Médio para Integrais, temos que existe  $c \in (x_0, x_1)$  tal que

$$\int_{x_0}^{x_1} \omega(x) f''(x) dx = f''(c) \int_{x_0}^{x_1} \omega(x) dx$$

Logo,

$$E_T = \frac{1}{2} \int_{x_0}^{x_1} \omega(x) f''(\alpha) dx = \frac{1}{2} f''(c) \int_{x_0}^{x_1} \omega(x) dx, \quad c \in (x_0, x_1)$$

mas

$$\begin{aligned} \int_{x_0}^{x_1} \omega(x) dx &= \int_{x_0}^{x_1} (x - x_0)(x - x_1) dx = \int_{x_0}^{x_1} [x^2 - (x_0 + x_1)x + x_0x_1] dx = \\ &= \left[ \frac{x^3}{3} - (x_0 + x_1) \frac{x^2}{2} + x_0x_1x \right]_{x_0}^{x_1} = \frac{x_1^3}{3} - (x_0 + x_1) \frac{x_1^2}{2} + x_0x_1^2 - \frac{x_0^3}{3} + (x_0 + x_1) \frac{x_0^2}{2} - x_0^2x_1 = \\ &= \frac{2x_1^3 - 3(x_0 + x_1)x_1^2 + 6x_0x_1^2 - 2x_0^3 + 3(x_0 + x_1)x_0^2 - 6x_0^2x_1}{6} = \\ &= \frac{1}{6} [2x_1^3 - 3x_0x_1^2 - 3x_1^3 + 6x_0x_1^2 - 2x_0^3 + 3x_0^3 + 3x_1x_0^2 - 6x_0^2x_1] = \\ &= \frac{1}{6} [-x_1^3 + 3x_0x_1^2 - 3x_0^2x_1 + x_0^3] = -\frac{(x_1 - x_0)^3}{6} = -\frac{h^3}{6} \end{aligned}$$

Logo

$$E_T = \frac{1}{2} f''(c) \int_{x_0}^{x_1} \omega(x) dx = \frac{1}{2} f''(c) \left[ -\frac{h^3}{6} \right]$$

$$E_T = -\frac{h^3}{12} f''(c)$$

Se utilizarmos a regra dos trapézios de forma repetida, para calcular o erro total devemos somar as contribuições elementares do erro para cada sub-intervalo, assim

$$E_T = \sum_{i=0}^{N-1} \left[ -\frac{h^3}{12} f''(c_i) \right]; \quad c_i \in (x_i, x_{i+1})$$

Se  $f''(x)$  é contínua em  $[a, b]$ , podemos dizer que existe um ponto  $\alpha \in (a, b)$  tal que

$$f''(\alpha) = \frac{\sum_{i=0}^{N-1} f''(c_i)}{N}$$

Logo,

$$E_T = \sum_{i=0}^{N-1} \left[ -\frac{h^3}{12} f''(c_i) \right] = -\frac{h^3}{12} \sum_{i=0}^{N-1} f''(c_i) = -\frac{h^3}{12} N f''(\alpha) \quad \alpha \in (a, b)$$

ou

$$E_T = -\frac{h^2}{12} \frac{(b-a)}{N} N f''(\alpha) = -\frac{h^2}{12} (b-a) f''(\alpha)$$

ou

$$|E_T| = \frac{h^2}{12} (b-a) |f''(\alpha)| \quad \alpha \in (a, b)$$

## Simpson

Procedendo de forma similar para a Regra de Simpson obtemos

$$E_S = -\frac{h^5 N}{90} f^{(4)}(\alpha) \quad \alpha \in (a, b)$$

ou

$$|E_S| = \frac{h^4}{180} (b-a) |f^{(4)}(\alpha)| \quad \alpha \in (a, b)$$

## Resumindo

Trapézios:

$$|E_T| = \frac{h^2}{12} (b-a) |f''(\alpha)| \quad \alpha \in (a, b)$$

Simpson:

$$|E_S| = \frac{h^4}{180} (b-a) |f^{(4)}(\alpha)| \quad \alpha \in (a, b)$$

Entretanto as expressões acima não são facilmente utilizadas devido ao desconhecimento do ponto  $\alpha$ . Logo podemos usar estas estimativas para determinar uma cota superior para o erro, isto é, Trapézios:

$$|E_T| \leq \frac{h^2}{12} (b-a) M_2$$

onde

$$M_2 = \max_{a \leq x \leq b} |f''(x)|$$

Simpson:

$$|E_S| \leq \frac{h^4}{180} (b-a) M_4$$

onde

$$M_4 = \max_{a \leq x \leq b} |f^{(4)}(x)|$$

## Observação:

Outra maneira de estimar se cometemos um grande erro numa regra de integração é calcularmos a integral com  $m$  sub-intervalos e a seguir aumentarmos o número de sub-intervalos, por exemplo  $2m$ , e compararmos os valores obtidos no dois casos, se a diferença é grande devemos refinar ainda mais a partição, senão usamos o resultado obtido como uma boa aproximação da integral.

### 6.3 Quadratura Gaussiana

As regras de Newton-Cotes que estudamos, Trapézios e Simpson, integram exatamente polinômios de grau 1 e 3, respectivamente. A regra dos trapézios usa para isto 2 pontos de integração e a regra de Simpson 3 pontos de integração. Queremos agora encontrar regras de integração tais que usando-se  $p$  pontos de integração calcula-se exatamente a integral de um polinômio de grau  $2p - 1$ . Para isto não mais exigimos que os pontos de integração sejam conhecidos e igualmente espaçados como nas fórmulas de Newton-Cotes. Já vimos que a fórmula geral de uma regra de integração numérica é:

$$\int_a^b f(x) dx \approx \sum_{i=1}^p \omega_i f(x_i) \quad (6.1)$$

onde:  $x_i$  e  $\omega_i$  são respectivamente os pontos e pesos de integração.

Nas regras de Newton-Cotes os pontos de integração são determinados *a priori* segundo uma partição uniforme definida sobre o intervalo de integração, e partir dos pontos de integração calcula-se os pesos associados a cada ponto.

Nas regras gaussianas como não conhecemos os pontos nem os pesos de integração temos que determiná-los simultaneamente.

Podemos construir estas regras de várias formas, a mais intuitiva é descrita a seguir.

Inicialmente nos restringiremos ao intervalo de integração  $[-1, 1]$ .

#### Regra de Gauss com um ponto de integração ( $p = 1$ )

$$\int_{-1}^1 f(x) dx \approx \omega_1 f(x_1)$$

Para que a regra acima seja exata para polinômios de grau  $2p - 1 = 2 \cdot 1 - 1$  temos que

$$\int_{-1}^1 (a_0 + a_1 x) dx = \omega_1 (a_0 + a_1 x_1)$$

ou

$$\int_{-1}^1 a_0 dx + \int_{-1}^1 a_1 x dx = \omega_1 (a_0 + a_1 x_1)$$

logo

$$a_0 \int_{-1}^1 1 dx + a_1 \int_{-1}^1 x dx = \omega_1 (a_0 + a_1 x_1)$$

$$2a_0 + 0a_1 = \omega_1 a_0 + \omega_1 a_1 x_1$$

como deve ser verdadeira para quaisquer  $a_0$  e  $a_1$ , temos

$$\omega_1 = 2$$

e

$$\omega_1 x_1 = 0 \quad \longrightarrow \quad x_1 = 0$$

Logo,

$$\int_{-1}^1 f(x) dx \approx \omega_1 f(x_1) = 2f(0)$$

**Regra de Gauss com 2 pontos de integração ( $p = 2$ )**

$$\int_{-1}^1 f(x) dx \approx \omega_1 f(x_1) + \omega_2 f(x_2)$$

Neste caso queremos que esta regra integre exatamente polinômios até grau  $2p - 1 = 2 \cdot 2 - 1 = 3$ .

Com  $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ . Neste caso:

$$\int_{-1}^1 p(x) dx = \omega_1 p(x_1) + \omega_2 p(x_2)$$

$$\int_{-1}^1 (a_0 + a_1x + a_2x^2 + a_3x^3) dx = \omega_1 p(x_1) + \omega_2 p(x_2)$$

ou

$$a_0 \int_{-1}^1 1 dx + a_1 \int_{-1}^1 x dx + a_2 \int_{-1}^1 x^2 dx + a_3 \int_{-1}^1 x^3 dx = \omega_1 p(x_1) + \omega_2 p(x_2)$$

$$a_0 2 + a_1 0 + a_2 \frac{2}{3} + a_3 0 = \omega_1 [a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3] + \omega_2 [a_0 + a_1 x_2 + a_2 x_2^2 + a_3 x_2^3]$$

$$\begin{aligned} 2a_0 + 0a_1 + \frac{2}{3}a_2 + 0a_3 &= a_0(\omega_1 + \omega_2) + a_1(\omega_1 x_1 + \omega_2 x_2) + a_2(\omega_1 x_1^2 + \omega_2 x_2^2) \\ &\quad + a_3(\omega_1 x_1^3 + \omega_2 x_2^3) \end{aligned}$$

