Fundação CECIERJ

consórcio cederj

# Cálculo Numérico

## Volume único

Dayse Haime Pastore
Roberto C. A. Thomé

Secretaria de Ciência, Tecnologia e Inovação

GOVERNO DO ESTADO RIO DE JANEIRO

UNIVERSIDADE ABERTA DO BRASIL

MINISTÉRIO DA EDUCAÇÃO

GOVERNO FEDERAL BRASIL UNIÃO E RECONSTRUÇÃO

Apoio:

FAPERJ
Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro

## Fundação Cecierj / Consórcio Cederj
www.cecierj.edu.br

**Presidente**
João de Melo Carrilho

**Vice-Presidente de Educação Superior a Distância**
Gerson Oliveira dos Anjos Junior

**Vice-Presidente Científico**
Régia Beatriz Santos de Almeida

**Coordenação do Curso de Engenharia de Produção**
CEFET - Igor Leão dos Santos
UFF - Ana Carolina Maia Angelo

## Material Didático

**Elaboração de Conteúdo**
Dayse Haime Pastore
Roberto C. A. Thomé

**Diretoria de Material Didático**
Ulisses Schnaider

**Diretoria de Design Instrucional**
Diana Castellani

**Diretoria de Material Impresso**
Bianca Giacomelli

**Design Instrucional**
Aroaldo Veneu
Felipe M. Castello-Branco
Luciana Sá Brito

**Preparação de Originais**
Rosane Lira

**Ilustração**
André Amaral

**Capa**
André Amaral

**Diagramação**
Maria Fernanda de Novaes

**Revisão**
Rosane Lira

**Produção Gráfica**
Equipe Cecierj

**Biblioteca**
Simone da Cruz Correa de Souza
Vera Vani Alves de Pinho





P293c
Pastore, Dayse Haime.
Cálculo numérico. Volume único / Dayse Haime Pastore, Roberto C. A. Thomé. – Rio de Janeiro : Fundação Cecierj, 2023.
216p.

ISBN : 978-85-458-0272-3

I. Matemática. II. Cálculo numérico. 1. Thomé, Roberto C. A. I. Título.

CDD: 510

Referências bibliográficas e catalogação na fonte, de acordo com as normas da ABNT.
Texto revisado segundo o novo Acordo Ortográfico da Língua Portuguesa.

## Governo do Estado do Rio de Janeiro

**Governador**
Cláudio Castro

**Secretário de Estado de Ciência, Tecnologia e Inovação**
Mauro Azevedo Neto

## Instituições Consorciadas

**CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca**
Diretor-geral: Maurício Aires Vieira

**FAETEC - Fundação de Apoio à Escola Técnica**
Presidente: Caroline Alves da Costa

**IFF - Instituto Federal de Educação, Ciência e Tecnologia Fluminense**
Reitor: Jefferson Manhães de Azevedo

**IFRJ - Instituto Federal do Rio de Janeiro**
Reitor: Rafael Barreto Almada

**UENF - Universidade Estadual do Norte Fluminense Darcy Ribeiro**
Reitor: Raul Ernesto Lopez Palacio

**UERJ - Universidade do Estado do Rio de Janeiro**
Reitor: Mario Sergio Alves Carneiro

**UFF - Universidade Federal Fluminense**
Reitor: Antonio Claudio Lucas da Nóbrega

**UFRJ - Universidade Federal do Rio de Janeiro**
Vice-reitor em exercício: Carlos Frederico Leão Rocha

**UFRRJ - Universidade Federal Rural do Rio de Janeiro**
Reitor: Roberto de Souza Rodrigues

**UNIRIO - Universidade Federal do Estado do Rio de Janeiro**
Reitor: Ricardo Silva Cardoso

# Sumário

# Aula 1

## Noções básicas sobre erros

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Metas

Apresentar a disciplina *cálculo numérico*; apresentar a aritmética utilizada pelo computador e estimar os possíveis erros cometidos por ele quando realizamos operações com a sua ajuda.

## Objetivos

Esperamos que, ao final dessa aula, você seja capaz de:

1. efetuar mudanças de bases numéricas;
2. estimar os erros gerados pelas operações realizadas em um computador.

## Introdução

Começaremos a primeira aula convidando você a pensar em como transformamos um problema real em um problema matemático. Ou seja, como pegamos um problema real e o transformamos em equações ou perguntas que estejam no universo da matemática.

Vamos olhar para um problema real. Estamos no aeroporto e a atendente da companhia aérea diz: "Começaremos o embarque. Por favor, queiram respeitar a ordem de entrada de acordo com o número do assento indicado em seus bilhetes".

A companhia aérea está tentando alocar os seus passageiros o mais rapidamente possível em seus assentos. Mas qual será a ordem de entrada no avião que torna isso mais eficiente? Será que a melhor ordem é dar preferência para as fileiras que estão na parte de trás da aeronave, ou seria melhor alocar primeiro as pessoas que sentarão junto às janelas?

Esse é um exemplo de um problema real para o qual se busca a melhor solução usando a matemática. Se ficou curioso com esse problema, você pode encontrar mais informações em:

http://www.economist.com/node/21528218.

O esquema a seguir nos apresenta uma forma de como podemos estruturar as fases para resolver problemas desse tipo.

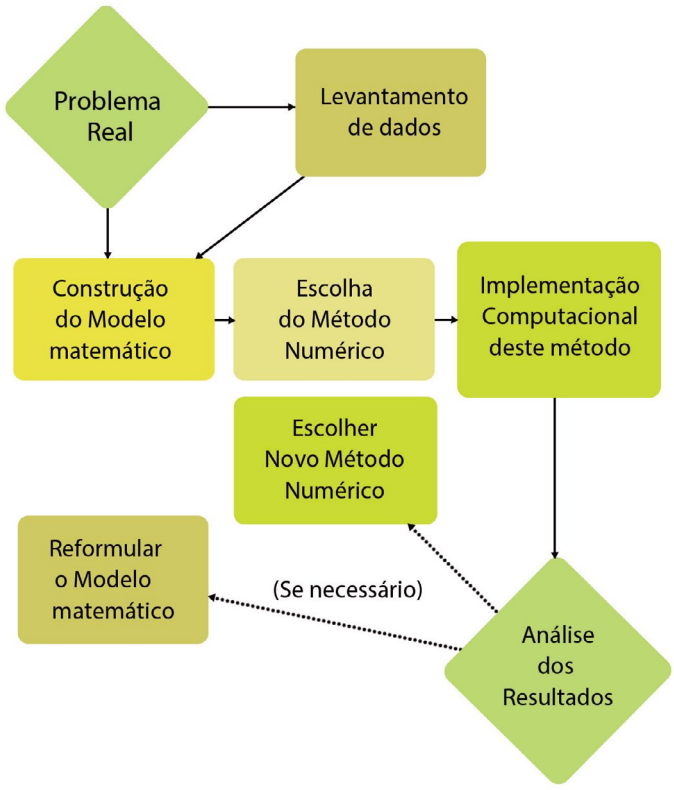



**Figura 1.1:** Diagrama das fases para resolver problemas.

Temos sempre que fazer uma análise rigorosa dos resultados encontrados, pois não é raro que estejam distantes do que se esperaria, ainda que todas as fases do processo tenham sido feitas de forma correta.

Os resultados obtidos dependem de vários fatores, como:

- a precisão dos dados de entrada;
- a forma como esses dados estão sendo representados no computador;
- a maneira como efetuamos as operações numéricas.

Quando obtemos os dados de entrada no computador, eles contêm uma imprecisão inerente, uma vez que esses dados foram coletados através de medidas, usando um equipamento. Podemos citar como exemplo os dados resultantes em uma pesquisa de opinião.

Começaremos estudando os erros que surgem da representação dos números em um computador e os erros resultantes das operações numéricas efetuadas.

## Representação numérica

Vamos começar propondo atividades, pois dessa forma será mais fácil entender como os erros surgem.

### Atividade 1

*Atende ao objetivo 2*

Calcule a área de uma circunferência de raio igual a 100 metros.

### Resposta comentada

Como sabemos, a área de uma circunferência é dada pela fórmula $A = \pi r^2$, onde $r$ é o raio da circunferência. O resultado depende do valor aproximado da constante $\pi$ que será utilizado nas contas. Observe três possíveis resultados abaixo:

(i) A = 31.400 $m^2$

(ii) A = 31.416 $m^2$

(iii) A = 31.415, 92654 $m^2$

Como explicar a diferença entre os três resultados? No resultado (i), consideramos $\pi \approx 3{,}14$; em (ii), $\pi \approx 3{,}1416$; e em (iii), $\pi \approx 3{,}141592654$. Como o número $\pi$ não pode ser representado através de uma quantidade finita de dígitos decimais, as três escolhas são válidas e nos retornam valores diferentes. Isto significa que para cada escolha que fizermos para $\pi$, teremos um erro. Como $\pi$ é um número irracional, a área nunca será obtida exatamente.

## O Maracanã

Esse estádio tem formato oval, medindo 317 m em seu eixo maior e 279 m no menor. Mede 32 metros de altura, o que corresponde a um prédio de seis andares; e a distância entre o espectador mais distante o centro do campo é de 126 m.

Fonte: https://memoria.ebc.com.br/esportes/2013/08/mudancas-no-projeto-do-maracana-levam-comite-olimpico-a-alterar-uso-do-complexo

## *Atividade 2*

***Atende ao objetivo 2***

Calcule os somatórios seguintes em uma calculadora simples e em um computador:

$$\sum_{i=1}^{30.000} x_i \text{ para } x_i = 0{,}5 \text{ e para } x_i = 0{,}11$$

### *Resposta comentada*

Para $x_i = 0{,}5$, na calculadora encontramos $S = 15.000$ e, no computador, $S = 15.000$. Já para $x_i = 0{,}11$, na calculadora encontramos $S = 3.300$ e, no computador, $S = 3.299{,}99691$. A justificativa para a diferença entre os resultados obtidos pelo computador e pela calculadora para $x_i = 0{,}11$ são os erros ocorridos devido à representação dos números em cada máquina.

Vimos nas duas atividades que a representação de um número numa máquina pode produzir erros, que dependem tanto da máquina utilizada, quanto da representação dos valores utilizados.

A representação de um número depende da base escolhida ou disponível na máquina em uso e do número máximo de dígitos usados na sua representação.

Como vimos no exemplo do cálculo da área, qualquer número que não possua representação finita na base escolhida, ou seja, que não possa ser representado através de um número finito de dígitos, apresentará erro no resultado. Note que, quanto maior for o número de dígitos, melhor será a aproximação da área.

A base decimal é a mais utilizada. Outras bases utilizadas são a base 12 e a base 60. Os computadores normalmente operam na base 2, ou seja, no sistema binário.

Convidamos você a pensar um pouco mais sobre a utilização dessas bases. O portal da coleção *M3 Matemática Multimídia* contém recursos educacionais multimídia em formatos digitais, desenvolvidos pela Unicamp para o Ensino Médio de Matemática no Brasil. Entre eles, há um interessante vídeo sobre a base binária, no endereço: http://m3.ime.unicamp.br/recursos/1116.

## Sistemas decimal e binário

Vamos pensar no que ocorre quando um usuário vai usar um computador para fazer contas.

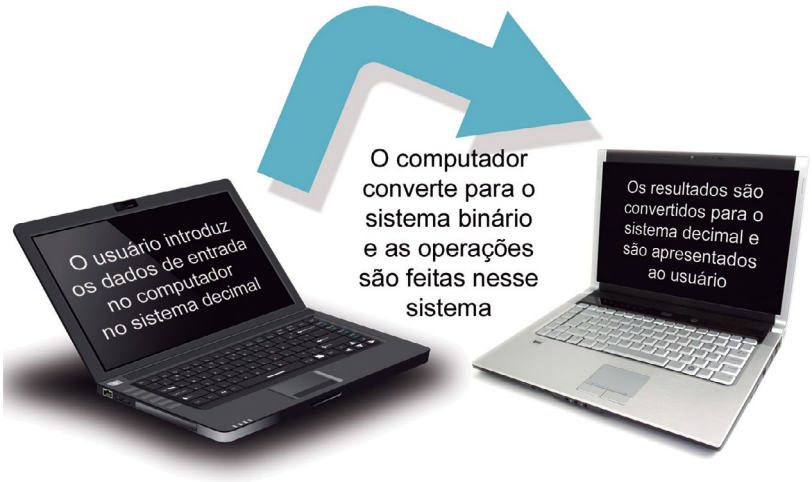



**Figura 1.2:** Modo como os computadores interpretam os dados numéricos.

Para termos uma ideia do que está acontecendo, vamos aprender a trabalhar os números no sistema binário.

No sistema decimal, utilizamos dez algarismos para representar os números: 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9. Já no sistema binário, utilizamos dois: 0 e 1.

O sistema decimal utiliza a base 10, ou seja, os números são escritos utilizando os dez algarismos citados acima e potências do número 10. Já no sistema binário, utilizamos somente dois algarismos e potências do número 2.

De uma forma geral, podemos representar um número em uma base $\beta$ como $(\alpha_n\alpha_{n-1}\ldots\alpha_2\alpha_1\alpha_0)_\beta$, onde $0 \leq \alpha_k \leq (\beta - 1)$, para $k = 0, 1, \ldots, n$, ou ainda da forma:

$$\alpha_n\beta^n + \alpha_{n-1}\beta^{n-}{}_1 + \ldots + \alpha_1\beta_1 + \alpha_0\beta^0.$$

Para que você possa compreender melhor a notação utilizada acima, vejamos como alguns números são representados na base binária e na base decimal:

$(100)_2 = 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0$

$(100)_{10} = 1 \times 10^2 + 0 \times 10^1 + 0 \times 10^0$

$(1101)_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$

$(13)_{10} = 1 \times 10^1 + 3 \times 10^0$

Agora que já estamos familiarizados com a notação, vamos efetuar algumas contas simples.

## Atividade 3

***Atende ao objetivo 1***

Mostre que $(100)_2 \neq (100)_{10}$:

### *Resposta comentada*

Vimos nos exemplos de representações numéricas anteriores que:

$(100)_2 = 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0$.

Dessa forma, podemos escrever que:

$(100)_2 = 4 + 0 + 0 = 4 = (4)_{10.}$

Por outro lado, temos que:

$(100)_{10} = 1 \times 10^2 + 0 \times 10^1 + 0 \times 10^0 = 100 + 0 + 0 = 100.$

Como $4 \neq 100$, então concluímos que $(100)_2 \neq (100)_{10}$.

## Atividade 4

*Atende ao objetivo 1*

Mostre que $(1101)_2 = (13)_{10}$:

### Resposta comentada

Vimos nos exemplos de representações numéricas anteriores que:

$(1101)_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$.

Dessa forma, podemos escrever que:

$(1101)_2 = 8 + 4 + 0 + 1 = 13$.

Por outro lado, temos que:

$(13)_{10} = 1 \times 10^1 + 3 \times 10^0 = 10 + 3 = 13$.

Como $(13)_{10} = 13$, então concluímos que $(1101)_2 = (13)_{10}$.

Converter um número do sistema binário para o sistema decimal agora deve ser fácil, pois basta efetuarmos as contas.

Para convertermos um número do sistema binário para o decimal de maneira prática, basta seguir os seguintes passos da figura abaixo. Observe como convertemos facilmente o número $(1101)_2$ da base binária para a base decimal, obtendo o resultado:

$$13 = (13)_{10}$$

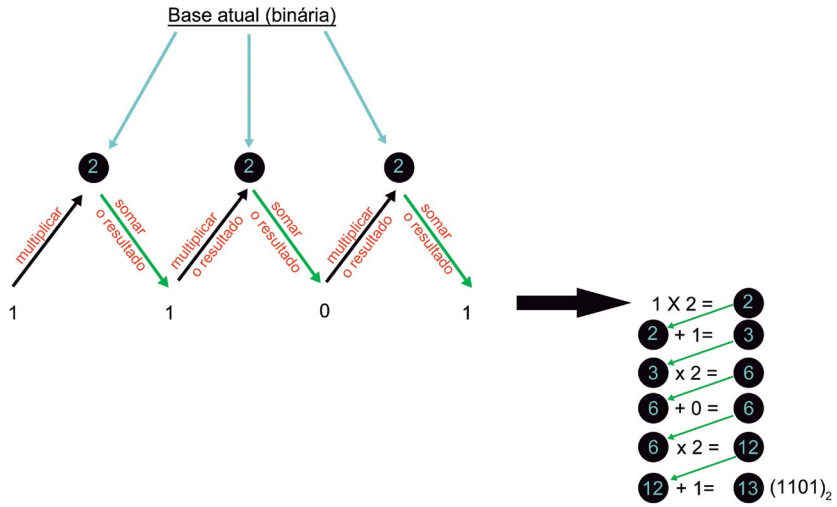



**Figura 1.3:** Esquema prático para conversão de um número do sistema binário para o decimal.

Agora que você já está familiarizado com a notação, iremos formalizar matematicamente o passo a passo. A forma de efetuarmos as contas nos exemplos anteriores nos dá um algoritmo para passarmos do sistema binário para o sistema decimal.

Considere o número $(a_n a_{n-1} \ldots a_2 a_1 a_0)_2$ no sistema binário. Vamos montar um processo recursivo onde, no último passo, teremos o número escrito no sistema decimal. O processo é dado da seguinte maneira:

$$b_n = a_n$$

$$b_{n-1} = a_{n-1} + 2 \times b_n$$

$$b_{n-2} = a_{n-2} + 2 \times b_{n-1}$$

$$\vdots$$

$$b_1 = a_1 + 2 \times b_2$$

$$b_0 = a_0 + 2 \times b_1$$

Ao final desse processo, teremos: $(a_n a_{n-1} \ldots a_2 a_1 a_0)_2 = (b_0)_{10}$.

## Atividade 5

*Atende ao objetivo 1*

Mostre que $(10111)_2 = (23)_{10}$ utilizando o processo recursivo acima.

### Resposta comentada

Se escrevermos $(10111)_2 = (a_4a_3a_2a_1a_0)_2$, então teremos as relações $a_4 = 1$, $a_3 = 0$, $a_2 = 1$, $a_1 = 1$ e $a_0 = 1$. Dessa maneira, temos:

$b_4 = a_4 = 1$

$b_3 = a_3 + 2 \times b_4 = 0 + 2 \times 1 = 2$

$b_2 = a_2 + 2 \times b_3 = 1 + 2 \times 2 = 5$

$b_1 = a_1 + 2 \times b_2 = 1 + 2 \times 5 = 11$

$b_0 = a_0 + 2 \times b_1 = 1 + 2 \times 11 = 23$

Logo, $(10111)_2 = (23)_{10} = 23$.

Agora vamos fazer o contrário, ou seja, vamos transformar um número inteiro do sistema decimal para um número inteiro do sistema binário.

Você deve ter notado que, para converter um número do sistema binário para o sistema decimal, realizamos sucessivas multiplicações pela base 2. É natural concluir que, para fazer a conversão oposta, teremos que desfazer esse processo realizando sucessivas divisões pela base 2. Sempre que dividimos um número na base decimal por 2, obtemos resto 0 ou 1, formando assim os dígitos do número binário.

Na atividade anterior, convertemos o número $(10111)_2$ da base binária para a base decimal, obtendo $(23)_{10}$ como resposta. Agora vamos fazer o oposto!

## Atividade 6

*Atende ao objetivo 1*

Converta o número 23 da base decimal para a base binária.

### Resposta comentada

Da atividade anterior, já sabemos que a resposta é $(10111)_2$. Mas como chegar nesse resultado? Precisamos desfazer as sucessivas multiplicações por 2. Se dividirmos 23 por 2, iremos obter resto 1 e quociente 11. Para continuar esse processo numérico, basta sempre pegarmos o quociente resultante e dividirmos por 2 até obter o quociente 1, quando não será mais possível continuarmos esse processo. Vejamos:

$23 = 2 \times (11) + 1$

$11 = 2 \times (5) + 1$

$5 = 2 \times (2) + 1$

$2 = 2 \times (1) + 0$

$1 = 2 \times (0) + 1$.

Repare que a resposta $(10111)_2$ aparece como os restos de baixo para cima dessas operações. Logo $23 = (23)_{10} = (10111)_2$.

Já para convertermos um número do sistema decimal para o binário de maneira prática, basta seguir os seguintes passos da figura abaixo. Observe como convertemos facilmente o número $(23)_{10} = 23$ da base decimal para a base binária, obtendo o resultado $(10111)_2$

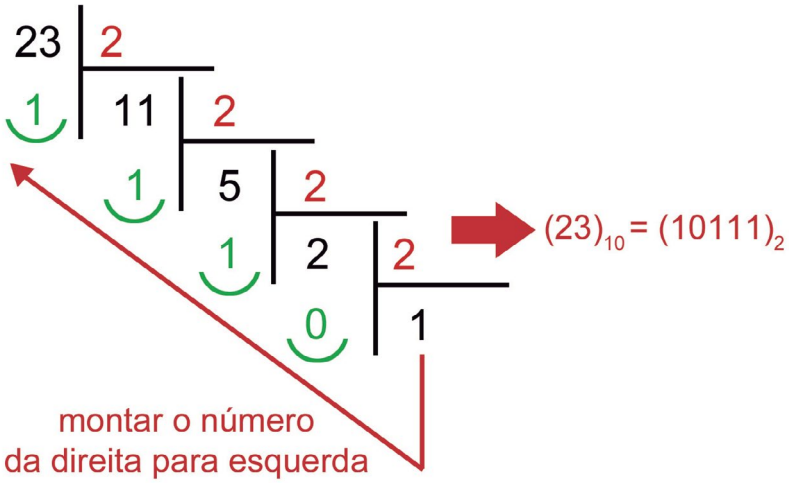



**Figura 1.4:** Esquema prático para conversão de um número do sistema decimal para o binário.

Acredito que você percebeu o que foi feito na última atividade: dividimos por 2 e tomamos o resto.