Como a igualdade acima deve ser válida para qualquer polinômio  $p(x)$ , temos o seguinte sistema não linear:

$$\left\{ \begin{array}{l} \omega_1 + \omega_2 = 2 \\ \omega_1 x_1 + \omega_2 x_2 = 0 \\ \omega_1 x_1^2 + \omega_2 x_2^2 = \frac{2}{3} \\ \omega_1 x_1^3 + \omega_2 x_2^3 = 0 \end{array} \right.$$

Resolvendo o sistema não-linear acima obtemos:

$$\omega_1 = 1, \omega_2 = 1, x_1 = -1/\sqrt{3} \text{ e } x_2 = 1/\sqrt{3}.$$

Em geral para determinar uma regra de quadratura gaussiana de  $p$  pontos temos que resolver um sistema não linear de ordem  $2p$ . Na verdade, os pesos e pontos de integração são determinados de outras formas que não envolvem a resolução de sistemas não-lineares, e são calculados através de raízes de polinômios ou resolvendo-se um problema de autovalor algébrico.

$$E_{Gauss} = \frac{2^{2p+1}(p!)^4}{(2p+1)[(2p)!]^3} f^{(2p)}(\xi) \quad \xi \in [-1, 1]$$

Somente deduzimos as regras de integração para um intervalo  $[-1, 1]$ . Se quisermos aplicar estas regras à integrais de  $a$  até  $b$  precisamos de uma transformação de  $[a, b]$  para  $[-1, 1]$ .

Temos, então,

$$\int_a^b f(t) dt \longrightarrow \int_{-1}^1 \mathcal{F}(x) dx$$

usando a seguinte relação:

$$t = \frac{b-a}{2}x + \frac{a+b}{2}$$

observe que para  $x = -1 \rightarrow t = a$  e para  $x = 1 \rightarrow t = b$ , logo,

$$\frac{dt}{dx} = \frac{b-a}{2} \quad \text{e} \quad dt = \frac{b-a}{2} dx$$

substituindo na integral

$$\int_a^b f(t) dt = \int_{-1}^1 f\left(\frac{b-a}{2}x + \frac{a+b}{2}\right) \frac{b-a}{2} dx = \int_{-1}^1 \mathcal{F}(x) dx$$

onde

$$\mathcal{F}(x) = \frac{b-a}{2} f\left(\frac{b-a}{2}x + \frac{a+b}{2}\right)$$

A seguir apresentamos uma tabela com alguns pontos e pesos de integração das regras de quadratura gaussiana.



Tabela — Abcissas e pesos das regras gaussianas

Abcissas	Pesos
p = 1	
0,00000 00000 00000	2,00000 00000 00000
p = 2	
+/-0,57735 02691 89626	1,00000 00000 00000
p = 3	
0,00000 00000 00000	0,88888 88888 88889
+/-0,77459 66692 41483	0,55555 55555 55556
p = 4	
+/-0,33998 10435 84856	0,65214 51548 62546
+/-0,86113 63115 94053	0,34785 48451 37454
p = 5	
0,00000 00000 00000	0,56888 88888 88889
+/-0,50846 93101 05683	0,47862 86704 99366
+/-0,90617 97459 38664	0,23692 68850 56189
p = 6	
+/-0,23861 91860 83197	0,46791 39345 72691
+/-0,66120 93864 66265	0,36076 15730 48139
+/-0,93246 95142 03152	0,17132 44923 79170
p = 7	
0,00000 00000 00000	0,41795 91836 73469
+/-0,40584 51513 77397	0,38183 00505 05119
+/-0,74153 11855 99394	0,27970 53914 89277
+/-0,94910 79123 42759	0,12948 49661 68870
p = 8	
+/-0,18343 46424 95650	0,36268 37833 78362
+/-0,52553 24099 16329	0,31370 66458 77887
+/-0,79666 64774 13627	0,22238 10344 53374
+/-0,96028 98564 97536	0,10122 85362 90376

**Exemplo:**

**Observações Finais:**

- Todas as regras que aqui estudamos podem ser aplicadas às integrais duplas ou triplas. Basta usar uma regra para cada variável de integração.
- Devemos ajustar as calculadoras para radianos quando utilizamos as funções trigonométricas e as regras numéricas de integração.

## Integração Numérica – Exercícios

1. Calcule as integrais abaixo pela Regra dos Trápezios e pela Regra de Simpson, usando quatro e seis divisões do intervalo de integração.

$$(a) \quad \int_1^2 e^x dx$$

$$(b) \quad \int_1^4 \sqrt{x} dx$$

$$(c) \quad \int_2^{14} \frac{dx}{\sqrt{x}}$$

2. Avalie a integral abaixo usando a regra dos trapézios com 5 sub-intervalos.

$$\int_0^1 \text{sen}(4x + 1) dx$$

3. Qual o erro máximo cometido na aproximação de

$$\int_0^4 (3x^3 - 3x + 1) dx$$

pela Regra de Simpson com quatro sub-intervalos?

Calcule a integral pela Regra dos Trapézios e compare os resultados.

4. Determinar  $h$ , a distância entre  $x_i$  e  $x_{i+1}$ , para que se possa avaliar

$$\int_0^{\pi/2} \cos x dx$$

com erro inferior a  $10^{-3}$  pela Regra de Simpson. Avalie esta integral por Simpson usando o número mínimo de sub-intervalos em função da análise acima. Calcule também esta integral por Quadratura Gaussiana com 3 pontos e compare os resultados obtidos com o valor exato.

**Solução:**

$$|E_S| \leq \frac{h^4}{180} (b - a) M_4$$

$$M_4 = \max_{0 \leq x \leq \pi/2} |f^{(iv)}(x)|$$

$$f^{(iv)}(x) = \cos x$$

logo

$$M_4 = 1 \quad \text{em} \quad \frac{\pi}{2}$$

$$10^{-3} = \frac{h^4}{180} \left( \frac{\pi}{a} - 0 \right) 1$$

$$h^4 = \frac{180 \times 10^{-3}}{\pi/2}$$

$$h = 0,581819345$$

mas

$$h = \frac{b-a}{n}$$

e

$$n \frac{b-a}{h} = \frac{\pi}{2h} = 2,699800789$$

como  $n$  tem que ser um número par faremos  $n = 4$ , assim,

$$h = \frac{b-a}{n} = \frac{\pi}{8}$$

logo,

$$x_0 = 0, \quad x_1 = \frac{\pi}{8}, \quad x_2 = \frac{\pi}{4}, \quad x_3 = \frac{3\pi}{8}, \quad x_4 = \frac{\pi}{2}.$$

$$\int_0^{\pi/2} \cos x \, dx \approx \frac{h}{3} [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + f(x_4)] =$$

$$\frac{\pi}{24} [\cos(0) + 4 \cos(\pi/8) + 2 \cos(\pi/4) + 4 \cos(3\pi/8) + \cos(\pi/2)] =$$

$$\frac{\pi}{24} [1,0000 + 3,695518130 + 1,414213562 + 1,530733729 + 0] =$$

$$= 1,000134585$$

5. Calcule pelo método dos trapézios

$$\int_0^{10} f(x) \, dx$$

onde  $f(x)$  é a função tabelada abaixo.

$x$	$f(x)$
0	4,1
1	6,5
2	9,0
7	13,2
8	17,1
10	20,1

6. Deduza uma fórmula de integração da forma,

$$\int_{-1}^1 f(x) dx \approx \omega_0 f(-0,5) + \omega_1 f(0) + \omega_2 f(0,5)$$

que integre exatamente polinômios de grau  $\leq 2$ .

7. Considere a integral

$$\int_0^{10} e^{-x} dx = 0,999955$$

Calcule por:

- (a) Quadratura gaussiana com 6 pontos;
- (b) Regra dos trapézios com 10 sub-intervalos;
- (c) Regra de Simpson com 10 sub-intervalos;

Compare os resultados.

8. Integre o polinômio de interpolação linear de  $x_i$  a  $x_{i+1}$  e mostre que o resultado é a Regra dos Trapézios para integração numérica.

9. Calcule o valor aproximado de

$$\int_0^{0,6} \frac{dx}{1+x}$$

com três casas decimais de precisão usando o número mínimo de subintervalos e a Regra de:

- a) Simpson;
- b) Trapézios.