Para generalizarmos o processo, podemos escrever o algoritmo para o caso geral:

- Algoritmo: convertendo um número da base decimal para binária.
  - Dados de entrada: $k = 0$, $D_k$ = número na base decimal.
  - Passo 1:

    Efetue a divisão de $D_k$ por 2 com resto

    $D_k = 2 \times (q_k) + r_k$

    Faça $a_k = r_k$
  - Passo 2:

    Se $q_k = 0$, então pare e escreva: $(D_0)_{10} = (a_k a_{k-1} \ldots a_2 a_1 a_0)_2$

    Senão, faça $k \leftarrow k+1$

    Escreva $D_k = q_{k-1}$ e volte para ao passo 1.

## Atividade 7

***Atende ao objetivo 1***

Converta o número 23 da base decimal para a base binária usando a notação e o passo a passo do algoritmo anterior.

### Resposta comentada

1ª iteração: $k = 0$ e $D_0 = 23$

(Passo 1) $D_0 = 23 = 2 \times (q_0) + r_0 = 2 \times (11) + 1 \rightarrow a_0 = r_0 \rightarrow a_0 = 1$

(Passo 2) Como $q_0 = 11 \neq 0$, então $k \leftarrow 0 + 1 \rightarrow k = 1$

$D_1 = q_0 \rightarrow D_1 = 11$.

2ª iteração: $k = 1$ e $D_1 = 11$

(Passo 1) $D_1 = 11 = 2 \times (q_1) + r_1 = 2 \times (5) + 1 \rightarrow a_1 = r_1 \rightarrow a_1 = 1$

(Passo 2) Como $q_1 = 5 \neq 0$, então $k \leftarrow 1 + 1 \rightarrow k = 2$

$D_2 = q_1 \rightarrow D_2 = 5$.

3ª iteração: $k = 2$ e $D_2 = 5$

(Passo 1) $D_2 = 5 = 2 \times (q_2) + r_2 = 2 \times (2) + 1 \rightarrow a_2 = r_2 \rightarrow a_2 = 1$

(Passo 2) Como $q_2 = 2 \neq 0$, então $k \leftarrow 2 + 1 \rightarrow k = 3$

$D_3 = q_2 \rightarrow D_3 = 2$.

4ª iteração: $k = 3$ e $D_3 = 2$

(Passo 1) $D_3 = 2 = 2 \times (q_3) + r_3 = 2 \times (1) + 0 \rightarrow a_3 = r_3 \rightarrow a_3 = 0$

(Passo 2) Como $q_3 = 1 \neq 0$, então $k \leftarrow 3 + 1 \rightarrow k = 4$

$D_4 = q_3 \rightarrow D_4 = 1$.

5ª iteração: $k = 4$ e $D_4 = 1$

(Passo 1) $D_4 = 1 = 2 \times (q_4) + r_4 = 2 \times (0) + 1 \rightarrow a_4 = r_4 \rightarrow a_4 = 1$

(Passo 2) Como $q_4 = 0$, então:

$(D_0)_{10} = (a_4 a_3 a_2 a_1 a_0)_2 \rightarrow (23)_{10} = (10111)_2$.

De maneira resumida, temos:

$D_0 = 23 = 2 \times (11) + 1 \rightarrow a_0 = 1$

$D_1 = 11 = 2 \times (5) + 1 \rightarrow a_1 = 1$

$D_2 = 5 = 2 \times (2) + 1 \rightarrow a_2 = 1$

$D_3 = 2 = 2 \times (1) + 0 \rightarrow a_3 = 1$

$D_4 = 1 = 2 \times (0) + 1 \rightarrow a_4 = 1$

Repare que a resposta $(10111)_2$ aparece como os restos de baixo para cima dessas operações. Logo $23 = (23)_{10} = (10111)_2$.

Vamos praticar um pouco mais!

## Atividade 8

***Atende ao objetivo 1***

Converta o número 123 da base decimal para a base binária usando a notação e o passo a passo do algoritmo anterior.

### *Resposta comentada*

De maneira resumida, temos:

$D_0 = 123 = 2 \times (q_0) + r_0 = 2 \times (61) + 1 \rightarrow a_0 = r_0 \rightarrow a_0 = 1$

$D_1 = 61 = 2 \times (q_1) + r_1 = 2 \times (30) + 1 \rightarrow a_1 = r_1 \rightarrow a_1 = 1$

$D_2 = 30 = 2 \times (q_2) + r_2 = 2 \times (15) + 0 \rightarrow a_2 = r_2 \rightarrow a_2 = 0$

$D_3 = 15 = 2 \times (q_3) + r_3 = 2 \times (7) + 1 \rightarrow a_3 = r_3 \rightarrow a_3 = 1$

$D_4 = 7 = 2 \times (q_4) + r_4 = 2 \times (3) + 1 \rightarrow a_4 = r_4 \rightarrow a_4 = 1$

$D_5 = 3 = 2 \times (q_5) + r_5 = 2 \times (1) + 1 \rightarrow a_5 = r_5 \rightarrow a_5 = 1$

$D_6 = 1 = 2 \times (q_6) + r_6 = 2 \times (0) + 1 \rightarrow a_6 = r_6 \rightarrow a_6 = 1$

Como $q_6 = 0$, então escrevemos:

$(D_0)_{10} = (a_6 a_5 a_4 a_3 a_2 a_1 a_0)_2 \rightarrow (123)_{10} = (1111011)_2$.

Tratamos aqui, apenas das conversões dos sistemas decimais e binários para números inteiros. Essas mudanças podem ser feitas também para outras bases e também para números fracionários. Você pode se aprofundar mais sobre o assunto lendo as obras apresentadas no boxe a seguir.

Para aprender mais sobre a conversão para outras bases, leia o capítulo 1 de: RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R. *Cálculo numérico:* aspectos teóricos e computacionais. São Paulo: Makron Books, 1996.

## Aritmética de ponto flutuante

Como um computador (ou uma calculadora) representa os números reais? Ele representa todos os números reais? Como ele representa o zero ou o infinito? Qual será o maior e o menor número para um computador?

As respostas para essas perguntas estão relacionadas com o modo como o computador efetua as operações. Antes de respondê-las, vamos aprender alguns conceitos.

A representação de números reais mais utilizada em máquinas é a do ponto flutuante. Esse número tem três partes: o sinal, a parte fracionária (*mantissa*) e o expoente. A principal vantagem da representação em ponto flutuante é que ela pode representar uma grande faixa de números, se comparada à representação de ponto fixo.

### *Ponto flutuante* ou vírgula flutuante?

O nome correto deste formato de representação na língua portuguesa deveria ser *vírgula flutuante*, pois em nossa língua é a vírgula que separa a parte inteira de um número de sua parte decimal. No entanto, como o conceito foi desenvolvido em língua inglesa, que usa o ponto,

e não a vírgula, para separar inteiros de decimais, a expressão *ponto flutuante* foi consagrada na literatura. Por isso, optamos por adotá-la.

---

O computador tem uma memória finita e só consegue representar um número finito de números. Por exemplo, o número 1/3 = 0,33333... terá que ter uma representação finita para o computador. Assim sendo, esse número terá que ser aproximado.

Primeiramente introduziremos o conceito de *mantissa*, pois ela é parte fundamental da representação dos números.

- Mantissa é a parte do ponto flutuante que contém os dígitos significativos. Dependendo da interpretação, pode ser considerada um número inteiro ou uma fração.

Agora sim, vamos ao conceito de *aritmética de ponto flutuante*.

- Na aritmética de ponto flutuante um número é representado na forma:

$$\pm (0, d_1 d_2 \ldots d_t) \times \beta^e$$

onde:

$\beta$ é a base em que a máquina opera;

$t$ é o número de dígitos na mantissa;

$e$ é o expoente que pertence a um intervalo.

Para compreendermos melhor a ideia do ponto flutuante, vamos ver como uma máquina trabalha com os números.

Considere uma máquina que opera no sistema com a base $\beta = 10$, o número de dígitos na mantissa $t = 3$ e o expoente $e \in [-3,3]$. Nesse caso, os números serão representados na seguinte forma, nesse sistema:

$$0, d_1 d_2 d_3 \times 10^{e},$$

onde para cada $j = 1, 2, 3$ temos $d_j \in \{0, 1, 2, \ldots, 9\}$, com $d_1 \neq 0$.

Com isso, temos:

| | |
|---|---|
| Menor número em valor absoluto | $m = 0{,}100 \times 10^{-3}$ $(= 10^{-4})$ |
| Maior número em valor absoluto | $M = 0{,}999 \times 10^{3}$ $(= 999)$ |
| Zero | $0{,}000 \times 10^{-3}$ $(= 0)$ |

Observação: repare que para o número *zero* do quadro acima escolhemos o expoente $e = -3$. Justificaremos essa escolha mais adiante, quando falarmos das operações.

Quais números conseguiremos representar nessa máquina? Para entendermos melhor e respondermos esta pergunta, consideramos o conjunto:

$$NM = \{ x \in \Re \mid m \leq |x| \leq M \}$$

Se $x \notin NM$, então não teremos a chance de representá-lo nessa máquina. Vejamos as duas situações abaixo.

- O número $x = 0{,}245 \times 10^{-4}$ não pode ser representado nesta máquina, pois o expoente $e = -4$ é menor do que $-3$ (menor expoente dessa máquina). Nesse caso, a máquina irá acusar a ocorrência de *underflow*;
- O número $x = 0{,}945 \times 10^{7}$ também não pode ser representado nesta máquina, pois o expoente $e = 7$ é maior do que 3 (maior expoente dessa máquina). Nesse caso, a máquina irá acusar a ocorrência de *overflow*.

Se $x \in NM$, então a máquina poderá representá-lo. Vejamos a seguinte situação.

- Considere o número $x = 43{,}275 = 0{,}432758 \times 10^{2}$. Observe que esse número tem 6 dígitos na sua mantissa, mas essa máquina só possui 3 dígitos. E agora, como representamos esse número? Teremos $0{,}432 \times 10^{2}$ ou $0{,}433 \times 10^{2}$? Se for usado arredondamento, o número será representado por $0{,}433 \times 10^{2}$. Por outro lado, se for usado truncamento, o número será representado por $0{,}432 \times 10^{2}$.

Observação: aprenderemos melhor a respeito de arredondamento e truncamento um pouco mais adiante.

Agora já estamos prontos para responder as perguntas feitas anteriormente. O computador representa os números usando a aritmética de ponto flutuante, pois essa foi a aritmética criada para as máquinas. O computador não representa todos os números reais, devido a sua

limitação de memória. Como acabamos ver, cada máquina terá o seu maior e o seu menor número, dependendo da quantidade de *bits* e de memória.

Algumas linguagens de programação consideram um número menor de casas decimais do que a máxima permitida pela aritmética de ponto flutuante da máquina. Nesses casos, essas linguagens permitem utilizarmos a *precisão dupla*.

Agora considere uma máquina que opera com base $\beta$, número $t$ na mantissa e expoente $e \in [-n, n]$, onde $n$ é um número natural. Assim, um número $y$ em ponto flutuante nessa máquina será denotado por $fl(y)$ e apresentado da forma:

$$fl(y) = \pm 0, d_1 d_2 \ldots d_t \times \beta^e.$$

Agora podemos definir o metódo de truncamento e de arredondamento.

- No *método de truncamento,* simplesmente cortamos os dígitos na quantidade $k$ de casas decimais que queremos ou que são permitidas. Ou seja, consiste em cortar os dígitos $d_{k+1}$, $d_{k+2}$,..., e dessa forma ficamos com:

$$fl(y) = \pm 0, d_1 d_2 \ldots d_k \times \beta^e.$$

- No *método de arredondamento,* adicionamos $5 \times \beta^{e-(k+1)}$ ao número $y$ e usamos o truncamento no resultado. Com isso, obtemos um número na forma:

$$fl(y) = \pm 0, \delta_1 \delta_2 \ldots \delta_k \times \beta^e.$$

De fato, quando fazemos o arredondamento, se $d_{k+1} \geq 5$, adicionamos 1 a $d_k$ para obter $fl(y)$ chamado de *arredondamento para cima*. Quando $d_{k+1} < 5$, simplesmente truncamos, ou seja, ficamos apenas com os $k$ dígitos, chamado de *arredondamento para baixo*. Se arredondarmos para baixo, então $\delta_i = d_i$ para cada $i = 1, 2, \ldots, k$. No entanto, se arredondarmos para cima, os algarismos (e até mesmo o expoente) podem mudar.

Para entender melhor essas duas definições, vejamos algumas situações. Considere uma máquina que opera no sistema com a base $\beta = 10$, o número de dígitos na mantissa $t = 3$ e o expoente $e \in [-5, 5]$.

Dessa forma, para cada $j = 1, 2, 3, 4, 5$, temos $d_j \in \{0, 1, 2, \ldots, 9\}$, com $d_1 \neq 0$. Acompanhe na tabela a seguir.

**Tabela 1.1:** Tabela comparativa entre as diferentes representações de um número

| Número *y* | Representação *fl(y)* por truncamento | Representação *fl(y)* por arredondamento |
|---|---|---|
| 2,15 | $0{,}215 \times 10^1$ | $0{,}215 \times 10^1$ |
| 32,114 | $0{,}321 \times 10^2$ | $0{,}321 \times 10^2$ |
| –2,634 | $-0{,}263 \times 10^1$ | $-0{,}263 \times 10^1$ |
| 354,7 | $0{,}354 \times 10^3$ | $0{,}355 \times 10^3$ |
| –56,78 | $-0{,}567 \times 10^2$ | $-0{,}568 \times 10^2$ |
| $\pi = 3{,}141$ | $0{,}314 \times 10^1$ | $0{,}314 \times 10^1$ |
| 0,0000003 | underflow | underflow |
| 1.029.543,54 | overflow | overflow |

## Erro absoluto e erro relativo

Vamos agora aprender sobre *erro absoluto* e *erro relativo*. O erro absoluto é simplesmente a diferença entre o valor e sua aproximação. Já o erro relativo é aquele que, além de calcularmos a diferença, ainda normalizamos pela aproximação, pois podemos ter ordens de grandeza diferentes envolvidas.

Quando estamos próximos a uma solução é difícil saber o quão pequeno é o erro. O erro relativo serve para termos uma ideia da porcentagem do erro que estamos cometendo. Por exemplo, 1% de erro relativo significa que o erro relativo é igual a 0,01.

Vejamos duas definições a seguir:

- Erro absoluto: considere $\overline{x}$ uma aproximação do valor $x$. Então o erro absoluto é dado por:

$$ErA_x = |\overline{x} - x|.$$

- Erro Relativo: considere $\overline{x}$ uma aproximação do valor $x$. Então o erro relativo é dado por $\dfrac{|\overline{x} - x|}{|\overline{x}|}$. Como, na prática, o que temos é o valor aproximado, então se considera o erro relativo como sendo:

$$ErR_x = \frac{\left|\bar{x} - x\right|}{\left|\bar{x}\right|}.$$

Para entender melhor essas duas definições, vejamos algumas situações:

**Tabela 1.2:** Tabela comparativa entre erro absoluto e erro relativo

| Valor $x$ | Aproximação $\bar{x}$ | Erro Absoluto $\lvert\bar{x}-x\rvert$ | Erro Relativo $\frac{\lvert\bar{x}-x\rvert}{\lvert\bar{x}\rvert}$ |
|---|---|---|---|
| $0{,}3000 \times 10^{1}$ | $0{,}3100 \times 10^{1}$ | 0,1 | $0{,}3225 \times 10^{-1}$ |
| $0{,}3000 \times 10^{-3}$ | $0{,}3100 \times 10^{-3}$ | $0{,}10 \times 10^{-4}$ | $0{,}3225 \times 10^{-1}$ |
| $0{,}3000 \times 10^{4}$ | $0{,}3100 \times 10^{4}$ | $0{,}10 \times 10^{3}$ | $0{,}3225 \times 10^{-1}$ |

Nesse exemplo, podemos ver como o erro absoluto pode variar dependendo da ordem de grandeza do valor $x$, mas o erro relativo se mantém.

## Operações e erros na aritmética de ponto flutuante

Como os erros se propagam quando operamos na aritmética de ponto flutuante? Precisamos analisar e entender o que acontece quando realizamos uma operação no computador. Quando somamos, subtraímos, multiplicamos ou dividimos, cada uma dessas operações podem gerar erro. Após uma sequência de operações, o erro final é composto pelo erro de cada operação e pelo erro no resultado da sequência de operações.

Para compreendermos isso melhor, vejamos alguns exemplos de operações e os erros que aparecem. Vamos trabalhar na base *10*, com três dígitos, $x = 4/3$, $y = 5/7$, $fl(x) = 0{,}133 \times 10^{1}$ e $fl(y) = 0{,}714 \times 10^{0}$.

Usando o método do truncamento, obtemos os resultados da tabela abaixo:

**Tabela 1.3:** Tabela comparativa entre erro absoluto e erro relativo em operações numéricas

| Operação | Resultado | Erro Absoluto $\lvert\bar{x}-x\rvert$ | Erro Relativo $\frac{\lvert\bar{x}-x\rvert}{\lvert\bar{x}\rvert}$ |
|---|---|---|---|
| fl(x) + fl(y) | $0{,}108 \times 10^{1}$ | $0{,}333 \times 10^{-3}$ | $0{,}307 \times 10^{-3}$ |
| fl(x) – fl(y) | $0{,}417 \times 10^{0}$ | $0{,}333 \times 10^{-3}$ | $0{,}799 \times 10^{-3}$ |
| fl(x) × fl(y) | $0{,}249 \times 10^{0}$ | $0{,}627 \times 10^{-1}$ | $0{,}251 \times 10^{0}$ |
| fl(x) ÷ fl(y) | $0{,}252 \times 10^{1}$ | $0{,}225 \times 10^{-2}$ | $0{,}100 \times 10^{-2}$ |

Se usarmos o método do arredondamento, vamos obter os seguintes resultados:

**Tabela 1.4:** tabela comparativa entre erro absoluto e erro relativo em operações numéricas.

| Operação | Resultado | Erro Absoluto $\lvert\bar{x}-x\rvert$ | Erro Relativo $\frac{\lvert\bar{x}-x\rvert}{\lvert\bar{x}\rvert}$ |
|---|---|---|---|
| fl(x) +fl(y) | $0{,}108 \times 10^{1}$ | $0{,}333 \times 10^{-3}$ | $0{,}308 \times 10^{-3}$ |
| fl(x) – fl(y) | $0{,}417 \times 10^{0}$ | $0{,}333 \times 10^{-3}$ | $0{,}799 \times 10^{-3}$ |
| fl(x) × fl(y) | $0{,}25 \times 10^{0}$ | $0{,}625 \times 10^{-1}$ | $0{,}25 \times 10^{0}$ |
| fl(x) ÷ fl(y) | $0{,}252 \times 10^{1}$ | $0{,}225 \times 10^{-2}$ | $0{,}100 \times 10^{-2}$ |

Vamos agora calcular as fórmulas para os erros absoluto e relativo nas operações aritméticas com erros nas parcelas ou fatores.

Sejam $x$ e $y$, tais que $x = fl(x) + ErA_x$ e $y = fl(y) + ErA_y$.

Logo, teremos:

**Adição:**

$$\begin{aligned} x + y &= (fl(x) + ErA_x) + (fl(y) + ErA_y) \\ &= (fl(x) + fl(y)) + (\,\mathrm{ErA}_x + ErA_y) \\ &= (fl(x) + fl(y)) + (ErA_{x+y}). \end{aligned}$$

Assim, o erro absoluto da soma, denotado por $ErA_{x+y}$ é a soma dos erros absolutos das parcelas:

$$ErA_{x+y} = ErA_x + ErA_y.$$

Vamos calcular, agora, o erro relativo nesta operação:

$$ErR_{x+y} = \frac{ErA_{x+y}}{fl(x)+fl(y)}$$
$$= \frac{ErA_x}{fl(x)}\left(\frac{fl(x)}{fl(x)+fl(y)}\right)+\frac{ErA_y}{fl(y)}\left(\frac{fl(y)}{fl(x)+fl(y)}\right)$$
$$= ErR_x\left(\frac{fl(x)}{fl(x)+fl(y)}\right)+ErR\left(\frac{fl(y)}{fl(x)+fl(y)}\right)$$

Analogamente, temos que os erros, absoluto e relativo, na diferença são dados por:

$$ErA_{x-y} = ErA_x - ErA_y$$

e

$$ErR_{x-y} = ErR_x\left(\frac{fl(x)}{fl(x)-fl(y)}\right)+ErR_y\left(\frac{fl(y)}{fl(x)-fl(y)}\right).$$

Agora vamos calcular o erro na multiplicação:

$$x \times y = (fl(x) + ErA_x) \times (fl(y) + ErA_x).$$

Fazendo a distributividade no produto, teremos:

$$x \times y = fl(x)\times fl(y) + fl(x) \times ErA_y + fl(y)\times ErA_x + ErA_x \times ErA_y.$$

Se considerarmos que $ErA_x \times ErA_y$ é relativamente pequeno e desprezá-lo, o erro absoluto do produto é dado por:

$$ErA_{x\times y} \approx fl(x) \times ErA_y + fl(y) \times ErA_x.$$

Já o erro relativo do produto será dado por:

$$ErR_{x\times y} \approx \left(\frac{fl(x)ErAy + fl(y)ErAx}{fl(x)fl(y)}\right) = \left(\frac{fl(x)ErAy}{fl(x)fl(y)}\right)+\left(\frac{fl(y)ErAx}{fl(x)fl(y)}\right)$$
$$= \frac{ErAy}{fl(y)}+\frac{ErAx}{fl(x)}.$$

Com isso, temos:

$$ErR_{x\times y} \approx ErR_x + ErR_y.$$

De maneira análoga, os erros, absoluto e relativo, na divisão são dados por:

$$ErA_{x/y} \approx \left( \frac{fl(y)ErAx - fl(x)ErAy}{fl(y)^2} \right)$$

e

$$ErR_{x/y} \approx ErR_x - ErR_y.$$

O cálculo dos erros relativos e absolutos para a divisão pode ser encontrado nas obras listadas na seção de Referências.