### Solução:

Como queremos três casas decimais de precisão logo o erro deve ser menor que 0,001.

a) Simpson;

$$|E_S| \leq \frac{h^4}{180}(b-a)M_4$$

onde

$$M_4 = \max |f^{(4)}(x)|, \quad a \leq x \leq b$$

$$f(x) = \frac{1}{x+1}$$

$$f^{(1)}(x) = -\frac{1}{(x+1)^2}$$

$$f^{(2)}(x) = 2\frac{1}{(x+1)^3}$$

$$f^{(3)}(x) = -6\frac{1}{(x+1)^4}$$

$$f^{(4)}(x) = 24\frac{1}{(x+1)^5}$$

$$M_4 = \max \left| 24\frac{1}{(x+1)^5} \right|, \quad 0 \leq x \leq 0,6$$

mas o máximo de  $f^{(4)}(x)$  entre 0 e 0,6 se dá em  $x = 0$ . Portanto,

$$M_4 = 24$$

substituindo na expressão para o erro

$$|E_S| \leq \frac{h^4}{180}(b-a)M_4 = \frac{h^4}{180}(0,6-0) \cdot 24$$

na situação mais desfavorável teremos a igualdade

$$|E_S| = \frac{h^4}{180}(0,6-0) \cdot 24$$

mas queremos que

$$|E_S| < 0,001$$

ou

$$\frac{h^4}{180}(0,6-0) \cdot 24 < 0,001$$

$$h^4 < \frac{0,001 \cdot 180}{0,6 \cdot 24}$$

$$h^4 < 0,0125$$

$$h < 0,334370153$$

mas

$$h = \frac{b-a}{n}$$

ou

$$n = \frac{b-a}{h} = \frac{0,6-0}{0,334370153} = 1,794418537$$

Devemos usar um número de divisões para o intervalo de integração maior que 1,794418537. Portanto usaremos  $n = 2$ .

$$\int_0^{0,6} \frac{dx}{1+x} \approx \sum_{i=0}^{n/2-1} \frac{h}{3} [f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})]$$

$$h = \frac{b-a}{n} = \frac{0,6-0}{2} = 0,3$$

$$x_i = a + ih = 0 + 0,3i = 0,3i, \quad i = 0, 1, 2$$

$$x_0 = 0 \quad x_1 = 0,3 \quad x_2 = 0,6$$

$$\begin{aligned} \int_0^{0,6} \frac{dx}{1+x} &\approx \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)] = \frac{h}{3} \left[ \frac{1}{1+x_0} + 4\frac{1}{1+x_1} + \frac{1}{1+x_2} \right] \\ &= \frac{0,3}{3} \left[ \frac{1}{1+0} + 4\frac{1}{1+0,3} + \frac{1}{1+0,6} \right] = \frac{0,3}{3} \left[ \frac{1}{1} + 4\frac{1}{1,3} + \frac{1}{1,6} \right] \\ &= 0,4701923 \end{aligned}$$

a) Trapézios;

$$|E_T| \leq \frac{h^2}{12}(b-a)M_2$$

onde

$$M_2 = \max |f^{(2)}(x)|, \quad a \leq x \leq b$$

$$f(x) = \frac{1}{x+1}$$

$$f^{(1)}(x) = -\frac{1}{(x+1)^2}$$

$$f^{(2)}(x) = 2\frac{1}{(x+1)^3}$$

$$M_2 = \max \left| 2\frac{1}{(x+1)^3} \right|, \quad 0 \leq x \leq 0,6$$

mas o máximo de  $f^{(2)}(x)$  entre 0 e 0,6 se dá em  $x = 0$ . Portanto,

$$M_2 = 2$$

substituindo na expressão para o erro

$$|E_T| \leq \frac{h^2}{12}(b-a)M_2 = \frac{h^2}{12}(0,6-0) \cdot 2$$

na situação mais desfavorável teremos a igualdade

$$|E_T| = \frac{h^2}{12}(0,6-0) \cdot 2$$

mas queremos que

$$|E_T| < 0,001$$

ou

$$\frac{h^2}{12}(0,6-0) \cdot 2 < 0,001$$

$$h^2 < \frac{0,001 \cdot 12}{0,6 \cdot 2}$$

$$h^2 < 0,01$$

$$h < 0,1$$

mas

$$h = \frac{b-a}{n}$$

ou

$$n = \frac{b-a}{h} = \frac{0,6-0}{0,1} = 6$$

Portanto usaremos  $n = 6$ .