Vimos que podem acontecer erros nas operações numéricas. Podemos ter erros absolutos grandes com erros relativos pequenos e vice-versa. Também podemos ter erros relativos e absolutos grandes. Note que, quando temos erros grandes, os nossos resultados não são confiáveis. No cálculo numérico, estamos sempre buscando maneiras de cometermos erros menores, para chegar a resultados melhores. Na meteorologia, por exemplo, erros grandes podem esconder uma possível catástrofe natural ou podem gerar pânico sem necessidade.

## Conclusão

Nesse ponto você já deve entender como as operações são realizadas pelo computador, e deve também imaginar quantos erros são gerados. Mas com o conhecimento adquirido, você também já está ciente de como o computador é fundamental para resolvermos problemas reais com a ajuda da matemática.

## Resumo

Nessa aula, você estudou os pontos listados a seguir.

- A representação de um número depende da base escolhida ou disponível na máquina em uso e do número máximo de dígitos usados na sua representação. A base decimal é a mais utilizada. Outras bases utilizadas são a base 12 e a base 60. Os computadores normalmente operam na base dois, ou seja, no sistema binário.
- No sistema decimal, utilizamos dez algarismos para representar os números: 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9. Já no sistema binário, utilizamos dois: 0 e 1. De uma forma geral, podemos representar um número

em uma base $\beta$ como $(\alpha_n\alpha_{n-1} \ldots \alpha_2\alpha_1\alpha_0)_\beta$, onde $0 \leq a_k \leq (\beta - 1)$, para $k = 0, 1\ldots, n$, ou ainda da forma: $\alpha_n\beta^n + \alpha_{n-1}\beta^{n-1} + \ldots + \beta^1 + \alpha_0\beta^0$.

- Para converter um número do sistema binário para o sistema decimal realizamos sucessivas multiplicações pela base 2. É natural concluir que, para fazer a conversão oposta, teremos que desfazer esse processo realizando sucessivas divisões pela base 2. Sempre que dividimos um número na base decimal por 2, obtemos resto 0 ou 1, formando assim os dígitos do número binário.
- A representação de números reais mais utilizada em máquinas é a do ponto flutuante. Esse número tem três partes: o sinal, a parte fracionária (*mantissa*) e o expoente.
- Na aritmética de ponto flutuante, um número é representado na forma $\pm (0, d_1 d_2 \ldots d_t) \times \beta^e$, onde $\beta$ é a base em que a máquina opera, $t$ é o número de dígitos na mantissa e $e$ é o expoente que pertence a um intervalo.
- No método de truncamento, simplesmente cortamos os dígitos na quantidade $k$ de casas decimais que queremos ou que são permitidas. Ou seja, o método consiste em cortar os dígitos $d_{k+1}, d_{k+2}, \ldots$, e dessa forma ficamos com $fl(y) = \pm 0, d_1, d_2, \ldots, d_k, \times \beta^e$.
- No método de arredondamento, adicionamos $5 \times \beta^{e-(k+1)}$ ao número $y$ e usamos o truncamento no resultado. Com isso, obtemos um número na forma $fl(y) = \pm 0, \delta_1, \delta_2, \ldots, \delta_k \times \beta^e$.
- O erro absoluto é dado por $ErA_x = |\overline{x}\text{-}x|$, onde $\overline{x}$ é uma aproximação do valor $x$.
- O erro relativo é dado $ErR_x = \dfrac{|\overline{x} - x|}{|\overline{x}|}$, onde $\overline{x}$ é uma aproximação do valor $x$.

## Referências

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R. *Cálculo numérico:* aspectos teóricos e computacionais. São Paulo: Makron Books, 1996.

BURDEN, Richard L.; FAIRES, Douglas. *Análise numérica*. São Paulo: Pioneira Thomson Learning, 2003.

# Aula 2

## Zeros reais de funções reais: método da bissecção e método da posição falsa

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Introduzir o conceito de zeros de função através de métodos iterativos como o *método da bissecção* e *método da posição falsa*.

## Objetivos

Esperamos que, ao final desta aula, você seja capaz de:

1. aplicar o *método da bissecção*;
2. aplicar o *método da posição falsa*.

## Introdução

Começaremos explicando o que seria o zero de uma função e o quanto é importante e útil encontrá-lo.

Definimos o zero de uma função $y = f(x)$ como sendo o valor real $x_0$, tal que $f(x_0) = 0$.

Dessa forma, para se determinar todos os zeros de uma função $y = f(x)$, basta igualar o valor de $y$ ou $f(x)$ a zero e resolver a equação encontrada. Graficamente, os zeros da função $f$ são os valores por onde a função intercepta o eixo $x$.

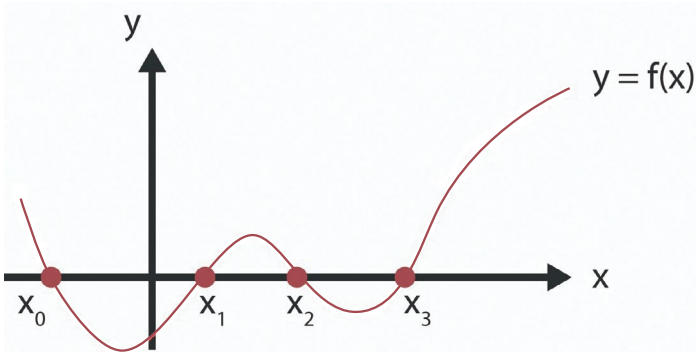



**Figura 2.1:** Gráfico de uma função $y = f(x)$ e seus zeros.

Os *zeros* de uma função polinomial $p_n(x) = a_n x^n + ... + a_2 x^2 + a_1 x + a_0$ são chamados de *raízes* do polinômio.

Algumas funções possuem zeros fáceis de se encontrar. Quando a função é uma reta do tipo $y = ax + b$, por exemplo, resolvemos a equação $ax + b = 0$. Dessa forma, com fáceis manipulações algébricas, encontramos o zero da reta dado por $x = -b/a$.

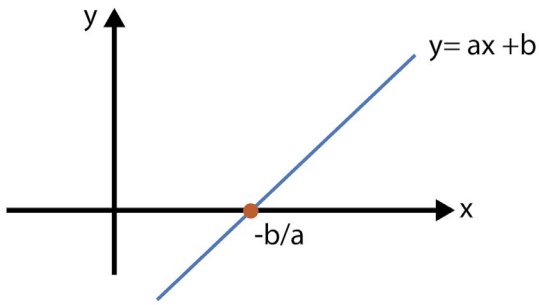



**Figura 2.2:** Gráfico da reta $y = ax + b$ e o seu zero.

Quando temos uma parábola do tipo $y = ax^2 + bx + c$, os seus dois zeros são dados pela famosa fórmula de Bhaskara:

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

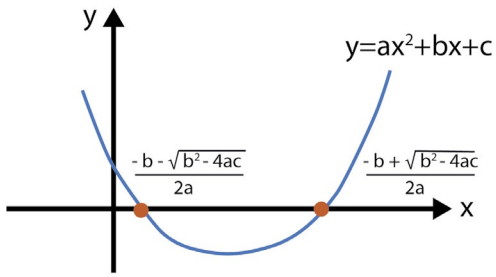



**Figura 2.3:** Gráfico da parábola $y = ax^2 + bx + c$ e seus dois zeros.

## A origem do nome *Fórmula de Bhaskara*

O nome da fórmula foi dado em homenagem a Bhaskara Akaria. Ele foi um matemático, professor, astrólogo e astrônomo indiano; considerado o mais importante matemático do século XII e o último matemático medieval importante da Índia.

Leia mais sobre Bhaskara em: <*https://www.estudopratico.com.br/formula-de-bhaskara-origem-importancia-e-exemplos/*>.

No geral, é difícil encontrar zeros de funções complicadas por meio de fórmulas. Por exemplo, se quiséssemos calcular os zeros da função

$$f(x) = x^5 + 2x^3 - 3x - cos(x) + e^x,$$

certamente teríamos bastante dificuldade de encontrar a resposta exata.

Muitos problemas matemáticos que aparentemente não estão atribuídos ao cálculo de zero de função podem ser resolvidos através do cálculo de zeros de função, como veremos.

Suponha que desejemos encontrar a intersecção entre as funções $y = f(x)$ e $y = g(x)$. Dessa forma, teríamos que resolver o problema de encontrar os valores de $x$ que satisfazem a equação $f(x) = g(x)$. Como, nesse caso, temos $f(x) - g(x) = 0$, então podemos interpretar este problema como a busca pelo zero da função $h(x)$ dada pela diferença entre $f$ e $g$, isto é: $h(x) = f(x) - g(x)$.

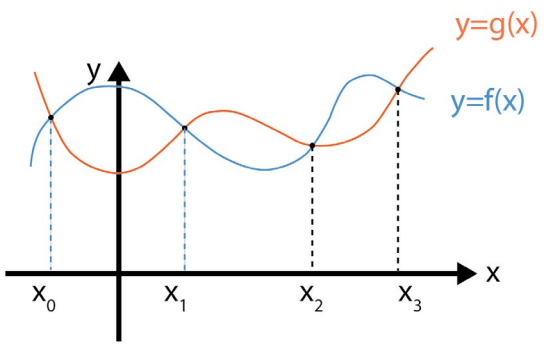



**Figura 2.4:** Gráfico da intersecção entre as funções $y = f(x)$ e $y = g(x)$

Para entendermos isso melhor, vejamos um típico problema da disciplina de Cálculo a Uma Variável. Se quiséssemos encontrar a área de regiões delimitada por funções, teríamos que encontrar as intersecções; como por exemplo, quando queremos encontrar a área da região delimitada pelas funções $f(x) = e^x$ e $g(x) = \frac{1}{x}$ e a reta vertical $x = 1$.

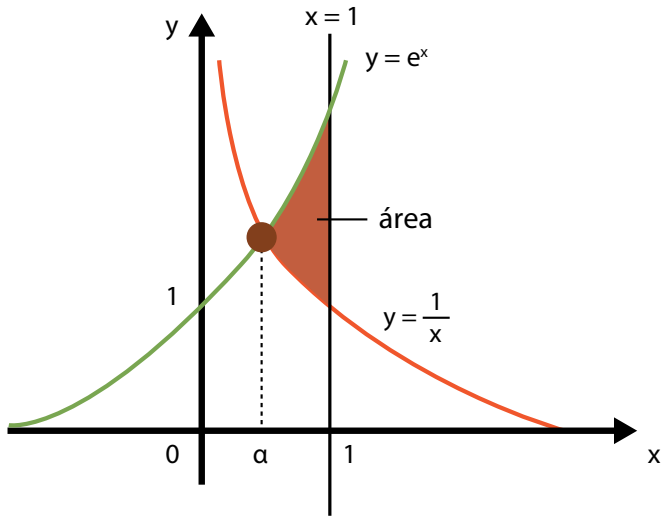



**Figura 2.5:** Gráfico da região delimitada pelas funções $f(x) = e^x$ e $g(x) = \frac{1}{x}$ e pela reta vertical $x = 1$.

Para resolver este problema, teríamos que calcular a integral dada por:

$$\int_a^1 (e^x - \frac{1}{x})dx,$$

em que o valor $x = a$ é a solução da equação $f(x) = g(x)$, isto é, $e^x = \frac{1}{x}$. Isso seria o mesmo que resolver a equação $e^x - \frac{1}{x} = 0$. Dessa forma, podemos definir a função $h(x) = e^x - \frac{1}{x}$ e encontrar o seu zero.

Vimos assim, a importância de termos estratégias bem definidas para encontrar os zeros de funções. Lembrando que, em primeiro lugar, é necessário localizar em que intervalo o zero da função se encontra. No exemplo anterior, não é difícil verificar que o valor $a \in [0, 1]$.

## Isolamento das Raízes

Uma análise gráfica será útil e muitas vezes necessária. Quando a função for contínua, uma coisa importante é fazer um estudo de sinais da função. Por exemplo, consideremos o polinômio de terceiro grau dado por

$$p(x) = x^3 - 9x + 3.$$

Nesse caso, podemos chamar os zeros de raízes do polinômio $p(x)$. O *Teorema Fundamental da Álgebra* nos diz que *todo polinômio de grau ímpar possui pelo menos um zero real*. Isso pode ser visto quando calculamos os limites abaixo:

$$\lim_{x\to+\infty}\left(x^3 - 9x + 3\right) = \lim_{x\to+\infty} x^3\left(1 - \frac{9}{x^2} + \frac{3}{x^3}\right) = \lim_{x\to+\infty} x^3 = +\infty$$

e

$$\lim_{x\to-\infty}\left(x^3 - 9x + 3\right) = \lim_{x\to-\infty} x^3\left(1 - \frac{9}{x^2} + \frac{3}{x^3}\right) = \lim_{x\to-\infty} x^3 = -\infty.$$

Como o polinômio $p(x)$ é uma função contínua, então, para ir de "$-\infty$" a "$+\infty$", $p(x)$ precisa cortar o eixo $x$ pelo menos uma vez. Observe o estudo de sinais do polinômio $p(x) = x^3 - 9x + 3$.

| $x$ | −4 | −3 | −2 | −1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|
| $p(x)$ | −25 | 3 | 13 | 11 | 3 | −5 | −7 | 3 |
| Sinal de $p(x)$ | − | + | + | + | + | − | − | + |

Observando a tabela de estudo de sinais, podemos identificar onde estão localizadas as três raízes do polinômio. Para isso, basta analisar a mudança de sinal de $p(x)$. Com isso, esse polinômio tem raízes $x_0 \in [-4, -3]$, $x_1 \in [0, 1]$ e $x_2 \in [2, 3]$.

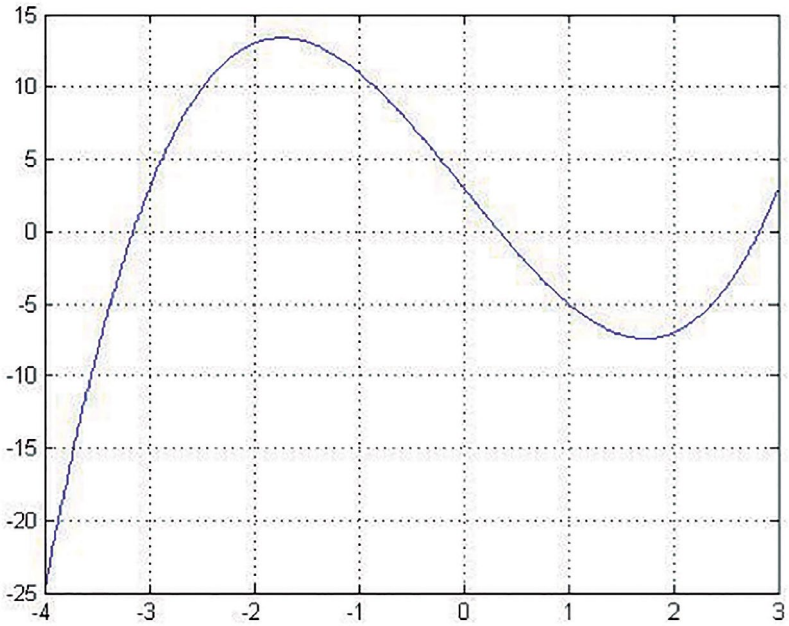



**Figura 2.6:** Gráfico do Polinômio $p(x) = x^3 - 9x + 3$. Observe que o polinômio cruza o eixo $x$ três vezes, mostrando as três raízes da equação $x^3 - 9x + 3 = 0$.

A explicação desse fato se deve ao Teorema de Bolzano, que é um caso particular do famoso Teorema do Valor Intermediário, visto na disciplina de Cálculo a Uma Variável. Vamos relembrar!

#Teorema de Bolzano: seja $f(x)$ uma função contínua em um intervalo $[a, b]$. Se

$$f(a) \times f(b) < 0,$$

então existe pelo menos um $x_0 \in (a, b)$ tal que $f(x_0) = 0$. Ou seja, a função tem pelo menos um zero entre $a$ e $b$.

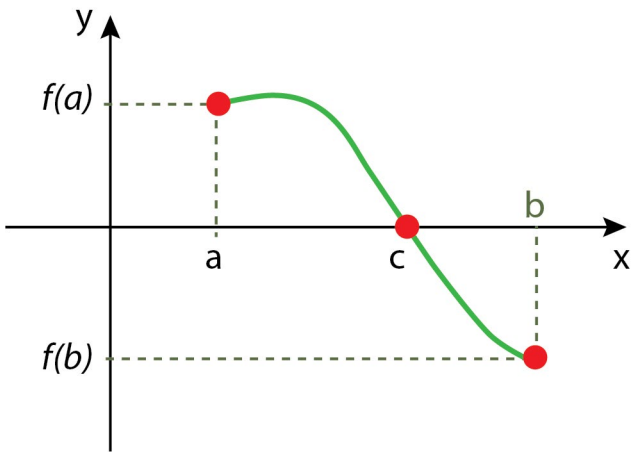



**Figura 2.7:** Teorema de Bolzano. A função *f(x)* atravessa o eixo *x* dentro do intervalo *[a, b]*.

Agora que estudamos maneiras para localizar um zero de uma função, vamos aprender alguns métodos numéricos, dentre os diversos que existem, para encontrar zeros de funções.

**Método iterativo**

Procedimento, em matemática computacional, que gera uma sequência de soluções aproximadas que vão melhorando conforme são executadas iterações, e resolvem uma classe de problemas estabelecida. Uma implementação específica de um método iterativo, incluindo o critério para a parada é um algoritmo iterativo. Um método iterativo é considerado convergente se a sequência correspondente converge, dado uma tolerância inicial de aproximação. Geralmente, é efetuada uma análise rigorosa de convergência de um método iterativo; no entanto, métodos iterativos baseados em heurísticas são comuns.

Fonte: <https://pt.wikipedia.org/wiki/M%C3%A9todo_iterativo>.

Os métodos apresentados nessa aula são conhecidos como **métodos iterativos**. Isso significa que estabelecemos uma expressão, a ser aplicada repedidas vezes, a partir de uma aproximação inicial ("*chute*" inicial), produzindo uma sequência de aproximações que convergem para a solução do problema. Cada vez que executamos um ciclo do método, damos o nome de *iteração*. Cada iteração começa a partir dos dados da iteração anterior, sucessivamente. Para que esse método tenha um fim, temos que adotar algum critério de parada, caso contrário o método entra em *looping* infinito (repetição infinita).

## Critérios de Parada

Após efetuar diversas iterações, precisamos saber se a solução aproximada está suficientemente próxima da solução exata do problema. Para isso, precisamos efetuar um teste; isto é, estabelecer um critério de parada para o método numérico. Para isso, precisamos aprender o significado de solução aproximada.

Existem basicamente duas interpretações para zero aproximado, mas que nem sempre levam ao mesmo resultado.

Seja $\overline{x}$ uma aproximação para o zero $r$ de uma função $f(x)$ com uma precisão $\varepsilon$. Então, podemos efetuar um dos dois testes:

$$\text{(i)} \; |\, \overline{x} - r \,| < \varepsilon$$

ou

$$\text{(ii)} \; |\, f(\overline{x}) \,| < \varepsilon.$$

Mas como efetuar o teste (i) se ainda não conhecemos a solução $r$, isto é, o zero da função $f(x)$? Uma maneira seria reduzir o intervalo que contém o zero a cada iteração. Além disso, nem sempre é possível ter as exigências (i) e (ii) satisfeitas, simultaneamente. Os gráficos a seguir ilustram algumas possibilidades:

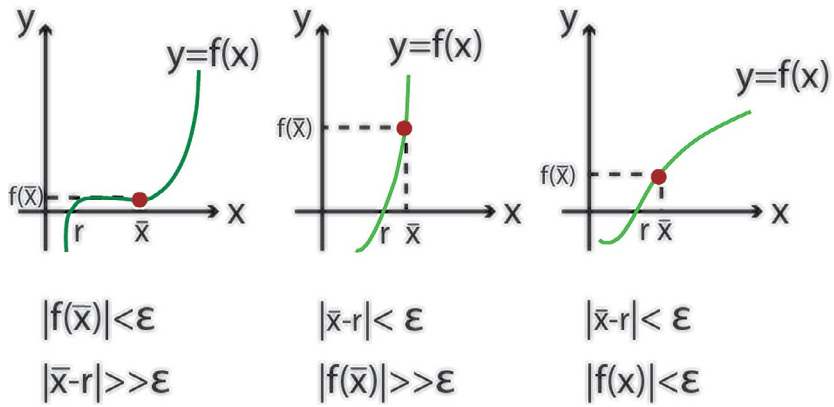



**Figura 2.8:** Análise gráfica dos possíveis critérios de parada.

Observe que, no primeiro caso, temos uma aproximação $\overline{x}$ distante do zero $r$, mas com imagem $f(\overline{x})$ próximo do valor zero. Já no segundo caso, temos o oposto; isto é, temos uma aproximação $\overline{x}$ perto do zero $r$, mas com imagem $f(\overline{x})$ distante do valor zero. Já no terceiro caso, temos uma situação em que os dois critérios de parada são satisfeitos, isto é, a aproximação $\overline{x}$ perto do zero $r$ e com imagem $f(\overline{x})$ próximo do valor zero.