$$\int_0^{0,6} \frac{dx}{1+x} \approx \sum_{i=0}^{n-1} \frac{h}{2} [f(x_i) + f(x_{i+1})]$$

$$h = \frac{b-a}{n} = \frac{0,6-0}{6} = 0,1$$

$$x_i = a + ih = 0 + 0,1i = 0,1i, \quad i = 0, 1, 2, 3, 4, 5, 6$$

$$x_0 = 0,0 \rightarrow x_1 = 0,1 \rightarrow x_2 = 0,2 \rightarrow x_3 = 0,3 \rightarrow x_4 = 0,4 \rightarrow x_5 = 0,5 \rightarrow x_6 = 0,6$$

$$\begin{aligned} \int_0^{0,6} \frac{dx}{1+x} \approx \frac{h}{2} [f(x_0) + f(x_1) + f(x_1) + f(x_2) + f(x_2) + f(x_3) + f(x_3) + f(x_4) \\ + f(x_4) + f(x_5) + f(x_5) + f(x_6)] \end{aligned}$$



$$\begin{aligned}
\int_0^{0,6} \frac{dx}{1+x} &\approx \frac{h}{2} [f(x_0) + 2f(x_1) + 2f(x_2) + 2f(x_3) + 2f(x_4) + 2f(x_5) + f(x_6)] \\
&= \frac{h}{2} \left[ \frac{1}{1+x_0} + 2\frac{1}{1+x_1} + 2\frac{1}{1+x_2} + 2\frac{1}{1+x_3} + 2\frac{1}{1+x_4} + 2\frac{1}{1+x_5} + \frac{1}{1+x_6} \right] \\
&= \frac{0,1}{2} \left[ \frac{1}{1+0} + 2\frac{1}{1+0,1} + 2\frac{1}{1+0,2} + 2\frac{1}{1+0,3} + 2\frac{1}{1+0,4} + 2\frac{1}{1+0,5} + \frac{1}{1+0,6} \right] \\
&= \frac{0,1}{2} \left[ \frac{1}{1} + 2\frac{1}{1,1} + 2\frac{1}{1,2} + 2\frac{1}{1,3} + 2\frac{1}{1,4} + 2\frac{1}{1,5} + \frac{1}{1,6} \right] \\
&= \frac{0,1}{2} \left[ \frac{1}{1} + 2 \left( \frac{1}{1,1} + \frac{1}{1,2} + \frac{1}{1,3} + \frac{1}{1,4} + \frac{1}{1,5} \right) + \frac{1}{1,6} \right] \\
&= 0,470510739
\end{aligned}$$

Resumindo temos

$$\int_0^{0,6} \frac{dx}{1+x} \approx 0,4701923 \quad (\text{Simpson } c/ 2 \text{ sub-intervalos})$$

$$\int_0^{0,6} \frac{dx}{1+x} \approx 0,470510739 \quad (\text{Trapézios } c/ 6 \text{ sub-intervalos})$$

$$\int_0^{0,6} \frac{dx}{1+x} = \ln(1+x)|_0^{0,6} = \ln 1,6 - \ln 1,0 = 0,470003629 \quad (\text{Exato})$$

10. Dada a integral abaixo escolha os pesos e pontos de integração de Gauss para cada variável de forma que a regra calcule exatamente o valor da integral. A seguir use sua escolha para calcular a integral.

$$\int_{-1}^1 \int_{-1}^1 \{x^2y^3 + x + xy\} dx dy$$

### Solução:

Nas regras de Gauss com  $p$  pontos integramos exatamente polinômios de grau  $2p - 1$ . Como a função dada acima é um polinômio podemos integrá-la exatamente.

Para a variável  $y$  temos um polinômio de grau 3. Logo temos que usar um regra de pelo menos 2 pontos de Gauss, posto que tal regra integra exatamente polinômios de grau 3 ( $2p - 1 = 2 \cdot 2 - 1 = 3$ ). Logo,

$$\begin{aligned}
&\int_{-1}^1 \int_{-1}^1 \{x^2y^3 + x + xy\} dx dy = \\
&= \int_{-1}^1 \sum_{i=1}^2 \omega_i \{x^2y_i^3 + x + xy_i\} dx =
\end{aligned}$$

$$= \int_{-1}^1 \left\{ \omega_1 [x^2 y_1^3 + x + x y_1] + \omega_2 [x^2 y_2^3 + x + x y_2] \right\} dx =$$

Usando os dados da tabela para  $p = 2$

$$= \int_{-1}^1 \left\{ -x^2 \left( \frac{\sqrt{3}}{3} \right)^3 + x - x \left( \frac{\sqrt{3}}{3} \right) + x^2 \left( \frac{\sqrt{3}}{3} \right)^3 + x + x \left( \frac{\sqrt{3}}{3} \right) \right\} dx =$$

$$= \int_{-1}^1 \{2x\} dx =$$

Para a variável  $x$  temos agora um polinômio de grau 1. Logo podemos usar um regra de 1 ponto de Gauss, que integra exatamente polinômios de grau 1 ( $2p - 1 = 2 \cdot 1 - 1 = 1$ ). Logo,

$$= \int_{-1}^1 \{2x\} dx = \sum_{i=1} 1 \omega_i(2x_i) = \omega_1(2x_1) = 2 \cdot (2 \cdot 0) = 0$$

11. Qual o erro máximo cometido na aproximação de  $\int_0^4 (3x^3 - 3x + 1) dx$  pela Regra de Simpson com quatro subintervalos?

Calcule por Trapézios e compare os resultados.

12. Deduza uma fórmula de integração para

$$I = \int_a^b f(x) dx$$

dividindo o intervalo em  $n$  sub-intervalos iguais de largura

$$h = \frac{b - a}{n}$$

Aproxime a integral para cada sub-intervalo pela área do retângulo cuja altura é o valor de  $f(x)$  na extremidade *esquerda* de cada sub-intervalo. A seguir use esta regra para aproximar

$$\int_0^{\pi/2} \cos x dx$$

com  $n = 4$ .

### Solução:

Teremos então  $n$  contribuições de retângulos, uma para cada sub-intervalo.

Para um único sub-intervalo temos:

$$\text{Área de cada retângulo} = h \cdot f(x_i); \quad i = 0, 1, 2, \dots, n - 1$$

A integral será aproximada pela soma das áreas de cada retângulo, isto é:

$$\int_a^b f(x) dx = \sum_{i=0}^{n-1} h \cdot f(x_i) = h \sum_{i=0}^{n-1} f(x_i)$$

Logo para a integral

$$\int_0^{\pi/2} \cos x \, dx$$

com  $n = 4$ .

temos,

$$\int_0^{\pi/2} \cos x \, dx = h \sum_{i=0}^3 f(x_i)$$

onde

$$h = \frac{b-a}{n} = \frac{\pi/2 - 0}{4} = \frac{\pi}{8}$$

e

$$\begin{aligned} h \sum_{i=0}^3 f(x_i) &= h [f(x_0) + f(x_1) + f(x_2) + f(x_3)] = \\ &= h [\cos x_0 + \cos x_1 + \cos x_2 + \cos x_3] = \\ &= \frac{\pi}{8} \left[ \cos 0 + \cos \frac{\pi}{8} + \cos \frac{\pi}{4} + \cos \frac{3\pi}{8} + \cos \frac{\pi}{2} \right] = \\ &= \frac{\pi}{8} [1 + 0,923879533 + 0,7071067811 + 0,382683432 + 0] = \\ &= 1,183465342 \end{aligned}$$

13. Calcule a integral abaixo pelo método dos trapézios com com 4 sub-intervalos.

$$\int_2^3 \frac{1}{x^2} dx$$

14. Uma das maneiras de calcular o logaritmo natural de um número  $N$  é calculando a integral

$$\int_1^N \frac{dx}{x}$$

usando integração numérica, porque, para isso, só é preciso calcular valores da função  $f(x) = 1/x$ . Se for usado o Método de Simpson, pede-se o número mínimo de subintervalos a dividir o intervalo de integração, para que tenhamos  $\ln N$  com erro inferior a  $10^{-4}$  em função de  $N$ .