Os métodos numéricos para encontrar zeros de função são desenvolvidos de modo a satisfazer pelo menos um dos critérios de parada.

Quando utilizamos um programa computacional, é aconselhável estipular um número máximo de iterações, além de um dos critérios de parada. Isso irá evitar que o algoritmo entre em um *looping*.

Estudaremos agora alguns métodos iterativos para encontrar o zero de funções. O método da *bissecção* e o *método da posição falsa* são conhecidos como métodos de quebra, pois se baseiam na quebra do intervalo que contém o zero de uma função.

Os métodos de quebra são os mais intuitivos, geometricamente, contudo, são os que convergem mais lentamente para a solução. A partir de um intervalo *[a, b]* que contenha um zero para a equação $f(x) = 0$, divide-se este intervalo em outros menores que ainda contenham pelo menos um zero da equação $f(x) = 0$. É necessário que a função $f$ troque de sinal no intervalo inicial e seja contínua dentro do intervalo.

## Método da Bissecção

**Bissecção**

Divisão em duas partes iguais.

Fonte: https://dicionario.priberam.org/bissecção.

O **método da bissecção** é um método de busca de raízes que bissectam um intervalo, isto é, que dividem repetidamente o intervalo que contém a raiz em duas partes iguais e, então, selecionam um subintervalo contendo a raiz para continuar o processo iterativo até que um critério de parada seja satisfeito.

É um método bastante simples e robusto, porém relativamente lento, quando comparado a outros, como o *método de Newton*, que veremos na próxima aula. Por este motivo, o método da bissecção é usado muitas vezes para encontrar uma primeira aproximação de uma solução e então utilizado como ponto inicial para métodos que convergem de forma mais rápida. O método também é chamado de método da pesquisa binária ou *método da dicotomia*.

De maneira mais formal, o método da bissecção, inspirado no Teorema de Bolzano, começa com um intervalo $[a, b]$ que contenha uma raiz para a equação

$$f(x) = 0,$$

em que a função contínua $f$ satisfaz $f(a) \times f(b) < 0$, ou seja, $f(x)$ corta o eixo $x$ em pelo menos um ponto do intervalo *[a, b]*. O objetivo deste método é dividir sucessivamente o intervalo ao meio, de maneira que o zero sempre esteja dentro do novo intervalo. O processo iterativo continua até que a amplitude do intervalo $[a, b]$ atinja a precisão requerida, dada por $(b - a) < \varepsilon$.

Vamos lá! Pense no épsilon como um número positivo e pequeno (algo do tipo 0,00000...001), isto é, perto de zero. Por mais próximos que os extremos *a* e *b* estejam um do outro, é sempre possível encontrar um número exatamente no meio dos dois que será dado por $(a + b)/2$. Se não colocarmos um fim nessa história (critério de parada), o método nunca terminará (*loop* infinito). Dizer que $(b - a) < \varepsilon$ é o mesmo que dizer que não vale a pena continuar, pois *a* e *b* estão muito próximos um do outro. Gastaremos tempo (mais iterações) e a qualidade na solução não irá melhorar. O épsilon é como se fosse zero para o computador.

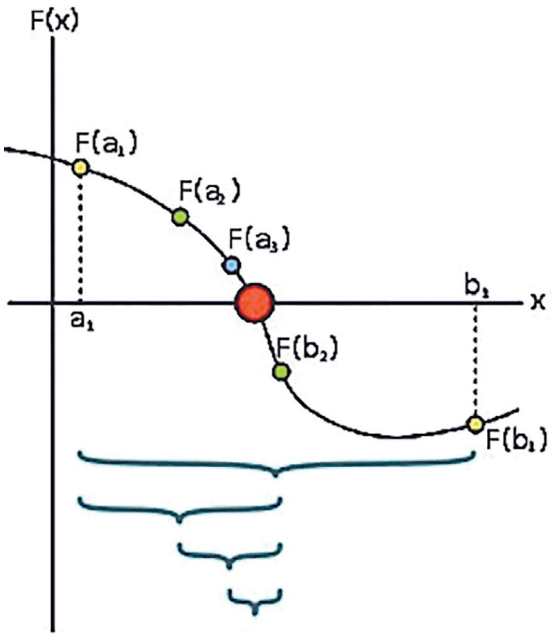



**Figura 2.9:** Análise gráfica do método da bissecção.

Para simplificar, vamos assumir que o intervalo *[a, b]* contenha apenas um único zero da equação $f(x) = 0$, em que $f(x)$ é contínua em *[a, b]*, com $f(a) \times f(b) < 0$. Os passos do algoritmo do método da bissecção são dados a seguir:

## Algoritmo: Método da Bissecção

- Dados iniciais: Seja $f(x)$ contínua no intervalo $[a, b]$ tal que $f(a) \times f(b) < 0$. Escolha a precisão $\varepsilon > 0$;
- Passo 1: se $(b - a) < \varepsilon$, então escolha $\overline{x} \in [a, b]$ e pare. Senão, vá para o Passo 2;
- Passo 2: faça $k = 1$;
- Passo 3: calcule $\overline{x} = \frac{a+b}{2}$;
- Passo 4: se $f(a) \times f(\overline{x}) > 0$, então faça $a = \overline{x}$ e vá ao passo 6. Senão, vá para o Passo 5;
- Passo 5: faça $b = \overline{x}$ e vá ao passo 6;
- Passo 6: se $(b - a) < \varepsilon$, então escolha $\overline{x} \in [a, b]$ e pare. Senão, vá para o Passo 7;
- Passo 7: faça $k = k + 1$ e volte ao Passo 3.

Observação: podemos incluir no algoritmo o critério de parada com o módulo da função e um número máximo de iterações.

Em cada iteração do algoritmo do método da bissecção, o intervalo $[a, b]$ será reduzido pela metade. O ponto médio $\overline{x} = \frac{a+b}{2}$, dado no passo 3 do algoritmo, ocupará o lugar do limitante inferior "$a$" ou do limitante superior "$b$" do intervalo. Tudo isso dependerá do sinal do valor de $f(\overline{x})$. De duas opções, teremos uma: ou $f(\overline{x})$ terá o mesmo sinal de $f(a)$ ou terá o mesmo sinal de $f(b)$.

Para garantir que o zero da função esteja dentro do novo intervalo reduzido pela metade, os limitantes do intervalo precisam ter sinais opostos.

Logo, se $f(a) \times f(\overline{x}) < 0$, isto é, $f(a)$ e $f(\overline{x})$ tem sinais opostos, então ficaremos com o intervalo $[a, \overline{x}]$. Nesse caso, descartaremos o limitante superior $b$ do intervalo, fazendo agora $b = \overline{x}$. Agora, por outro lado, se acontecer de $f(a) \times f(\overline{x}) > 0$, então significa que $f(a)$ e $f(\overline{x})$ têm o mesmo sinal (ambos negativos ou ambos positivos). Essa situação não nos daria garantia de que existe algum zero no intervalo $[a, \overline{x}]$. Mas, consequentemente, teremos garantido que o zero estaria no intervalo $[\overline{x}, b]$, pois quando $f(a) \times f(\overline{x}) > 0$ acontecer, $f(\overline{x}) \times f(b) < 0$ será verdadeiro. Nesse caso, descartaremos o limitante inferior $a$ do intervalo, fazendo agora $a = \overline{x}$. Vejamos essa explicação na seguinte ilustração:

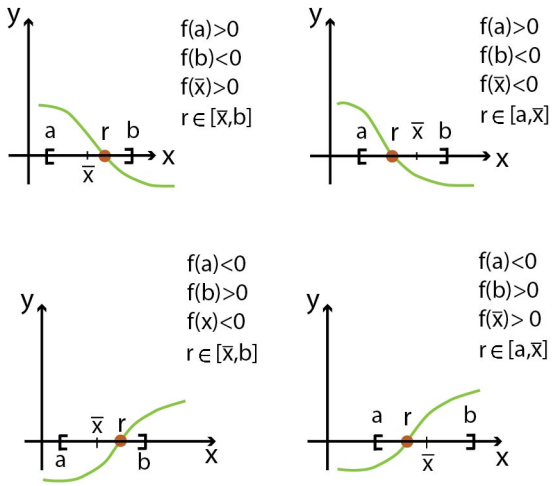



**Figura 2.10:** Análise gráfica do algoritmo do método da bissecção.

Vejamos como encontrar uma aproximação para $\sqrt{2}$, utilizando o método da bissecção.

## *Atividade 1*

***Atende ao objetivo 1***

Encontre uma aproximação para $\sqrt{2}$ utilizando o método da bissecção. Use quatro casas decimais.

### *Resposta comentada*

Observe que podemos modelar esse problema como sendo um problema de encontrar zero de uma função. Para isto, basta fazer $x = \sqrt{2}$. Então precisamos encontrar uma aproximação para $x$. Observe que, se $x = \sqrt{2}$, então $x^2 = 2$ ou seja, $x^2 - 2 = 0$. Dessa forma, basta escolhermos a função $f(x) = x^2 - 2$ e resolvermos o problema de encontrar $x$ que satisfaça a equação $f(x) = 0$.

Além disso, sabemos que $\sqrt{1} < \sqrt{2} < \sqrt{4}$, ou seja, $1 < \sqrt{2} < 2$. Nesse momento, já temos os dados iniciais para utilizar o algoritmo do método da bissecção, pois $f(x)$ é contínua (função polinomial do segundo grau) no intervalo [1, 2] com $f(1) = 1^2 - 2 = -1$ e $f(2) = 2^2 - 2 = 2$, isto é, $f(1) \times f(2) < 0$ (imagens com sinais opostos). Se fôssemos utilizar um computador precisaríamos definir uma precisão $\varepsilon = 10^{-3} = 0{,}001$, por exemplo. Mas como se trata de um exercício em que faremos as contas com uso de uma calculadora científica, vamos nos limitar a utilizar quatro casas decimais e fazer apenas sete iterações, para entendermos como funciona o método da bissecção. Vamos às contas:

$k = 1$ (primeira iteração)

$I_1 = [1; 2]$

$a_1 = 1 \Rightarrow f(1) = -1 < 0$

$b_1 = 2 \Rightarrow f(2) = +2 > 0$

$x_1 = \dfrac{1 + 2}{2} = 1{,}5 \Rightarrow f(1{,}5) = +0{,}25 > 0$

O ponto médio $x = 1{,}5$ do intervalo [1, 2] ocupará o lugar do limite inferior $a = 1$ ou do limite superior $b = 2$. Ou seja, devemos ficar com o intervalo [1; 1,5] ou o intervalo [1,5; 2]. Como $f(1{,}5) > 0$, então $x_1 = 1{,}5$ ocupará o lugar de $b = 2$, pois a imagem tem o mesmo sinal, isto é, $f(2) > 0$. Dessa forma, ficaremos com o intervalo [1; 1,5], pois $f(1) \times f(1{,}5) < 0$. Com isso, basta repetir o processo para esse novo intervalo, e daí por diante.

$k = 2$ (segunda iteração)

$I_2 = [1; 1{,}5]$

$a_2 = 1 \Rightarrow f(1) = -1 < 0$

$b_2 = 1{,}5 \Rightarrow f(1{,}5) = +0{,}25 > 0$

$x_2 = \dfrac{1 + 1{,}5}{2} = 1{,}25 \Rightarrow f(1{,}25) = -0{,}4375 < 0$

$k = 3$ (terceira iteração)

$I_3 = [1{,}25;\ 1{,}5]$

$a_3 = 1{,}25 \Rightarrow f(1{,}25) = -0{,}4375 < 0$

$b_3 = 1{,}5 \Rightarrow f(1{,}5) = +0{,}25 > 0$

$x_3 = \dfrac{1{,}25 + 1{,}5}{2} = 1{,}375 \Rightarrow f(1{,}375) \approx -0{,}1094 < 0$

$k = 4$ (quarta iteração)

$I_4 = [1{,}375;\ 1{,}5]$

$a_4 = 1{,}375 \Rightarrow f(1{,}375) \approx -0{,}1094 < 0$

$b_4 = 1{,}5 \Rightarrow f(1{,}5) = +0{,}25 > 0$

$x_4 = \dfrac{1{,}375 + 1{,}5}{2} = 1{,}4375 \Rightarrow f(1{,}4375) \approx +0{,}0664 > 0$

$k = 5$ (quinta iteração)

$I_5 = [1{,}375;\ 1{,}4375]$

$a_5 = 1{,}375 \Rightarrow f(1{,}375) \approx -0{,}1094 < 0$

$b_5 = 1{,}4375 \Rightarrow f(1{,}4375) \approx +0{,}0664 > 0$

$x_5 = \dfrac{1{,}375 + 1{,}4375}{2} \approx 1{,}4063 \Rightarrow f(1{,}4063) \approx -0{,}0223 < 0$

$k = 6$ (sexta iteração)

$I_6 = [1{,}4063;\ 1{,}4375]$

$a_6 = 1{,}4063 \Rightarrow f(1{,}4063) \approx -0{,}0223 < 0$

$b_6 = 1{,}4375 \Rightarrow f(1{,}4375) \approx +0{,}0664 > 0$

$x_6 = \dfrac{1{,}4063 + 1{,}4375}{2} \approx 1{,}4219 \Rightarrow f(1{,}4219) \approx +0{,}0218 > 0$

$k = 7$ (sétima iteração)

$I_7 = [1{,}4063;1{,}4219]$

$a_7 = 1{,}4063 \Rightarrow f(1{,}4063) \approx -0{,}0223 < 0$

$b_7 = 1{,}4219 \Rightarrow f(1{,}4219) \approx +0{,}0218 > 0$

$x_7 = \dfrac{1{,}4063 + 1{,}4219}{2} \approx 1{,}4141 \Rightarrow f(1{,}4141) \approx -0{,}0003 < 0$

Após sete iterações, concluímos que $\sqrt{2} \approx 1{,}4141$. Obviamente, precisaríamos de mais iterações para encontrar uma melhor aproximação. Observe que a amplitude do último intervalo encontrado

$$I_7 = [a_7,\ b_7] = [1{,}4063;\ 1{,}4219]$$

é dada por $b_7 - a_7 = 1{,}4219 - 1{,}4063 = 0{,}0156 > \varepsilon = 0{,}001$. Se tivéssemos adotado o critério de parada $b_k - a_k < \varepsilon$, ainda precisaríamos realizar

mais iterações, até que este critério fosse satisfeito. Por outro lado, se tivéssemos adotado o critério de parada $|f(x_k)| < \varepsilon$, então poderíamos parar nessa sétima iteração, pois:

$$|f(x_7) = |f(1{,}4141)| \approx 0{,}0003 < \varepsilon$$

Nesse momento, podemos nos indagar se é possível estimar o número de iterações a ser realizada no método da bissecção, caso optássemos pelo critério de parada $b_k - a_k < \varepsilon$. Vejamos agora como podemos fazer essa estimativa.

## Estimativa do Número de Iterações no Método da Bissecção

Repare que cada intervalo possui a metade da amplitude do intervalo anterior. Então, dada uma precisão $\varepsilon > 0$ e um intervalo inicial $|a, b|$, é possível estimar o número de iterações $k$ a ser realizada pelo método da bissecção, de modo que o critério de parada $b_k - a_k < \varepsilon$ seja satisfeito.

Seguindo os passos do Algoritmo do Método da Bissecção, podemos escrever:

$$b_k - a_k = \frac{b_{k-1} - a_{k-1}}{2} = \frac{b_{k-2} - a_{k-2}}{2^2} = \frac{b_{k-3} - a_{k-3}}{2^3} = \ldots \frac{b-a}{2^k} < \varepsilon.$$

O método da bissecção corta o intervalo sempre ao meio, onde iremos escolher (baseado no método) se ficaremos com a primeira ou a segunda metade do intervalo (faremos isso em cada iteração). Chamamos o valor (b − a) de amplitude de um [a, b]. As desigualdades anteriores significam que serão sucessivas divisões por 2, isto é, 2, 4, 8, 16, etc. Haverá um momento em que a amplitude do intervalo atual será tão pequena (menor que $\varepsilon$), que pararemos (critério de parada).

Daí, temos que:

$$2^k > \frac{b-a}{\varepsilon} \Rightarrow log\left(2^k\right) > log\left(\frac{b-a}{\varepsilon}\right) \Rightarrow k \cdot log\left(2\right) > log\left(b-a\right) - log\left(\varepsilon\right)$$

$$\Rightarrow k > \frac{log\left(b-a\right) - log\left(\varepsilon\right)}{log\left(2\right)}.$$

Portanto, se *k* satisfaz a relação acima, teremos a estimativa do número mínimo de iterações necessárias para atingir uma precisão $\varepsilon$. Isso se deve ao fato de que se é o zero da função, o critério de parada $|\overline{x} - r| < \varepsilon$ será atendido, pois temos que:

$$|\overline{x} - r| < |b_k - a_k| < \varepsilon$$

Graficamente ,temos:

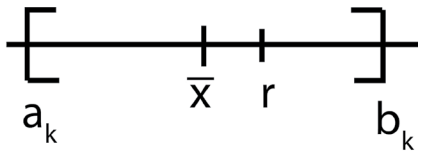



**Figura 2.11:** Análise gráfica do critério de parada do método da bissecção.

Na **Atividade 1**, se tivéssemos optado pelo critério de parada $|\overline{x} - r| < \varepsilon$ com $\varepsilon = 10^{-3} = 0{,}001$, por exemplo, teríamos que fazer muito mais iterações. Vejamos como descobrir quantas iterações seriam necessárias, no exemplo a seguir:

## *Atividade 2*

***Atende ao objetivo 1***

Calcule o número mínimo de iterações necessárias para encontrar o zero da função $f(x) = x^2 - 2$ no intervalo [1, 2], utilizando a precisão $\varepsilon = 10^{-3}$.

***Resposta comentada***

Vimos que o número $k$ mínimo de iterações necessárias para atingir uma precisão ε é dado por $k > \frac{log(b-a) - log(\varepsilon)}{log(2)}$. Com $a = 1$, $b = 2$ e $\varepsilon = 10^{-3}$, teremos:

$$k > \frac{log(2-1) - log(10^{-3})}{log(2)} = \frac{log(1) + 3log(10)}{log(2)} = \frac{3}{log(2)} \approx \frac{3}{0,3010}$$

$$\approx 9,9668 \Rightarrow k = 10.$$

Logo, precisaríamos realizar pelo menos 10 iterações.

Uma vez satisfeita a hipótese de continuidade da função $f(x)$ no intervalo $[a, b]$ com $f(a) \times f(b) < 0$, o método da bissecção irá gerar uma sequência que aproximações que convergirá para a solução do problema sem realizar cálculos mirabolantes. No entanto, a convergência desse método é bastante lenta se comparada outros métodos para encontrar zero de função.

O *método da posição falsa* foi inspirado no método da bissecção, com o objetivo de acelerar a convergência realizando um número menor de iterações.

## Método da Posição Falsa

Vimos que o método da Bissecção consiste em reduzir pela metade um intervalo $[a, b]$ que contém um zero de uma função $f(x)$. Para isso, o método utiliza-se de uma média aritmética entre $a$ e $b$, isto é, $\overline{x} = \frac{a+b}{2}$. Essa média não leva em consideração nenhuma propriedade da função $f(x)$ para *tentar descobrir se o zero está mais perto de a ou de b.* É exatamente isso que o *método da posição falsa* tenta fazer, utilizando uma média ponderada, ao invés de aritmética. Para simplificar, vamos novamente assumir que o intervalo $(a, b)$ contenha apenas um único zero da equação $f(x) = 0$, onde $f(x)$ é contínua em $[a, b]$ com $f(a) \times f(b) < 0$. Se $|f(a)|$ estiver mais próximo de zero do que $|f(b)|$, então é provável que o zero da função esteja mais próximo de $a$ do que de $b$. O contrário também seria esperado, isto é, se $|f(b)|$ estiver mais

próximo de zero do que $|f(a)|$, então é provável que o zero da função esteja mais próximo de $b$ do que de $a$ (pelo menos, isso aconteceria se a função fosse uma reta). Observe o gráfico ilustrando isso:

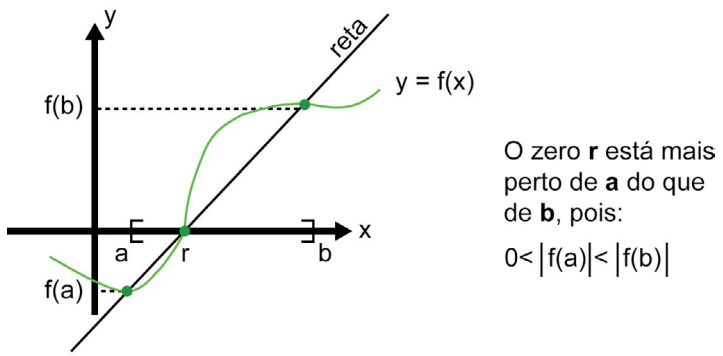



**Figura 2.12:** Análise gráfica do método da posição falsa.

Dessa forma, ao invés de utilizar a média aritmética como no caso do método da bissecção, o método da posição falsa utiliza a média ponderada dada por:

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

Observe que $|f(a)|$ e $|f(b)|$ são usados como pesos na média ponderada. Quanto maior for o valor de $|f(b)|$, isso fará com a média $\overline{x}$ fique mais próxima do limite inferior $a$ do intervalo $[a, b]$. Por outro lado, quanto maior for o valor de $|f(a)|$ mais próxima ficará a média $\overline{x}$ do limite superior $b$ do intervalo $[a, b]$. Além disso, como por hipótese $f(a)$ e $f(b)$ possuem sinais opostos, então podemos reescrever a média ponderada como:

$$\overline{x} = \frac{af(b) - bf(a)}{f(b) - f(a)}$$

Provavelmente você já se perguntou por que o método chama-se *posição falsa*. Isso se deve ao fato de que, dependendo da curvatura da função $f(x)$ no intervalo $[a, b]$ que contém o zero da função, ao invés de nos aproximarmos do zero, podemos nos afastar dele. Isso pode ser observado no gráfico a seguir:

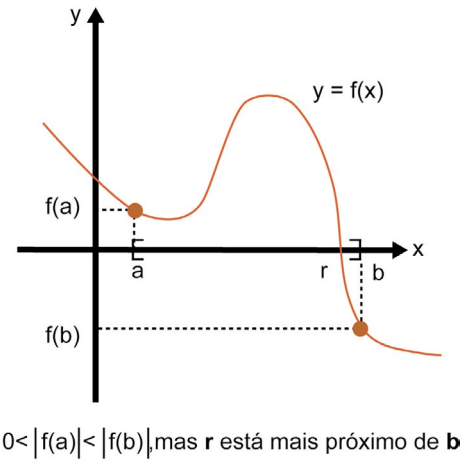



**Figura 2.13:** O motivo do nome *método da posição falsa*.

Os passos do algoritmo são dados a seguir:

## Algoritmo: Método da Posição Falsa

- Dados iniciais: Seja $f(x)$ contínua no intervalo $[a, b]$ tal que $f(a) \times f(b) < 0$.