**Solução:**

O erro na Regra de Simpson é dado por

$$E_S = \frac{h^4}{180} (b-a) f^{(4)}(\xi) \quad a \leq \xi \leq b$$

Neste caso temos

$$a = 1$$

$$b = N$$

$$h = \frac{b-a}{n} = \frac{N-1}{n}$$

Ou ainda

$$|E_S| \leq \frac{h^4}{180}(b-a)M_4$$

onde

$$M_4 = \max_{a \leq x \leq b} |f^{(4)}(x)|$$

mas

$$f(x) = \frac{1}{x}$$

$$f'(x) = -\frac{1}{x^2}$$

$$f''(x) = \frac{2}{x^3}$$

$$f^{(3)}(x) = -\frac{6}{x^4}$$

$$f^{(4)}(x) = \frac{24}{x^5}$$

portanto

$$M_4 = \max_{1 \leq x \leq N} \left| \frac{24}{x^5} \right| = 24$$

A expressão para o erro toma então a seguinte forma:

$$|E_S| \leq \frac{h^4}{180}(b-a)M_4 = \left(\frac{N-1}{n}\right)^4 \frac{(N-1) \cdot 24}{180} = \frac{(N-1)^5}{n^4} \frac{24}{180}$$

No caso limite teríamos a igualdade

$$|E_S| = \frac{(N-1)^5}{n^4} \frac{24}{180}$$

ou

$$10^{-4} = \frac{(N-1)^5}{n^4} \frac{24}{180}$$

logo

$$n^4 = \frac{(N-1)^5}{10^{-4}} \frac{24}{180} = 1333.333333333(N-1)^5$$

e

$$n = 6.042750794(N-1)^{5/4}$$

Finalmente

$$n \geq 6.042750794(N-1)^{5/4}$$

15. Para a função tabelada abaixo, aproxime  $\int_1^5 f(x) dx$  pela Regra de Simpson.

$x$	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
$f(x)$	0,00	0,41	0,69	0,92	1,10	1,25	1,39	1,50	1,61

**Solução:**

Vemos que as abcissas estão uniformemente espaçadas no intervalo  $[1, 5]$ , logo podemos utilizar diretamente a Regra de Simpson com  $h = 0.5$ . Temos então:

$$\int_a^b f(x) dx \approx \frac{h}{3} \sum_{i=0}^{\frac{n}{2}-1} (f(x_{2i}) + f(x_{2i+1}) + f(x_{2i+2}))$$

Neste problema,

$$\int_1^5 f(x) dx \approx \frac{h}{3} \sum_{i=0}^2 \{f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})\} = \frac{h}{3} \{f(x_0) + 4f(x_1) + 2f(x_2) +$$

$$+4f(x_3) + 2f(x_4) + 4f(x_5) + 2f(x_6) + 4f(x_7) + f(x_8)\}$$

$$\int_1^5 f(x) dx \approx \frac{0,5}{3} \{0,00 + 4 \cdot 0,41 + 2 \cdot 0,69 + 4 \cdot 0,92 + 2 \cdot 1,10 + 4 \cdot 1,25 + 2 \cdot 1,39 +$$

$$+4 \cdot 1,50 + 1,61\} = \frac{0,5}{3} \{24,29\} = 4,048333333$$

Tabela — Abcissas e pesos das regras gaussianas

Abcissas	Pesos
p = 1	
0,00000 00000 00000	2,00000 00000 00000
p = 2	
+/-0,57735 02691 89626	1,00000 00000 00000
p = 3	
0,00000 00000 00000	0,88888 88888 88889
+/-0,77459 66692 41483	0,55555 55555 55556
p = 4	
+/-0,33998 10435 84856	0,65214 51548 62546
+/-0,86113 63115 94053	0,34785 48451 37454
p = 5	
0,00000 00000 00000	0,56888 88888 88889
+/-0,50846 93101 05683	0,47862 86704 99366
+/-0,90617 97459 38664	0,23692 68850 56189
p = 6	
+/-0,23861 91860 83197	0,46791 39345 72691
+/-0,66120 93864 66265	0,36076 15730 48139
+/-0,93246 95142 03152	0,17132 44923 79170
p = 7	
0,00000 00000 00000	0,41795 91836 73469
+/-0,40584 51513 77397	0,38183 00505 05119
+/-0,74153 11855 99394	0,27970 53914 89277
+/-0,94910 79123 42759	0,12948 49661 68870
p = 8	
+/-0,18343 46424 95650	0,36268 37833 78362
+/-0,52553 24099 16329	0,31370 66458 77887
+/-0,79666 64774 13627	0,22238 10344 53374
+/-0,96028 98564 97536	0,10122 85362 90376