  Escolha a precisão $\varepsilon > 0$;
- Passo 1: se $(b - a) < \varepsilon$, então escolha $\overline{x} \in [a, b]$ e pare;

  Senão,
- Passo 2: faça $k = 1$;
- Passo 3: calcule $\overline{x} = \dfrac{af(b) - bf(a)}{f(b) - f(a)}$;
- Passo 4: se $|f(\overline{x})| < \varepsilon$, então escolha $\overline{x}$ como a solução do problema e pare;

  Senão,
- Passo 4: se $f(a) \times f(\overline{x}) > 0$, então faça $a = \overline{x}$ e vá ao passo 6;

  Senão,
- Passo 5: faça $b = \overline{x}$ e vá ao passo 6;
- Passo 6: se $(b - a) < \varepsilon$, então escolha $\overline{x} \in [a, b]$ como solução e pare;

  Senão,
- Passo 7: faça $k = k + 1$ e volte ao passo 3.

No início desta aula, quando fizemos um estudo de sinais do polinômio do terceiro grau dado por $p(x) = x^3 - 9x + 3$, vimos que existe um zero no intervalo [0, 1]. Vamos utilizar o método da posição falsa para encontrar uma aproximação para esse zero.

## Atividade 2

*Atende ao objetivo 2*

Utilize o *método da posição falsa* para encontrar o zero da função $p(x) = x^3 - 9x + 3$ que está no intervalo [0, 1] utilizando a precisão $\varepsilon = 10^{-3}$:

### Resposta comentada

Já temos os dados iniciais para utilizar o algoritmo do Método da Posição Falsa, pois $f(x)$ é contínua (função polinomial do terceiro grau) no intervalo [0, 1] com $f(0) = 3 > 0$ e $f(1) = -5 < 0$, isto é, $f(0) \times f(1) < 0$ (imagens com sinais opostos). Vamos às contas:

$$k = 1 \left(\text{primeira iteração}\right)$$
$$I_1 = \left[0;1\right]$$
$$a_1 = 0 \Rightarrow f\left(0\right) = +3 > 0$$
$$b_1 = 1 \Rightarrow f\left(1\right) = -5 < 0$$
$$x_1 = \frac{\left(0\right)\left(-5\right) - \left(1\right)\left(3\right)}{\left(-5\right) - \left(3\right)} = \frac{-3}{-8} = 0{,}375 \Rightarrow f\left(0{,}375\right) \approx -0{,}3223 < 0$$

A média ponderada $x_1 = 0{,}375$ do intervalo [0, 1] ocupará o lugar do limite inferior $a = 0$ ou do limite superior $b = 1$. Ou seja, devemos ficar com o intervalo [0; 0,375] ou o intervalo [0,375; 1]. Como $f(0{,}375) < 0$, então $x_1 = 0{,}375$ ocupará o lugar de $b = 1$, pois a imagem tem o mesmo sinal, isto é, $f(1) < 0$. Dessa forma, ficaremos com o intervalo [0; 0,375], pois $f(0) \times f(0{,}375) < 0$. Com isto, basta repetir o processo para esse novo intervalo e daí por diante.

$k = 2 \left(\text{segunda iteração}\right)$
$I_2 = \left[0; 0{,}375\right]$
$a_2 = 0 \Rightarrow f\left(0\right) = +3 > 0$
$b_2 = 0{,}375 \Rightarrow f\left(0{,}375\right) \approx -0{,}3223 < 0$
$$x_2 = \frac{\left(0\right)\left(-0{,}3223\right) - \left(0{,}375\right)\left(3\right)}{\left(-0{,}3223\right) - \left(3\right)} \approx 0{,}3386$$
$\Rightarrow f\left(0{,}3386\right) \approx -0{,}0086 < 0$

$k = 3 \left(\text{terceira iteração}\right)$
$I_3 = \left[0; 0{,}3386\right]$
$a_3 = 0 \Rightarrow f\left(0\right) = +3 > 0$
$b_3 = 0{,}3386 \Rightarrow f\left(0{,}3386\right) \approx -0{,}0086 < 0$
$$x_3 = \frac{\left(0\right)\left(-0{,}0086\right) - \left(0{,}3386\right)\left(3\right)}{\left(-0{,}0086\right) - \left(3\right)} = 0{,}3376$$
$\Rightarrow f\left(0{,}3376\right) \approx -0{,}0001 < 0$

E, portanto, $\overline{x}$ = 0,3376 é uma aproximação para o zero da função que está no intervalo [0, 1], pois $|f(0{,}3376)| = 0{,}0001 < 0{,}001 = \varepsilon$.

## Conclusão

Nesta aula, vimos dois métodos para encontrar zero de função. Um deles é o *método da bissecção*, que tem enorme simplicidade, mas que é bastante lento na convergência. O *método da posição falsa* é uma adaptação do *método da bissecção*, com o objetivo que acelerar a convergência, mas, como o próprio nome diz, podemos nos afastar da solução dependendo da curvatura da função do problema. Existem muitos outros métodos que podemos utilizar para encontrar o zero de uma função. Na próxima aula, estudaremos um dos métodos mais famosos para esse tipo de problema. Estamos falando do *método de Newton-Raphson*.

## Resumo

Nessa aula você estudou:

- o método da bissecção, que consiste em reduzir o intervalo [*a, b*]em cada iteração, utilizando para isso, a média aritmética:

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

- o *método da posição falsa*, que consiste em reduzir o intervalo [*a, b*] em cada iteração, utilizando para isso a média ponderada:

$$\bar{x} = \frac{af(b) - bf(a)}{f(b) - f(a)}$$

## Referências

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R. *Cálculo Numérico:* aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

BURDEN, Richard L.; FAIRES, Douglas. *Análise numérica*. São Paulo: Pioneira Thomson Learning, 2003.

# Aula 3

## Zeros reais de funções reais: método de Newton-Raphson

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

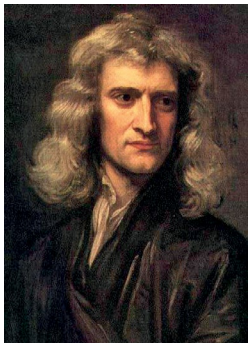
Apresentar o Método de Newton-Raphson para o cálculo de zeros de funções.

## Objetivo

Esperamos que, ao final dessa aula, você seja capaz de calcular zeros de função através do Método de Newton-Raphson.

## Introdução

Na aula anterior estudamos dois métodos iterativos, entre os diversos existentes, para encontrar zeros de uma função. Nesta aula, estudaremos o mais famoso e um dos mais eficientes métodos para solucionar esse tipo de problema. Estamos falando do Método de Newton-Raphson, desenvolvido por Isaac Newton e Joseph Raphson. Na maioria das vezes, este método é chamado apenas de Método de Newton devido ao famoso matemático inglês Isaac Newton, que publicou seu método para encontrar zeros de equações não lineares em 1687. Para fazer jus ao também inglês Joseph Raphson, que apresentou em 1690 a sistematização do método, foi dado o nome de Método de Newton-Raphson. A versão de Raphson do método é mais simples do que a de Newton e é, por essa razão, considerada superior e encontrada nos livros atualmente.



## Isaac Newton

**(1643 – 1727)**

Newton é considerado um dos maiores gênios da história da humanidade. Ele nasceu em Woolsthorpe, Lincolnshire, Inglaterra, e estudou no Trinity College, em Cambridge. Foi o maior cientista europeu desde a época de Arquimedes até a de Albert Einstein. Antes de completar 24 anos de idade, ele já havia criado o teorema dos binômios e o cálculo funcional, descoberto o espectro da luz e escrito sua Teoria da Gravitação. Há quem afirme que ele teria elaborado esta teoria em 1665, ao observar a queda de uma maçã. Enquanto esteve em Cambridge, desenvolveu as famosas

Três Leis do Movimento, uma façanha espetacular para uma pessoa ainda tão jovem. Em 1687, Newton publicou seus *Princípios Matemáticos da Filosofia Natural*. Neste trabalho, universalmente conhecido como *Principia*, ele demonstrou a estrutura do Universo, o movimento dos planetas e calculou a massa do Sol, dos planetas e de algumas luas.

Fonte: http://www.sohistoria.com.br/biografias/newton/.

## Método de Newton-Raphson

O Método de Newton-Raphson combina duas ideias básicas muito comuns nas aproximações numéricas, que são a *linearização* e *iteração*. Quando estudamos cálculo diferencial, o primeiro conceito que temos de derivada é construir a equação de uma reta tangente a uma função em um determinado ponto. Tal reta tangente é uma aproximação linear da função na vizinhança do ponto tangente e é exatamente com essa linearização da função que o método de Newton trabalha.

Dessa forma, ao invés de calcular o zero da função, o método de Newton encontra o zero da reta que é muito mais fácil se encontrar através de operações básicas. Então, na vizinhança do ponto de tangência, substituímos um problema complicado por um problema simples de lidar.

A parte iterativa do método consiste em repetir esse processo toda vez que encontramos uma aproximação. Em suma, damos um "chute" (aproximação) inicial $x_0$ para o zero da função *f(x)* e calculamos o zero da reta tangente à $f(x)$ em $x_0$ chamando-o de $x_1$. Tal equação é dada por:

$$y = f'(x_0) \cdot (x - x_0) + f(x_0),$$

Para calcular a nova aproximação $x_1$, basta fazer $y = 0$ na equação da reta tangente acima e isolar a variável $x$, resolvendo o problema:

$$f'(x_0) \cdot (x - x_0) + f(x_0) = 0,$$

e encontrando, assim:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

A figura a seguir ilustra essa situação:

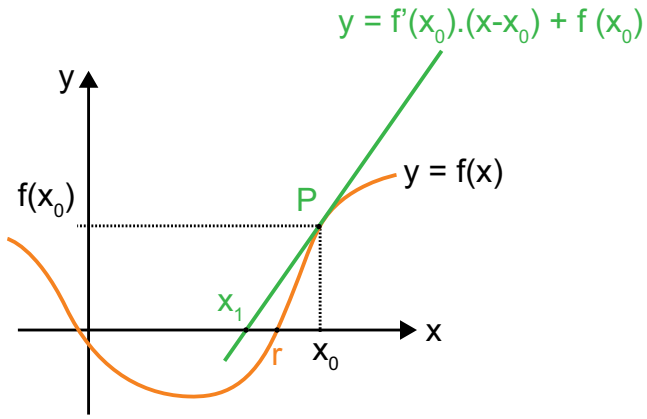



**Figura 3.1:** Reta tangente em $P(x_0, f(x_0))$.

Daí é só repetir o processo calculando o zero da reta tangente à $f(x)$ em $x_1$ chamando-o de $x_2$, isto é:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)},$$

e assim por diante. Dessa forma, iremos construir uma sequência de aproximações $x_k$, utilizando a aproximação $x_k$ para encontrar $x_{k+1}$ através da fórmula:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \text{ com } k = 0, 1, 2,\ldots$$

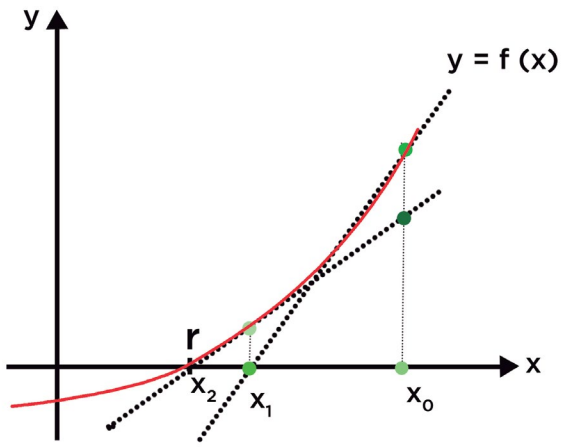



**Figura 3.2:** Processo iterativo do Método de Newton-Raphson.

Repetiremos esse processo até que a precisão desejada seja atingida, atendendo ao critério de parada escolhido. Por exemplo, podemos parar quando $f(x_k)$ for suficientemente próximo de zero ou quando a diferença entre os dois iterados, ou seja, $|x_{k+1} - x_k|$, for muito pequena, demonstrando assim pouca possibilidade de avanço nos cálculos.

O Método de Newton-Raphson está esquematizado no algoritmo a seguir.

## Algoritmo: Método de Newton-Raphson

- Dados iniciais: escolha a aproximação inicial $x_0$ e a precisão $\varepsilon > 0$. Calcule $f'(x)$ a partir de $f(x)$;
- Passo 1: se $|f(x_0)| < \varepsilon$, então faça $\overline{x} = x_0$ e pare. Senão, siga para o Passo 2;
- Passo 2: faça $k = 1$;
- Passo 3: calcule $x_1 = x_0 - \dfrac{f(x_0)}{f'(x_0)}$;
- Passo 4: se $|f(x_1)| < \varepsilon$ ou se $|x_1 - x_0| < \varepsilon$, então faça $\overline{x} = x_1$ e pare. Senão, siga para o Passo 5;
- Passo 5: faça $x_0 = x_1$;
- Passo 6: faça $k = k + 1$ e volte ao Passo 3.

Observação: Podemos incluir no Algoritmo uma salvaguarda: caso a derivada $|f'(x_k)| < \varepsilon$, para alguma iteração $k$, então devemos parar e escolher um novo chute $x_o$.

Essa observação destaca a possível falha no Método de Newton-Raphson. De forma algébrica, se $f'(x_k) = 0$, então não poderíamos calcular $x_{k+1}$, pois não podemos dividir por zero o lado direito da equação do Passo 3 ($x_{k+1} = x_k - f(x_k)/f'(x_k)$).

De forma geométrica, toda que vez que a derivada é nula, a reta tangente é paralela ao eixo $x$ e, dessa forma, não irá interceptá-lo. De modo prático, toda vez que a derivada estiver próximo de zero ($f'(x_k) \approx 0$), ou seja, $|f'(x_k)| < \varepsilon$, a reta tangente irá interceptar o eixo $x$ muito longe do zero $r$ da função.

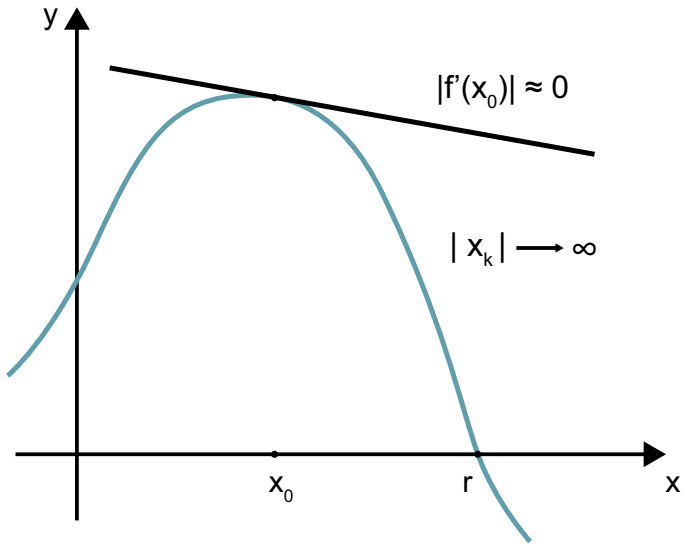



**Figura 3.3:** Possível falha no Método de Newton-Raphson.

Na primeira atividade da Aula 2, encontramos uma aproximação para $\sqrt{2}$ utilizando o método da Bissecção. Agora vamos fazer o mesmo, mas utilizando o Método de Newton-Raphson.

## Atividade 1

***Atende ao objetivo 1***

Encontre uma aproximação para $\sqrt{2}$ utilizando o Método de Newton-Raphson. Utilize quatro casas decimais.

Dica: para resolver o problema pelo método proposto, podemos continuar utilizando a mesma função que encontramos na primeira atividade da Aula 2, $f(x) = x^2 - 2$, da qual $\sqrt{2}$ é raiz.

### *Resposta comentada*

Na primeira atividade da Aula 2, vimos que $\sqrt{2} \in [1,2]$ é um zero da função $f(x) = x^2 - 2$.

Além disso, a derivada da função é dada por $f'(x) = 2x$. Escolheremos $x_0 = 1,5$ como aproximação inicial para $\sqrt{2}$ (zero da função). Com isso, teremos:

$$\begin{cases} k = 1 \ (\text{primeira iteração}) \\ x_0 = 1,5 \\ f(x_0) = f(1,5) = 0,25 \\ f'(x_0) = f'(1,5) = 3 \\ x_1 = x_0 - \dfrac{f(x_0)}{f'(x_0)} = 1,5 - \dfrac{(0,25)}{(3)} \approx 1,4167 \end{cases}$$

$$\begin{cases} k = 2 \ (\text{segunda iteração}) \\ x_1 = 1,4167 \\ f(x_1) = f(1,4167) \approx 0,0070 \\ f'(x_1) = f'(1,4167) = 2,8334 \\ x_2 = x_1 - \dfrac{f(x_1)}{f'(x_1)} = 1,4167 - \dfrac{(0,0070)}{(2,8334)} \approx 1,4142 \end{cases}$$

$$\begin{cases} k = 3 \ (\text{terceira iteração}) \\ x_2 = 1,4142 \\ f(x_2) = f(1,4142) \approx 0,0000 \\ f'(x_2) = f'(1,4142) = 2,8284 \\ x_3 = x_2 - \dfrac{f(x_2)}{f'(x_2)} = 1,4142 - \dfrac{(0,0000)}{(2,8284)} = 1,4142 \end{cases}$$

Note que podemos parar na terceira iteração, pois não houve avanço na aproximação ($x_3 = x_2$). Com isso, utilizando apenas quatro casas decimais, temos $\sqrt{2} \approx 1,4142$. Na verdade, já poderíamos ter parado na segunda iteração, pois $f(x_2) = f(1,4142) \approx 0$.

Na terceira atividade da Aula 2, calculamos uma aproximação para o zero do polinômio do terceiro grau dado por $p(x) = x^3 - 9x + 3$ que está no intervalo [0, 1]. Agora vamos utilizar o Método de Newton-Raphson para encontrar uma aproximação para esse zero.

## Atividade 2

***Atende ao objetivo 1***

Utilize o Método de Newton-Raphson para encontrar o zero da função $p(x) = x^3 - 9x + 3$ que está no intervalo [0,1], utilizando a precisão $\varepsilon = 10^{-3}$:

### Resposta comentada

A derivada da função é dada por $f'(x) = 3x^2 - 9$. Escolheremos $x_0 = 0,5$ como aproximação inicial. Como isso, temos:

$$\begin{cases} k = 1 \text{ (primeira iteração)} \\ x_0 = 0,5 \\ f(x_0) = f(0,5) = -1,375 \\ f'(x_0) = f'(0,5) = -8,25 \\ x_1 = x_0 - \dfrac{f(x_0)}{f'(x_0)} = 0,5 - \dfrac{(-1,375)}{(-8,25)} \approx 0,3333 \end{cases}$$

$$\begin{cases} k = 2 \left(\text{segunda iteração}\right) \\ x_1 = 0,3333 \\ f\left(x_1\right) = f\left(0,3333\right) \approx 0,0373 \\ f'\left(x_1\right) = f'\left(0,3333\right) \approx -8,6667 \\ x_2 = x_1 - \dfrac{f\left(x_1\right)}{f'\left(x_1\right)} = 0,3333 - \dfrac{\left(0,0373\right)}{\left(-8,6667\right)} \approx 0,3376 \end{cases}$$

Já podemos parar na segunda iteração com $\overline{x} = 0,3376$, pois já teremos o seguinte critério de parada atendido:

$$|f(x_2)| = |f(0,3376)| \approx 0,0001 < 0,001 = 10^{-3} = \varepsilon.$$

## Conclusão

Nesta aula, vimos o Método de Newton-Raphson para encontrar zero de função, um dos mais famosos e eficientes quando se trata desse tipo de problema.

Para que o Método de Newton-Raphson funcione bem, é necessário que o chute inicial esteja relativamente próximo da solução. Além disso, é muito importante que a derivada calculada nas aproximações não fique próxima de zero. Caso isso aconteça, basta recomeçar o método escolhendo um novo chute inicial.

## Resumo

Nesta aula você estudou o Método de Newton-Raphson, que consiste em calcular uma sequência de aproximações dada por a partir de um chute inicial .

## Referências

BURDEN, Richard L.; FAIRES, Douglas. *Análise numérica*. São Paulo: Pioneira Thomson Learning, 2003.

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R., *Cálculo numérico*: aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

# Aula 4

## Resolução de sistemas lineares: métodos diretos – método da eliminação de Gauss

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Apresentar os conceitos da resolução de sistemas lineares da forma direta, apresentando o método da eliminação de Gauss.

## Objetivos

Esperamos que, ao final desta aula, você seja capaz de:

1. identificar o que é um sistema linear e reescreve-lo na forma matricial;
2. entender cada passo do algoritmo numérico para o método da eliminação de Gauss;
3. resolver o sistema linear através do método direto da eliminação de Gauss;
4. escrever um algoritmo para o método de Gauss.

## Pré-requisitos

Para se ter um bom aproveitamento desta aula, é importante você relembrar os conceitos de *matrizes e sistemas lineares* apresentados nas disciplinas de Álgebra Linear I e II.

## Introdução

Existem dois tipos de métodos numéricos para encontrar soluções de um sistema linear, são eles:

- *Métodos diretos:* são aqueles que a menos de erros de arredondamento, fornecem a solução exata do sistema.
- *Métodos iterativos:* são aqueles que procuram a solução do sistema considerando uma aproximação inicial e gerando uma sequência iterativa $\{X_N\}$. Essa sequência pode ou não convergir para a solução do sistema.

## Sistemas lineares

Dados $a_1, ..., a_n$ e $b \in \mathbb{R}$ e a equação $a_1x_1 + ... + a_nx_n = b$ onde $x_1, ..., x_n$ são variáveis ou incógnitas, é denominada *equação linear nas variáveis* $x_1, ..., x_n$. Os números $a_1, ..., a_n$ são denominados *coeficientes das variáveis* $x_1, ..., x_n$, respectivamente, e $b$ é denominado de *termo independente*.

Um *sistema linear sobre* $\mathbb{R}$ com *m equações* e *n incógnitas* é um conjunto de $m \geq 1$ equações lineares com $n \geq 1$ variáveis, e é representado por:

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \\ \vdots \quad\quad\quad\quad\quad \vdots \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases}$$

com $a_{ij} \in \mathbb{R}$ para todo $i = 1, \ldots, m$ e $j = 1, ..., n$. Ou ainda, podemos representá-lo como um produto de matrizes:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

onde:

a matriz $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ é denominada *matriz dos coeficientes*;

a matriz $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ é denominada *matriz das variáveis*;

a matriz $B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ é denominada *matriz dos termos independentes*;

a matriz $\left[A \mid b\right] = \begin{bmatrix} a_{11} & \cdots & a_{1n} & \vdots & b_1 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & \vdots & b_m \end{bmatrix}$ é denominada *matriz ampliada do sistema*.

## Atividade 1

***Atende ao objetivo 1***

Identifique se os sistemas abaixo são lineares e, em caso afirmativo, escreva o sistema matricial correspondente.

1. $\begin{cases} 2x+4y &= 0 \\ x+y+z &= 2 \\ x-y &= 3 \\ x+z &= 1 \end{cases}$

2. $\begin{cases} x^2+y+z &= 2 \\ x-y &= 0 \\ x+z &= 1 \end{cases}$

3. $\begin{cases} x+y+z &= 20 \\ x-y &= 3 \\ x+z &= 10 \end{cases}$

### Resposta comentada

Os sistemas 1 e 3 são sistemas lineares, porém o sistema 2 tem uma variável quadrática, o que não permite que ele seja linear.

A forma matricial para os sistemas 1 e 3 são, respectivamente:

$$\begin{bmatrix} 2 & 4 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 3 \\ 1 \end{bmatrix} \text{ e } \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 20 \\ 3 \\ 10 \end{bmatrix}.$$

## Método da eliminação de Gauss

Diremos que uma matriz $A = (a_{ij})_{m \times n}$ é *escalonada* quando o primeiro elemento não-nulo de cada uma de suas linhas está a esquerda do primeiro elemento não-nulo de cada uma das linhas subsequentes e, além disso, as linhas nulas (se houver) estão abaixo das demais.

Na imagem a seguir, podemos ver um exemplo da matriz escalonada, destacando o primeiro elemento não-nulo de cada uma de suas linhas e o primeiro elemento não nulo de cada uma das linhas subsequentes.

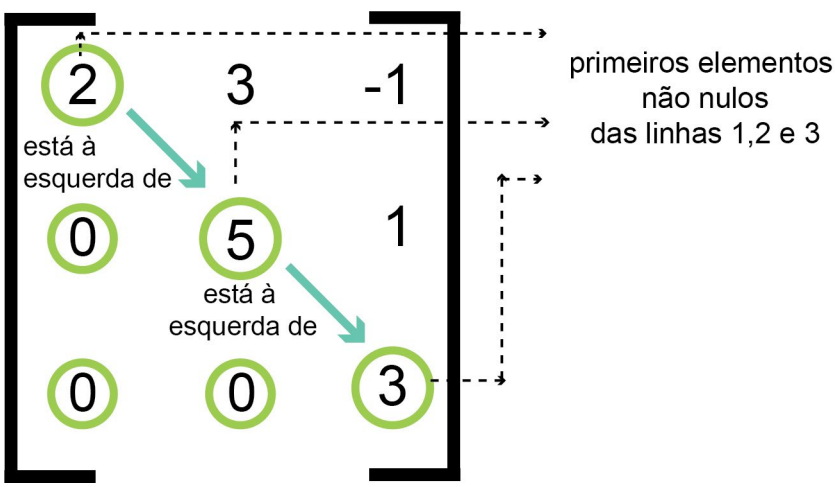



**Figura 4.1:** Matriz na forma escalonada.

Para aplicar o método da eliminação de Gauss, é importante dominar as seguintes operações elementares com matrizes:

1. trocar a posição de duas linhas;
2. somar a uma linha um múltiplo de outra linha;
3. multiplicar uma linha por um número diferente de zero.

Descreveremos a seguir o processo de *eliminação* (ou *escalonamento*), o qual, mediante aplicações sucessivas dessas três operações elementares às linhas de uma matriz, produz uma matriz escalonada.

### Procedimento:

a) Se $a_{11} \neq 0$, o processo começa deixando a primeira linha intacta e somando a cada linha $L_i$, com $i \geq 2$, a primeira linha multiplicada por $l_{i1} = \frac{-a_{i1}}{a_{11}}$. Com isto se obtém uma matriz cuja primeira coluna é $(a_{11}, 0, \ldots, 0)$.

b) Se $a_{11} = 0$, uma troca de linhas fornece uma matriz com $a_{11} \neq 0$, desde que a primeira coluna não seja nula.

c) Se, porém todos os elementos de primeira coluna são iguais a zero, passa-se para a segunda coluna ou, mais comumente, para a coluna mais próxima, à direita da primeira, onde haja algum elemento não-nulo e opera-se como antes, de modo a obter uma matriz cuja primeira coluna não-nula começa com um elemento diferente de zero, mas todos os demais são iguais a zero.

d) A partir daí, não se mexe mais na primeira linha. Recomeça-se o processo, trabalhando-se com as linhas a partir da segunda, até obter uma matriz escalonada.

### Exemplo 1

Para gerarmos uma matriz escalonada, antes de tudo, devemos ter em mente qual é o resultado final que queremos. Ora, uma matriz escalonada, como já dissemos antes, é uma matriz que possui uma forma de escada, em que, na primeira coluna, somente o primeiro elemento é diferente de zero; na segunda coluna, o segundo elemento é diferente de zero e todos abaixo dele são iguais a zero; na terceira coluna, o terceiro elemento é diferente de zero e todos abaixo dele são iguais a zero, e assim por diante.

Para conseguir isso, passamos então a dividir o problema em partes menores (*baby steps*) até conquistarmos o resultado esperado.

a) Objetivo 1: transformar o primeiro elemento da linha $L_2$ em zero.

- Estratégias:

  - Transformar a linha $L_2$ em $L_2 - 5L_1$

  - Transformar a linha $L_3$ em $L_3 - 9L_1$

$$\begin{bmatrix} 1 & 2 & 3 & \vdots & 4 \\ 5 & 6 & 7 & \vdots & 8 \\ 9 & 10 & 11 & \vdots & 12 \end{bmatrix} \begin{matrix} L_2 \to L_2 - 5L_1 \\ \text{------} \to \\ L_3 \to L_3 - 9L_1 \end{matrix} \begin{bmatrix} 1 & 2 & 3 & \vdots & 4 \\ 0 & -4 & -8 & \vdots & -12 \\ 0 & -8 & -16 & \vdots & -24 \end{bmatrix}$$

b) Objetivo 2: transformar o segundo elemento da linha 3 em zero.

- Estratégia:

  -Transformar a linha $L_3$ em $L_3 - 2L_2$.

$$\begin{matrix} \text{------} \to \\ L_3 \to L_3 - 2L_2 \end{matrix} \begin{bmatrix} 1 & 2 & 3 & \vdots & 4 \\ 0 & -4 & -8 & \vdots & -12 \\ 0 & 0 & 0 & \vdots & 0 \end{bmatrix}$$

Agora temos a nossa matriz escalonada, através do método da eliminação de Gauss, adotando *baby steps* para nos ajudarem a avançar na resolução do problema de forma mais segura.

## Exemplo 2

Considere o sistema:

$$\begin{cases} 2x + y + z &= 5 \\ 4x - 6y &= -2 \\ -2x + 7y + 2z &= 9 \end{cases}$$

Vamos aplicar a eliminação na matriz ampliada do sistema:.

Atenção: Vá acompanhando o passo a passo no seu caderno, como se você estivesse fazendo um exercício. Assim você treina o método!

$$\begin{bmatrix} 2 & 1 & 1 & \vdots & 5 \\ 4 & -6 & 0 & \vdots & -2 \\ -2 & 7 & 2 & \vdots & 9 \end{bmatrix} \begin{matrix} L_2 \to L_2 - \frac{4}{2}L_1 \\ \text{------} \to \\ L_3 \to L_3 - \left(-\frac{2}{2}\right)L_1 \end{matrix} \begin{bmatrix} 2 & 1 & 1 & \vdots & 5 \\ 0 & -8 & -2 & \vdots & -12 \\ 0 & 8 & 3 & \vdots & 14 \end{bmatrix}$$

$$\begin{matrix} \text{------} \to \\ L_3 \to L_3 - \left(-\frac{8}{8}\right)L_2 \end{matrix} \begin{bmatrix} 2 & 1 & 1 & \vdots & 5 \\ 0 & -8 & -2 & \vdots & -12 \\ 0 & 0 & 1 & \vdots & 2 \end{bmatrix}$$

Essa matriz está escalonada. Logo,

$$\begin{cases} 2x+y+z &= 5 \\ -8y-2z &= -12 \\ z &= 2 \end{cases}$$

substituindo $z = 2$ na segunda equação,

$$8y = 12 - 4 = 8 \Rightarrow y = 1$$

substituindo $y = 1$ e $z = 2$ na primeira equação

$$2x = 5 - 1 - 2 = 2 \Rightarrow x = 1$$

logo, a solução do sistema é $x = y = 1$ e $z = 2$.

Os elementos da diagonal principal de uma matriz escalonada são chamados de *pivôs*. Note que todos os pivôs da matriz acima são diferentes de zero e que esse sistema tem solução única.

**Carl Friedrich Gauss (1777-1855)**

Johann Carl Friedrich Gauss, conhecido popularmente como o "príncipe dos matemáticos", foi uma referência incontornável na matemática, na geometria, na física e na astronomia. Entre as suas maiores conquistas acadêmicas está a invenção do telégrafo, desenhou o heptadecágono, definiu o conceito de números complexos e criou a geometria diferencial

Fonte: https://www.ebiografia.com/carl_friedrich_gauss/

## Atividade 2

*Atende ao objetivo 2*

Resolva o sistema linear a seguir, usando o método da eliminação de Gauss.

$$\begin{cases} x+y+z &= 1 \\ 4x-6z &= -2 \\ 2x+y+2z &= 3 \end{cases}$$

### Resposta comentada

Vamos aplicar a eliminação na matriz ampliada do sistema.

$$\begin{bmatrix} 1 & 1 & 1 & \vdots & 1 \\ 4 & 0 & -6 & \vdots & -2 \\ 2 & 1 & 2 & \vdots & 3 \end{bmatrix} \begin{matrix} L_2 \to L_2 - 4L_1 \\ \text{------} \to \\ L_3 \to L_3 - 2L_1 \end{matrix} \begin{bmatrix} 1 & 1 & 1 & \vdots & 1 \\ 0 & -4 & -10 & \vdots & -6 \\ 0 & -1 & 0 & \vdots & 1 \end{bmatrix}$$

$$\begin{matrix} \text{------} \to \\ L_3 \to L_3 - \left(\frac{1}{4}\right) L_2 \end{matrix} \begin{bmatrix} 1 & 1 & 1 & \vdots & 1 \\ 0 & -4 & -10 & \vdots & -6 \\ 0 & 0 & \frac{10}{4} & \vdots & \frac{10}{4} \end{bmatrix}$$

Essa matriz está escalonada logo,

$$\begin{cases} x+y+z &= 1 \\ -4y-10z &= -6 \\ \frac{10}{4}z &= \frac{10}{4} \end{cases}$$

substituindo $z = 1$ na segunda equação

$4y = 6 - 10 = -4 \Rightarrow y = -1$

substituindo $y = -1$ e $z = 1$ na primeira equação

$x = 1 + 1 - 1 = 1 \Rightarrow x = 1$

Logo, a solução do sistema é $x = z = 1$ e $y = -1$.

---

Para entendermos melhor o que é um sistema não-singular e um sistema singular, vamos apresenta-los através de um exemplo.

### Sistema não-singular

É aquele que sempre possui uma única solução.

$$\begin{cases} x+y+z = a \\ 2x+2y+5z = b \\ 4x+6y+8z = c \end{cases} \Rightarrow \begin{cases} x+y+z = a \\ 3z = b-2a \\ 2y+4z = c-4a \end{cases} \Rightarrow \begin{cases} x+y+z = a \\ 2y+4z = c-4a \\ 3z = b-2a \end{cases}$$

Sendo os pivôs todos diferentes de zero, então, podemos assumir que logo esse sistema tem solução para qualquer valor de *a, b* e *c*.

### Sistema singular

É aquele que não possui solução ou possui infinitas soluções.

$$\begin{cases} x+y+z = a \\ 2x+2y+5z = b \\ 4x+4y+8z = c \end{cases} \Rightarrow \begin{cases} x+y+z = a \\ 3z = b-2a \\ 4z = c-4a \end{cases}$$

Esse sistema tem um pivô ($a_{21}$, *depois ter sido* aplicada a eliminação Gaussiana) igual a zero. Não podemos concluir nada sobre esse sistema, a princípio, pois $3z = b - 2a$ e $4z = c - 4a \Rightarrow z = \frac{b-2a}{3}$ e $z = \frac{c-4a}{4}$, ou seja, se $\frac{b-2a}{3} \neq \frac{c-4a}{4}$, o sistema não tem solução.

Porém, se $\frac{b-2a}{3} = \frac{c-4a}{4}$, ou seja, $4a + 4b - 3c = 0$ implica que o sistema tem infinitas soluções.

## Algoritmo numérico para o método da eliminação de Gauss

Para implementarmos o método da eliminação de Gauss em um programa numérico devemos primeiro escrever um algoritmo genérico, que possa ser passado para qualquer linguagem de programação. Para isso:

1. Considere que você queira resolver um sistema linear na forma $Ax = b$, onde a matriz $A$ tem ordem $n \times n$, e o vetor $b$ tem ordem $n \times 1$.

   Primeiro temos que triangularizar o sistema, ou seja, deixá-lo na forma triangular superior, que é o mesmo que escalonar, como já explicamos anteriormente, ou seja, nesta forma:

   Exemplo de matriz triangular superior:

$$\begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$$

2. Vamos supor que os elementos $a_{kk}$ (que formam a diagonal) de todas as linhas sejam diferentes de zero. Caso contrário, temos que trocar toda essa linha por uma cujos elemento $a_{kk}$ sejam diferentes de zero.

   Agora que já temos o nosso sistema linear escrito em formato de matriz ampliada e devidamente escalonado, podemos passar à sugestão de nosso *algoritmo numérico* para a resolução de sistemas lineares. Vamos juntos?

### Sugestão para o algoritmo 1:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} & \vdots & b_1 \\ a_{21} & \cdots & a_{2n} & \vdots & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n-1,\,1} & \cdots & a_{n-1,\,n} & \vdots & b_{n-1} \\ a_{n1} & \cdots & a_{nn} & \vdots & b_n \end{bmatrix}$$

- **Passo 1:** primeiro temos que garantir que vamos percorrer todas as colunas com um loop de $k = 1, \ldots, n - 1$;
- **Passo 2:** depois, em cada linha, temos que calcular o número $(m = \frac{a_{ik}}{a_{kk}})$ que auxilia a zerar os elementos abaixo da diagonal principal e zerá-los ($a_{ik} = 0$). Para isso, precisamos fazer um outro loop de $i = k + 1, \ldots, n$.
- **Passo 3:** devemos modificar os elementos acima da diagonal principal, calculando os seus novos valores ($a_{ij} = a_{ij} - ma_{kj}$ e $b_i = b_i - mb_k$). Para isso, precisamos fazer um outro loop de $j = k + 1, \ldots, n$.

  No final do Passo 3, teremos a matriz na forma triangular superior:

$$\left[\begin{array}{ccccc|c} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & b_1 \\ 0 & a_{22} & \cdots & a_{2n} & \vdots & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1,\,n} & \vdots & b_{n-1} \\ 0 & 0 & \cdots & a_{nn} & \vdots & b_n \end{array}\right]$$

## Atividade 3

*Atende ao objetivo 3*

Utilize o sistema linear $\begin{cases} x + y = 1 \\ x - y = 3 \end{cases}$ como exemplo para aplicar o algoritmo para o método da eliminação de Gauss e chegue no sistema triangular superior equivalente.

### *Resposta comentada*

- Passo 1: $k \to 1$
- Passo 2: $i \to 2$

$$m = \frac{a_{21}}{a_{11}} = \frac{1}{1} = 1$$

$$a_{21} = 0$$

- Passo 3: $j \to 2$

$$a_{22} = a_{22} - ma_{12} = -1 - 1 \times 1 = -2$$

$$b_2 = b_2 - mb_1 = 3 - 1 \times 1 = 2$$

Como estamos em um sistema de ordem 2, temos apenas um *loop*. E o sistema triangular superior equivalente é $\begin{cases} x + y = 1 \\ -2y = 2 \end{cases}$.

---

Note que ainda não resolvemos o sistema; simplesmente chegamos no sistema triangular superior equivalente.

Agora temos que substituir os valores de baixo para cima para acharmos a solução do sistema.

### Ideia para o algoritmo 2:

- **Passo 1:** $x_n = \frac{b_n}{a_{nn}}$
- **Passo 2:** depois, temos que substituir cada $x_n$, um a um, porém um de cada vez. Para isso, vamos precisar fazer um *loop* para as linhas de $k = n - 1, \ldots, 2, 1$. E vamos precisar de uma variável auxiliar $s$ que começará em zero.
- **Passo 3**: precisaremos de um *loop* nas colunas de $j = k + 1, \ldots, n$. Atualize a variável auxiliar $s = s + a_{kj}x_j$ e calcule $x_k = \frac{b_k - s}{a_{kk}}$.

## Atividade 4

*Atende ao objetivo 3*

Ache a solução do sistema triangular superior $\begin{cases} x+y &= 1 \\ -2y &= 2 \end{cases}$, utilizando o algoritmo 2.

### Resposta comentada

- Passo 1: $x_2 = \dfrac{b_2}{a_{22}} = \dfrac{2}{-2} = -1$
- Passo 2: $k \to 1$

  $s = 0$
- Passo 3: $j \to 2$

  $s = s + a_{12}x_2 = 0 + 1 \times (-1) = -1$

  $x_1 = \dfrac{b_1 - s}{a_{11}} = \dfrac{1-(-1)}{1} = 2.$

Logo, a solução do sistema é $x = 2$ e $y = -1$.

Agora vamos escrever os algoritmos:

### Algoritmo 1:

- Passo 1: para $k = 1, \ldots, n-1$
- Passo 2: para $i = k+1, \ldots, n$

  $m = \dfrac{a_{ik}}{a_{kk}}$

  $a_{ik} = 0$

- Passo 3: para $j = k + 1, \ldots, n$

$$a_{ij} = a_{ij} - ma_{kj}$$
$$b_i = b_i - mb_k$$

### Algoritmo 2:

Passo 1: $x_n = \dfrac{b_n}{a_{nn}}$

Passo 2: para $k = (n - 1), \ldots, 2, 1$

$$s = 0$$

Passo 3: para $j = k + 1, \ldots, n$

$$s = s + a_{kj}x_j$$
$$x_k = \frac{b_k - s}{a_{kk}}$$

## Atividade 5

*Atende ao objetivo 4*

Escreva o algoritmo 1 em uma linguagem de programação (MatLab, C++, Fortran, ...) com a qual você esteja familiarizado.

### *Resposta comentada*

Utilizaremos como exemplo a linguagem do MatLab para escrever o algoritmo 1.

```
function S=triangular(A)
% x = triangular(A)
%  A é a matriz ampliada do sistema que você quer resolver
% S é a matriz ampliada para o sistema triangular superior equivalente.
%  Esta função calcula a matriz ampliada para o sistema triangular
%  superior equivalente ao que você está tentando resolver.
%
n=size(A); % calcula o tamanho da matriz ampliada a variável n tem
%duas entradas n=[numero de linhas, numero de colunas]
for k=1:n(1)−1
for i=k+1:n(1)
      m=A(i,k)/A(k,k);
      A(i,k)=0;
for j=k+1:n(1)
        A(i,j)=A(i,j)-m*A(k,j);
        A(i,n(2))=A(i,n(2))-m*A(k,n(2));
end
end
end
S=A;
```

## Informações sobre a próxima aula

Na próxima aula, estudaremos um outro método direto para resolução de sistemas lineares. Até lá!

## Resumo

Nesta aula, você estudou que:

- métodos diretos fornecem a solução exata do problema;

- métodos iterativos fornecem uma solução numérica aproximada do problema;
- sistemas lineares são um conjunto de equações lineares;
- matriz escalonada é uma matriz onde todos os elementos acima ou abaixo da diagonal principal são nulos;
- o processo de eliminação (ou escalonamento), produz uma matriz escalonada.
- algoritmos numéricos podem realizar a eliminação de Gauss.

## Referências

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R., *Cálculo numérico*: aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

BURDEN, Richard L.; FAIRES, Douglas. *Análise numérica*. São Paulo: Pioneira Thomson Learning, 2003.

# Aula 5

## Resolução de sistemas lineares: métodos diretos – fatoração LU

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Apresentar o método da fatoração LU para matrizes.

## Objetivos

Esperamos que, ao final desta aula, você seja capaz de:

1. identificar uma matriz triangular superior e uma matriz triangular inferior;
2. utilizar o método da fatoração LU;
3. descrever cada passo do algoritmo numérico para o método da fatoração LU;
4. escrever um algoritmo para o método da fatoração LU.

## Pré-requisitos

Para um bom aproveitamento desta aula, é importante relembrar o Método da Eliminação de Gauss, ensinado na Aula 4.

## Introdução

Na aula anterior, você aprendeu o que é um sistema não-singular, cuja matriz é uma matriz não-singular. A fatoração LU consiste em representar uma matriz não-singular como o produto de duas matrizes, uma triangular superior e outra triangular inferior.

O nome da fatoração vem do inglês L, de *lower*, e U, de *upper*. Essa fatoração é importante para resolvemos sistemas lineares numericamente, de forma mais eficiente.

## Matrizes triangulares

Diremos que uma matriz $A = (a_{ij})_{n\times n}$ está na *forma triangular superior* ou é uma *matriz triangular superior*, se todos os elementos abaixo da diagonal principal forem iguais a zero, ou seja, se $a_{ij} = 0$ para todo $j < i$.

$$\begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$$

Diremos ainda que uma matriz $A = (a_{ij})_{n\times n}$ está na *forma triangular inferior* ou é uma *matriz triangular inferior*, se todos os elementos acima da diagonal principal for igual a zero, ou seja, se $a_{ij} = 0$ para todo $j > i$.

$$\begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}$$

### Atividade 1

***Atende ao objetivo 1***

Identifique quais das matrizes abaixo são triangulares superiores ou triangulares inferiores.

1. $\begin{bmatrix} 2 & 4 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

4. $\begin{bmatrix} 1 & 0 & 0 \\ 20 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}$

### *Resposta comentada*

A matriz 1 não é quadrada, logo, não pode ser triangular, já que não possui o mesmo número de linhas e colunas.

As matrizes 2 e 3 estão na forma triangular superior, uma vez que possuem somente zeros abaixo da sua diagonal principal.

A matriz 4 está na forma triangular inferior, pois possui somente zeros acima da sua diagonal principal.

## *Atividade 2*

*Atende ao objetivo 1*

Dê um exemplo de uma matriz triangular superior e de uma matriz triangular inferior.

***Resposta comentada***

Matriz triangular superior: $\begin{bmatrix} 1 & -2 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$; uma vez que possui somente zero, abaixo da sua diagonal principal.

Matriz triangular inferior: $\begin{bmatrix} 1 & 0 & 0 \\ 20 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}$; uma vez que possui somente zeros acima da sua diagonal principal.

## Método da fatoração LU

A fatoração LU consiste em escrever uma matriz através do produto de duas matrizes triangulares. Escreveremos uma matriz $M$ como o produto de uma matriz $L$ (matriz inferior), e uma matriz $U$ (triangular superior).

Encontraremos esse produto de matrizes triangulares usando o método da eliminação de Gauss, estudado na Aula 4.

Para você entender melhor o que vamos aprender, observe o exemplo a seguir:

**Exemplo 1:**

$$A = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = LU.$$

Você pode estar se perguntando, agora, como encontramos as matrizes L e U. Por meio de um exemplo, descreveremos como encontramos essas matrizes usando a eliminação Gaussiana.

Você se lembra que aprendeu a escalonar uma matriz, usando a eliminação Gaussiana, na Aula 4? O resultado é exatamente uma matriz triangular superior, a sua matriz *U*.

Observe o **Exemplo 2** desta aula. Vamos usá-lo para exemplificar como montamos a matriz *L*.

**Exemplo 2:**

O que vamos fazer aqui é basicamente escalonar a matriz $M = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{bmatrix}$, transformando-a na matriz U, escalonada.

$$\begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{bmatrix} \begin{matrix} L_2 \to L_2 - \boldsymbol{l_{21}} L_1 = L_2 - \boldsymbol{5} L_1 \\ \text{----------} \to \\ L_3 \to L_3 - \boldsymbol{l_{31}} L_1 = L_3 - \boldsymbol{9} L_1 \end{matrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & -8 & -16 \end{bmatrix}$$

$$\begin{matrix} \text{----------} \to \\ L_3 \to L_3 - \boldsymbol{l_{32}} L_2 = L_3 - \boldsymbol{2} L_2 \end{matrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & 0 & 0 \end{bmatrix}$$

$U = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & 0 & 0 \end{bmatrix}$ é a matriz final escalonada. Perceba que utilizamos o coeficiente $l_{21} = 5$ para fazer a transformação da linha $L_2$; $l_{31} = 9$ para fazer a primeira transformação da linha $L_3$; e $l_{32} = 2$ para fazer a segunda transformação da linha $L_3$.

$L = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 9 & 2 & 1 \end{bmatrix}$, então, é a matriz formada pelos coeficientes que usamos para zerarmos os elementos abaixo da diagonal principal. Sim! Exatamente os coeficientes $l_{21} = 5$, $l_{31} = 9$ e $l_{32} = 2$, que destacamos na eliminação gaussiana que resultou em *U*. Tome um tempo e torne a verificar o que fizemos, para fixar a aprendizagem. Pegue seu lápis, papel e borracha e vá em frente!

Já verificou? Agora, aproveite e faça a prova real da decomposição LU, por meio da multiplicação das matrizes L e U. O resultado esperado é a matriz M, naturalmente.

$$\begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 9 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & 0 & 0 \end{bmatrix} = LU$$

## Atividade 3

*Atende ao objetivo 2*

Encontre a fatoração LU para as matrizes a seguir:

1. $\begin{bmatrix} 2 & 4 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix}$
2. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$
3. $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$
4. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

### *Resposta comentada*

1.

$$\begin{bmatrix} 2 & 4 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix} \xrightarrow[L_3 \to L_3 - \mathbf{0}L_1]{L_2 \to L_2 - \mathbf{1}/\mathbf{2}\, L_1} \begin{bmatrix} 2 & 4 & 0 \\ 0 & -2 & 1 \\ 0 & \mathbf{1} & 3 \end{bmatrix}$$

$$\xrightarrow[L_3 \to L_3 - \left(-\mathbf{1}/\mathbf{2}\right)L_2]{} \begin{bmatrix} 2 & 4 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & 7/2 \end{bmatrix}$$

Logo,

$$\begin{bmatrix} 2 & 4 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 0 & -1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & 7/2 \end{bmatrix} = LU.$$

2. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = LU.$

3. $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} = LU.$

4.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \xrightarrow[L_3 \to L_3 - \mathbf{1}L_1]{L_2 \to L_2 - \mathbf{0}L_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[L_3 \to L_3 - \mathbf{0}L_2]{} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Logo,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = LU.$$

Uma das operações elementares usadas no processo de eliminação é a de trocar a posição de duas linhas.

Registramos essa operação quando estamos realizando a fatoração LU, trocando a mesma linha em uma matriz identidade; ou seja, guardamos uma matriz de permutação P. Dessa forma, nossa fatoração LU, quando necessitamos trocar uma linha de posição, fica da forma, $PA = LU$ ou $A = P^{-1}LU$; em que $P$ é uma matriz de permutação, $L$ é uma matriz triangular inferior e $U$ é a matriz triangular superior. Para entender melhor, vejamos o exemplo seguinte.

**Exemplo 3:**

Vamos calcular a fatoração LU para a matriz $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix}$, acompanhe as operações a seguir:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix} \xrightarrow{L_1 \leftrightarrow L_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix} \xrightarrow{L_2 \leftrightarrow L_3} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

Logo,

$$PA = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} = LU.$$

## Atividade 4

*Atende ao objetivo 2*

Encontre a fatoração LU para as matrizes a seguir:

1. $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$
2. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$

***Resposta comentada***

1. $$PA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = LU.$$

2.
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \xrightarrow{L_2 \leftrightarrow L_3} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[L_3 \to L_3 - 0L_1]{L_2 \to L_2 - 1L_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{L_3 \to L_3 - 1L_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$PA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = LU.$$

## Algoritmo numérico para a fatoração LU

Vamos relembrar o algoritmo para o método da eliminação de Gauss, feito na aula passada; porém, aqui, só consideraremos a matriz $A$, e não o sistema linear.

Vamos supor que o elemento $a_{kk}$ de todas as linhas sejam diferente de zero. Caso contrário, temos que trocar toda essa linha por uma cujo elemento $a_{kk}$ seja diferente de zero; teremos que guardar uma matriz de permutação que corresponda à troca dessas linhas. Caso a matriz tenha linhas de zeros, estas devem ser colocada nas linhas de baixo da matriz.

**Ideia para o algoritmo:**

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n-1,\,1} & \cdots & a_{n-1,\,n} \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

**Passo (1)**: primeiramente, temos que garantir que vamos percorrer todas as colunas com um loop de $k = 1, \ldots, n - 1$.

**Passo (2)**: depois, em cada linha, temos que calcular o número ( $l_{ik} = \frac{a_{ik}}{a_{kk}}$ ), que auxilia a zerar os elementos abaixo da diagonal principal, e zerá-los ($a_{ik} = 0$). Para isso precisamos fazer um outro loop de $i = k + 1, \ldots, n$.

**Passo (3)**: devemos, então, modificar os elementos acima da diagonal principal, calculando os seus novos valores ($a_{ij} = a_{ij} - l_{ik}a_{kj}$). Para isso, precisamos fazer um outro loop de $j = k + 1, \ldots, n$.

No final do passo (3), teremos a matriz U na forma triangular superior e teremos guardado a matriz dos elementos da matriz triangular inferior L.

$$U = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1,\,n} \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \text{ e } L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ l_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ l_{n-1,\,1} & l_{n-1,\,2} & \cdots & 0 \\ l_{n1} & l_{n2} & \cdots & 1 \end{bmatrix}$$

## Atividade 5

*Atende ao objetivo 3*

Utilize a matriz $\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ como exemplo para aplicar o algoritmo para a fatoração LU e encontrar a sua decomposição.

### *Resposta comentada*

Passo (1): $k \rightarrow 1$

$$a_{11} = 2 \neq 0$$

Passo (2): $i \rightarrow 2$

$$l_{21} = \frac{a_{21}}{a_{11}} = \frac{0}{2} = 0$$

$$a_{21} = 0$$

Passo (3): $j \rightarrow 2$

$$a_{22} = a_{22} - l_{21}a_{12} = 1 - 0 \times 1 = 1$$

$$j \rightarrow 3$$

$$a_{23} = a_{23} - l_{21}a_{13} = 1 - 0 \times 1 = 1$$

Passo (2): $i \rightarrow 3$

$$l_{31} = \frac{a_{31}}{a_{11}} = \frac{1}{2}$$

$$a_{31} = 0$$

Passo (3): $j \rightarrow 2$

$$a_{32} = a_{32} - l_{31}a_{13} = 1 - \frac{1}{2} \times 1 = \frac{1}{2}$$

$$j \rightarrow 3$$

$$a_{33} = a_{33} - l_{31}a_{13} = 1 - \frac{1}{2} \times 1 = \frac{1}{2}$$

Passo (1): $k \rightarrow 2$

$$a_{22} = 1 \neq 0$$

Passo (2): $i \rightarrow 3$

$$l_{32} = \frac{a_{32}}{a_{22}} = \frac{\frac{1}{2}}{1} = \frac{1}{2}$$

$$a_{32} = 0$$

Passo (3): $j \rightarrow 3$

$$a_{33} = a_{33} - l_{32}a_{23} = \frac{1}{2} - \frac{1}{2} * 1 = 0$$

Como estamos em uma matriz de ordem 3, temos dois (2) *loops* para $k$. O resultado é:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} = LU.$$

---

Note que a matriz de permutação do sistema anterior é a identidade, pois não precisamos efetuar trocas de linhas.

Agora, vamos escrever o algoritmo:

**Algoritmo:**

Passo (1): Para $k = 1, \ldots, n - 1$.

Passo (2): Para $i = k + 1, \ldots,$ n

$$l_{ik} = \frac{a_{ik}}{a_{kk}}$$

$$a_{ik} = 0$$

Passo (3): Para $j = k + 1, \ldots, n$

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

$$U = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1,n} \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \text{ e } L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ l_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ l_{n-1,1} & l_{n-1,2} & \cdots & 0 \\ l_{n1} & l_{n2} & \cdots & 1 \end{bmatrix}$$

## Atividade 6

*Atende ao objetivo 4*

Escreva o algoritmo em uma linguagem de programação (MatLab, C++, Fortran,...) com a qual você esteja familiarizado.

### Resposta comentada

Utilizaremos como exemplo a linguagem do MatLab para escrever o algoritmo.

```
function [L,U]=lu(A)
% [L,U]=lu(A)
% A é a matriz quadrada que será fatorada
% L é a matriz triangular inferior
% U é a matriz triangular superior
% Esta função calcula a fatoração LU de uma matriz quadrada A
%
n=size(A); % calcula o tamanho da matriz A
L=eye(n); % começa com L sendo uma matriz identidade de ordem igual a matriz A
for k=1:n(1)-1
for i=k+1:n(1)
        L(i,k)=A(i,k)/A(k,k);
        A(i,k)=0;
for j=k+1:n(1)
        A(i,j)=A(i,j)-L(i,k)*A(k,j);
```

```
end
end
end
U=A;
```

## Solução para sistema linear usando a fatoração LU

Considere que você já conheça a fatoração LU da matriz $A = (a_{ij})_{n\times n}$ e queira resolver o sistema linear da forma $Ax = b$, usando a fatoração LU conhecida.

Pela fatoração LU conhecida, $PA = LU$, só conseguiremos resolver o sistema, se $U$ for uma matriz invertível, ou seja: uma matriz não-singular. Sabemos, pela construção da matriz $L$, que ela é uma matriz não-singular: sua diagonal é formada por uns (1); sendo assim, ela é invertível.

Note que, como as matrizes $L$ e $U$ são triangulares, calcular sua inversa não dará muito trabalho. Vamos usar o método de Gauss-Jordan para fazer isso.

Suponha, por agora, que você já tenha calculado as inversas de $L$ e $U$

Observe a conta a seguir:

$$Ax = b \Rightarrow PAx = Pb \Rightarrow LUx = Pb \Rightarrow L^{-1}LUx = L^{-1}Pb \Rightarrow Ux = L^{-1}Pb$$

$$\Rightarrow U^{-1}Ux = U^{-1}L^{-1}Pb \Rightarrow x = U^{-1}L^{-1}Pb$$

## Método de Gauss-Jordan

Vamos usar um exemplo para que você entenda o método de Gauss-Jordan em uma matriz triangular de ordem 3.

**Exemplo 4:**

Você vai começar escrevendo uma matriz identidade ao lado da matriz que você quer inverter.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 5 & 1 & 0 & 0 & 1 & 0 \\ 9 & 10 & 1 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} L_2 \to L_2 - \mathbf{5}L_1 \\ \text{------} \to \\ L_3 \to L_3 - \mathbf{9}L_1 \end{array} \left[\begin{array}{ccc|ccc} 1 & \mathbf{0} & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -5 & 1 & 0 \\ 0 & 10 & 1 & -9 & 0 & 1 \end{array}\right]$$

$$\begin{array}{l} \text{-------} \to \\ L_3 \to L_3 - \mathbf{10}L_2 \end{array} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & \mathbf{0} \\ 0 & 1 & 0 & -5 & 1 & 0 \\ 0 & 0 & 1 & 41 & -10 & 1 \end{array}\right]$$

$$A^{-1} = \left[\begin{array}{ccc} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 41 & -10 & 1 \end{array}\right] \text{ matriz inversa.}$$

$$AA^{-1} = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 9 & 10 & 1 \end{array}\right]\left[\begin{array}{ccc} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 41 & -10 & 1 \end{array}\right] = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

## Conclusão

Nesta aula, aprendemos como fatorar uma matriz quadrada na forma LU. Aprendemos também a programar a fatoração LU, ou seja, como ensinar o computador a executar a fatoração LU. Aprendemos, ainda, como usar a fatoração LU para resolver um sistema linear. Com o método de Gauss-Jordan, apreendemos, por último, como inverter uma matriz usando as técnicas da fatoração LU.

## Resumo

Nesta aula você estudou que :

- uma matriz $A = (a_{ij})_{n\times n}$ é uma *matriz triangular superior* se todos os elementos abaixo da diagonal principal for igual a zero, ou seja, se $a_{ij} = 0$ para todo $j < i$;
- uma matriz $A = (a_{ij})_{n\times n}$ é uma *matriz triangular inferior* se todos os elementos acima da diagonal principal for igual a zero, ou seja, se $a_{ij} = 0$ para todo $j > i$;
- a fatoração LU consiste em escrever uma matriz como o produto de uma matriz triangular inferior $L$ e uma matriz triangular superior $U$;
- pode usar um exemplo de algoritmo numérico, para realizar a fatoração LU;

- para resolver um sistema linear usando a fatoração LU, você vai escrever $x = U^{-1}L^{-1}Pb$.

## Informações sobre a próxima aula

Na próxima aula, estudaremos os métodos iterativos para a resolução de sistemas lineares. Até lá!

## Referências

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R. *Cálculo numérico:* aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

BURDEN, Richard L.; FAIRES, Douglas. *Análise numérica*. São Paulo: Pioneira Thomson Learning, 2003.

# Aula 6

## Resolução de sistemas lineares: métodos iterativos – Gauss-Jacobi e Gauss-Seidel

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Apresentar os métodos iterativos para resolução de sistemas lineares.

## Objetivos

Esperamos que, ao final desta aula, você seja capaz de:

1. aplicar o método de Gauss-Jacobi na resolução de sistemas lineares;
2. aplicar o método de Gauss-Seidel na resolução de sistemas lineares.

## Pré-requisitos

Para um bom aproveitamento desta aula, é importante que você relembre os conceitos de sistemas lineares apresentados na Aula 4.

## Introdução

Preste atenção no sistema linear a seguir:

$$\begin{cases} 2x_1 + x_2 = 1 \\ 2x_2 - x_3 = 1 \\ x_2 + 2x_3 = 1 \end{cases} \qquad (1)$$

Vamos fazer uma brincadeira e isolar a variável $x_1$ na primeira equação, a variável $x_2$ na segunda equação e a variável $x_3$ na terceira equação.

$$\begin{cases} x_1 = (1 - x_2)/2 \\ x_2 = (1 + x_3)/2 \\ x_3 = (1 - x_2)/2 \end{cases} \qquad (2)$$

Agora, ao invés de resolvermos o sistema linear, vamos "chutar" que $x_1 = 0$, $x_2 = 0{,}6$ e $x_3 = 0{,}2$. Quando substituirmos esses valores "chutados" no lado direito do sistema representado em (2), teremos que:

$$\begin{cases} x_1 = (1 - 0{,}6)/2 = 0{,}2 \\ x_2 = (1 + 0{,}2)/2 = 0{,}6 \\ x_3 = (1 - 0{,}6)/2 = 0{,}2 \end{cases} \qquad (3)$$

Conseguimos, assim, novos valores para $x_1$, $x_2$ e $x_3$.

E vamos continuar a brincadeira, substituindo no campo, outra vez, esses novos valores no sistema!

$$\begin{cases} x_1 = (1 - 0{,}6)/2 = 0{,}2 \\ x_2 = (1 + 0{,}2)/2 = 0{,}6 \\ x_3 = (1 - 0{,}6)/2 = 0{,}2 \end{cases}$$

Note que, já na última substituição, repetimos o valor encontrado na anterior. Será que isso aconteceu ao acaso, ou será que existe uma forma de garantir que sempre poderemos fazer essas iterações e isso nos dará solução do sistema?

Note que iteramos o sistema apenas duas vezes.

Na primeira iteração, "chutamos" valores para $x_1$, $x_2$ e $x_3$ e, na segunda iteração, substituímos os valores encontrados no lado direito da equação. Como consequência, descobrimos os valores que satisfazem o sistema.

Agora, vamos explicar de forma mais estruturada essa ideia, ensinando o método iterativo para sistemas lineares.

## Método iterativo para sistemas lineares

Vimos na Aula 4 que *métodos iterativos* são aqueles que procuram a solução do sistema considerando uma aproximação inicial e gerando uma sequência iterativa $\{X_N\}$. Mas atenção: essa sequência pode ou não convergir para a solução do sistema. Vamos ver como funciona?

Seja um sistema linear $AX = B$, onde $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$ é a matriz dos coeficientes, $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ é a matriz das variáveis e $b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ é a matriz dos termos independentes.

Para iniciarmos, devemos escrever uma lei de recorrência, para que cada passo do processo (iteração) dependa do passo anterior. Então, na forma matricial do sistema, teremos:

$$X^{(K+1)} = \overline{A}X^{(K)} + \overline{b} = \varphi(X^{(K)}).$$

Aqui, fazemos uma paradinha rápida, para explicar o que é esse K: ele é referente ao momento em que você se encontra na iteração. Ou seja, começamos com $K = 0$, o nosso "chute" inicial é $X^{(0)}$. Assim, usando "o chute", podemos calcular a próxima iteração $X^{(0+1)} = \overline{A}X^{(0)} + \overline{b} = \varphi(X^{(0)})$.

Sim. À primeira vista parece incompreensível. Precisaremos de um passo a passo metódico para demonstrar como se chega à equação a seguir.

Para isso, vamos criar um sistema linear cujos termos do lado direito sejam iguais a 1. Vale ressaltar que esse sistema é o mesmo que vimos na Introdução desta aula.

$$\text{Sistema linear} \rightarrow \begin{cases} 2x_1 + x_2 = 1 \\ 2x_2 - x_3 = 1 \\ x_2 + 2x_3 = 1 \end{cases} \rightarrow \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Em seguida, vamos construir a função $\varphi$ isolando $x_1$, $x_2$ e $x_3$ do lado esquerdo das equações que compõem o nosso sistema:

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

Agora, então, podemos escrever o sistema em forma de uma equação composta por matrizes, que iremos chamar, finalmente, de função $\varphi$.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^{(K+1)} = \underbrace{\begin{bmatrix} 0 & -1/2 & 0 \\ 0 & 0 & 1/2 \\ 0 & -1/2 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^{(K)} + \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}}$$

$$\bar{A}X^{(K)} + \bar{b} = \boldsymbol{\varphi}\left(\boldsymbol{X}^{(\boldsymbol{K})}\right)$$

### 1ª iteração

A partir da compreensão da função $\varphi(X^{(K)})$, vamos fazer a primeira iteração, considerando o sistema linear (1) e $X^{(0)} = \begin{bmatrix} 0 \\ 0{,}6 \\ 0{,}2 \end{bmatrix}$ "nosso chute inicial".

Vamos começar com o nosso sistema linear:

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

A partir dele, isolamos as variáveis $x_1$, $x_2$ e $x_3$ em cada uma das equações, tal e qual na equação (2). Assim:

$$\begin{cases} x_1 &= \left(1-x_2\right)/2 \\ x_2 &= \left(1+x_3\right)/2 \\ x_3 &= \left(1-x_2\right)/2 \end{cases}$$

Então, substituímos os valores de $X^{(0)} = \begin{bmatrix} 0 \\ 0{,}6 \\ 0{,}2 \end{bmatrix}$ nas variáveis $x_1$, $x_2$ e $x_3$ que estão do lado direito das equações do sistema. Assim, temos, tal e qual na equação (3):

$$\begin{cases} x_1 &= (1-0,6)/2 = 0,2 \\ x_2 &= (1+0,2)/2 = 0,6 \\ x_3 &= (1-0,6)/2 = 0,6 \end{cases} \qquad (4)$$

Dessa forma, podemos escrever a função $\varphi(X^{(K)})$ para a primeira iteração:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^{(1)} = \begin{bmatrix} 0 & -1/2 & 0 \\ 0 & 0 & 1/2 \\ 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^{(0)} + \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -1/2 & 0 \\ 0 & 0 & 1/2 \\ 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0,6 \\ 0,2 \end{bmatrix} + \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 0,2 \\ 0,6 \\ 0,2 \end{bmatrix}$$

**2ª iteração**

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^{(2)} = \begin{bmatrix} 0 & -1/2 & 0 \\ 0 & 0 & 1/2 \\ 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^{(1)} + \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -1/2 & 0 \\ 0 & 0 & 1/2 \\ 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} 0,2 \\ 0,6 \\ 0,2 \end{bmatrix} + \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 0,2 \\ 0,6 \\ 0,2 \end{bmatrix}$$

Como $X^{(1)} = X^{(2)}$, se continuarmos fazendo iterações, **não** sairemos desse ponto. Observe que essa é a solução do sistema.

Isso nos leva à seguinte pergunta: quando devemos parar de iterar? Quais são os critérios de parada dos métodos iterativos? É isso o que estudaremos a seguir.

## Critério de parada

Para pararmos um método iterativo para sistemas lineares, repetimos as iterações até que o vetor $X^{(K)}$ esteja suficientemente próximo do vetor $X^{(K-1)}$. A forma que vamos usar aqui, para medir o que é suficientemente próximo, será calcular a distância entre $X^{(K-1)}$ e $X^{(K)}$ usando a fórmula:

$$d^{(K)} = \max_{1 \leq i \leq n} \left| x_i^{(K)} - x_i^{(K-1)} \right|.$$

Como você já viu no curso de álgebra linear, essa fórmula é o meio de sabermos se dois vetores estão próximos. O que ela significa em palavras: a nossa distância $d^{(K)}$ representa a maior distância entre cada coordenada dos dois vetores. Você calcula a diferença de cada coordenada e depois pega a maior dessas diferenças.

Definida a fórmula que usaremos para distância e dada uma precisão ε, escolhemos o vetor $X^{(K)}$ para terminarmos o processo e achar a solução $\overline{X}$ aproximada do sistema, quando $d^{(K)} < \varepsilon$.

## Método iterativo de Gauss-Jacobi

Considere um sistema linear de ordem $n$,

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad & \vdots \quad \vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= b_n \end{cases}, \text{ suponha que } a_{ii} \neq 0 \text{ para } i = 1, \ldots, n.$$

Isole $x_i$ na linha $i$. Dessa forma teremos:

$$\begin{cases} x_1 &= \dfrac{b_1 - a_{12}x_2 - a_{13}x_3 - \cdots - a_{1n}x_n}{a_{11}} \\ x_2 &= \dfrac{b_2 - a_{21}x_1 - a_{23}x_3 - \cdots - a_{2n}x_n}{a_{22}} \\ \vdots \quad & \vdots \qquad\qquad \vdots \\ x_n &= \dfrac{b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{n,n-1}x_{n-1}}{a_{nn}}. \end{cases}$$

Na forma matricial, teremos:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 & \frac{-a_{12}}{a_{11}} & \frac{-a_{13}}{a_{11}} & \cdots & \frac{-a_{1n}}{a_{11}} \\ \frac{-a_{22}}{a_{22}} & 0 & \frac{-a_{23}}{a_{22}} & \cdots & \frac{-a_{2n}}{a_{22}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{-a_{n1}}{a_{nn}} & \frac{-a_{n2}}{a_{nn}} & \frac{-a_{n3}}{a_{nn}} & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \frac{b_1}{a_{11}} \\ \frac{b_2}{a_{22}} \\ \vdots \\ \frac{b_n}{a_{nn}} \end{bmatrix}.$$

Começamos com o chute inicial $X^{(0)}$. Chamamos esse chute de *aproximação inicial*. Usamos essa aproximação inicial para calcular a primeira iteração $X^{(1)}$ e assim, sucessivamente. Usamos a iteração $X^{(k)}$ para calcular a iteração $X^{(k+1)}$. Esse é o método de Gauss-Jacobi.

## Atividade 1

***Atende ao objetivo 1***

Resolva o sistema linear $\begin{cases} 2x_1 + x_2 = 1 \\ 2x_2 - x_3 = 1 \\ x_2 + 2x_3 = 1 \end{cases}$ usando o método de Gauss-Jacobi, considerando $X^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ e precisão $\varepsilon = 0{,}5$.

### Resposta comentada

Usando o método de Gauss-Jacobi, temos que:

$$1^a \text{ iteração} \rightarrow \begin{cases} x_1^{(1)} = \dfrac{1 - x_2^{(0)}}{2} = \dfrac{1-1}{2} = 0 \\ x_2^{(1)} = \dfrac{1 + x_3^{(0)}}{2} = \dfrac{1+1}{2} = 1 \\ x_3^{(1)} = \dfrac{1 - x_2^{(0)}}{2} = \dfrac{1-1}{2} = 0 \end{cases}$$

$$d^{(1)} = max\left\{|1-0|, |1-1|, |1-0|\right\} = 1 > 0{,}5$$

$$2^a \text{ iteração} \rightarrow \begin{cases} x_1^{(2)} = \dfrac{1-x_2^{(1)}}{2} = \dfrac{1-1}{2} = 0 \\ x_2^{(2)} = \dfrac{1+x_3^{(1)}}{2} = \dfrac{1+0}{2} = \dfrac{1}{2} \\ x_3^{(2)} = \dfrac{1-x_2^{(1)}}{2} = \dfrac{1-1}{2} = 0 \end{cases}$$

$$d^{(2)} = max\left\{\left|0-0\right|, \left|1-\frac{1}{2}\right|, \left|0-0\right|\right\} = 0,5 = \varepsilon$$

$$3^a \text{ iteração} \rightarrow \begin{cases} x_1^{(3)} = \dfrac{1-x_2^{(2)}}{2} = \dfrac{1-1/2}{2} = \dfrac{1}{4} \\ x_2^{(3)} = \dfrac{1+x_3^{(2)}}{2} = \dfrac{1+0}{2} = \dfrac{1}{2} \\ x_3^{(3)} = \dfrac{1-x_2^{(2)}}{2} = \dfrac{1-1/2}{2} = \dfrac{1}{4} \end{cases}$$

$$d^{(3)} = max\left\{\left|0-\frac{1}{4}\right|, \left|\frac{1}{2}-\frac{1}{2}\right|, \left|0-\frac{1}{4}\right|\right\} = 0,25 < 0,5$$

Pelo método de Gauss-Jacobi : $\overline{X} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ \frac{1}{4} \end{bmatrix}$.

## Critério de convergência

Uma pergunta natural é: para quais condições o método de Gauss-Jacobi converge? Essa pergunta será respondida pelo teorema a seguir.

### Teorema (Critério das Linhas)

Considere o sistema linear $Ax = b$ e

$$\alpha_k = \left( \sum_{\substack{j=1 \\ j \neq k}}^{n} \left|a_{kj}\right| \right) / \left|a_{kk}\right| = \frac{|a_{k1}| + |a_{k2}| + \ldots + |a_{kk-1}| + |a_{kk+1}| + \ldots + |a_{kn}|}{\left|a_{kk}\right|}.$$

Se $\alpha = \max_{1 \le k \le n} \alpha_k < 1$, então a sequência $\{X^{(k)}\}$ gerada pelo método de Gauss-Jacobi converge para a solução do sistema, para qualquer aproximação inicial $X^{(0)}$.

Para os mais curiosos, a demonstração desse teorema pode ser encontrada no capítulo 9 da última referência da aula.

## Atividade 2

*Atende ao objetivo 1*

Quais das matrizes abaixo satisfaz o Critério das Linhas?

1. $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 11 & 3 & 1 \\ -1 & 12 & 3 \\ 1 & 0 & 2 \end{bmatrix}$

### Resposta comentada

1. $\alpha_1 = \dfrac{|a_{12}|}{|a_{11}|} = \dfrac{1}{2}$ e $\alpha_2 = \dfrac{|a_{21}|}{|a_{22}|} = \dfrac{0}{1} = 0$, assim $\alpha = max\left\{\dfrac{1}{2}, 0\right\} = 0,5 < 1$.

A matriz $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$ satisfaz o Critério das Linhas.

2. $\alpha_1 = \dfrac{|a_{12}|+|a_{13}|}{|a_{11}|} = \dfrac{2+4}{1} = 6$, $\alpha_2 = \dfrac{|a_{21}|+|a_{23}|}{|a_{22}|} = \dfrac{0+1}{2} = 0{,}5$ e

$\alpha_3 = \dfrac{|a_{31}|+|a_{32}|}{|a_{33}|} = \dfrac{0+0}{1} = 0$, assim $\alpha = max\left\{6, \dfrac{1}{2}, 0\right\} = 6 > 1$.

A matriz $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ não satisfaz o Critério das Linhas.

3. $\alpha_1 = \dfrac{|a_{12}|+|a_{13}|}{|a_{11}|} = \dfrac{3+1}{11} = \dfrac{4}{11}$, $\alpha_2 = \dfrac{|a_{21}|+|a_{23}|}{|a_{22}|} = \dfrac{1+3}{12} = \dfrac{1}{3}$ e

$\alpha_3 = \dfrac{|a_{31}|+|a_{32}|}{|a_{33}|} = \dfrac{1+0}{2} = \dfrac{1}{2}$, assim $\alpha = max\left\{\dfrac{4}{11}, \dfrac{1}{3,}, \dfrac{1}{2}, \right\} = \dfrac{1}{2} < 1$.

A matriz $\begin{bmatrix} 11 & 3 & 1 \\ -1 & 12 & 3 \\ 1 & 0 & 2 \end{bmatrix}$ satisfaz o Critério das Linhas.

## Atividade 3

***Atende ao objetivo 1***

Resolva o sistema linear $\begin{cases} x_1 + x_2 = 2 \\ x_1 - 2x_2 = -2 \end{cases}$ usando o método de Gauss-Jacobi, com $X^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ e precisão $\varepsilon = 2^{-4}$. Porém, observe que o Critério das Linhas não é satisfeito.

### *Resposta comentada*

Considerando, no método de Gauss-Jacobi, a precisão indicada, lembre-se de que $\varepsilon = 2^{-4} = 1/16 = 0{,}0625$. Assim, temos que:

$$1^{a}\text{ iteração } \rightarrow \begin{cases} x_1^{(1)} = 2 - x_2^{(0)} = 2 - 1 = 1 \\ x_2^{(1)} = \dfrac{2 + x_1^{(0)}}{2} = \dfrac{2+1}{2} = \dfrac{3}{2}, \end{cases}$$

$$d^{(1)} = max\left\{\left|1-1\right|, \left|1 - \frac{3}{2}\right|\right\} = \frac{1}{2} > 0{,}0625$$

$$2^{a}\text{ iteração } \rightarrow \begin{cases} x_1^{(2)} = 2 - x_2^{(1)} = 2 - \dfrac{3}{2} = \dfrac{1}{2} \\ x_2^{(2)} = \dfrac{2 + x_1^{(1)}}{2} = \dfrac{2+1}{2} = \dfrac{3}{2} \end{cases}$$

$$d^{(2)} = max\left\{\left|1 - \frac{1}{2}\right|, \left|\frac{3}{2} - \frac{3}{2}\right|\right\} = \frac{1}{2} > 0{,}0625$$

$$3^{a}\text{ iteração } \rightarrow \begin{cases} x_1^{(3)} = 2 - x_2^{(2)} = 2 - \dfrac{3}{2} = \dfrac{1}{2} \\ x_2^{(3)} = \dfrac{2 + x_1^{(2)}}{2} = \dfrac{2 + 1/2}{2} = \dfrac{5}{4} \end{cases}$$

$$d^{(3)} = max\left\{\left|\frac{1}{2} - \frac{1}{2}\right|, \left|\frac{3}{2} - \frac{5}{4}\right|\right\} = \frac{1}{4} > 0{,}0625$$

$$4^{a}\text{ iteração } \rightarrow \begin{cases} x_1^{(4)} = 2 - x_2^{(3)} = 2 - \dfrac{5}{4} = \dfrac{3}{4} \\ x_2^{(4)} = \dfrac{2 + x_1^{(3)}}{2} = \dfrac{2 + 1/2}{2} = \dfrac{5}{4} \end{cases}$$

$$d^{(4)} = max\left\{\left|\frac{1}{2} - \frac{3}{4}\right|, \left|\frac{5}{4} - \frac{5}{4}\right|\right\} = \frac{1}{4} > 0{,}0625$$

$$5^{a}\text{ iteração } \rightarrow \begin{cases} x_1^{(5)} = 2 - x_2^{(4)} = 2 - \dfrac{5}{4} = \dfrac{3}{4} \\ x_2^{(5)} = \dfrac{2 + x_1^{(4)}}{2} = \dfrac{2 + 3/4}{2} = \dfrac{11}{8} \end{cases}$$

$$d^{(5)} = max\left\{\left|\frac{3}{4} - \frac{3}{4}\right|, \left|\frac{11}{8} - \frac{5}{4}\right|\right\} = \frac{1}{8} > 0{,}0625$$

$$6^{a}\text{ iteração} \rightarrow \begin{cases} x_1^{(6)} = 2 - x_2^{(5)} = 2 - \frac{11}{8} = \frac{5}{8} \\ x_2^{(6)} = \frac{2 + x_1^{(5)}}{2} = \frac{2 + 3/4}{2} = \frac{11}{8} \end{cases}$$

$$d^{(6)} = max\left\{\left|\frac{3}{4} - \frac{5}{8}\right|, \left|\frac{11}{8} - \frac{11}{8}\right|\right\} = \frac{1}{8} > 0{,}0625$$

$$7^{a}\text{ iteração} \rightarrow \begin{cases} x_1^{(7)} = 2 - x_2^{(6)} = 2 - \frac{11}{8} = \frac{5}{8} \\ x_2^{(7)} = \frac{2 + x_1^{(6)}}{2} = \frac{2 + 5/8}{2} = \frac{21}{16} \end{cases}$$

$$d^{(7)} = max\left\{\left|\frac{5}{8} - \frac{5}{8}\right|, \left|\frac{11}{8} - \frac{21}{16}\right|\right\} = \frac{1}{16} = 0{,}0625 = \varepsilon$$

$$8^{a}\text{ iteração} \rightarrow \begin{cases} x_1^{(8)} = 2 - x_2^{(7)} = 2 - \frac{21}{16} = \frac{11}{16} \\ x_2^{(8)} = \frac{2 + x_1^{(7)}}{2} = \frac{2 + 5/8}{2} = \frac{21}{16} \end{cases}$$

$$d^{(8)} = max\left\{\left|\frac{5}{8} - \frac{11}{16}\right|, \left|\frac{21}{16} - \frac{21}{16}\right|\right\} = \frac{1}{16} = 0{,}0625 = \varepsilon$$

$$9^{a}\text{ iteração} \rightarrow \begin{cases} x_1^{(9)} = 2 - x_2^{(8)} = 2 - \frac{21}{16} = \frac{11}{16} \\ x_2^{(9)} = \frac{2 + x_1^{(8)}}{2} = \frac{2 + 11/16}{2} = \frac{43}{32} \end{cases}$$

$$d^{(9)} = max\left\{\left|\frac{11}{16} - \frac{11}{16}\right|, \left|\frac{21}{16} - \frac{43}{32}\right|\right\} = \frac{1}{32} = \frac{1}{2^5} < \frac{1}{2^4} = \varepsilon$$

Pelo método de Gauss-Jacobi: $\overline{X} = \begin{bmatrix} \frac{11}{16} \\ \frac{43}{32} \end{bmatrix}$.

Porém, temos que $\alpha_1 = \frac{|a_{12}|}{|a_{11}|} = \frac{1}{1} = 1$ e $\alpha_2 = \frac{|a_{21}|}{|a_{22}|} = \frac{1}{2}$, assim $\alpha = max\left\{1, \frac{1}{2}\right\} = 1$. A matriz $\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$ não satisfaz o Critério das Linhas.

Observe que a condição do Critério das Linhas é apenas suficiente.

## Atividade 4

*Atende ao objetivo 1*

Qual mudança você pode fazer no sistema linear $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 1 \\ 2 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, sem alterar a sua solução, para que a matriz satisfaça o Critério das Linhas?

### Resposta comentada

Note $\alpha_1 = \frac{|a_{12}|+|a_{13}|}{|a_{11}|} = \frac{2+4}{1} = 6$, $\alpha_2 = \frac{|a_{21}|+|a_{23}|}{|a_{22}|} = \frac{0+1}{2} = 0{,}5$ e $\alpha_3 = \frac{|a_{31}|+|a_{32}|}{|a_{33}|} = \frac{2+0}{1} = 2$, assim $\alpha = max\left\{6, \frac{1}{2}, 2\right\} = 6 > 1$ ou seja, a matriz $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 2 & 4 \end{bmatrix}$ não satisfaz o Critério das Linhas.

Porém, o sistema linear é $\begin{cases} x + 2y + 4z = 1 \\ 2y + z = 1 \\ 2x + z = 1 \end{cases}$, que é equivalente ao sistema

$\begin{cases} 2x + z = 1 \\ 2y + z = 1 \\ x + 2y + 4z = 1 \end{cases}$, cuja forma matricial é $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 2 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Aplicando o Critério das Linhas para o novo sistema, temos que:

$\alpha_1 = \frac{|a_{12}|+|a_{13}|}{|a_{11}|} = \frac{1+0}{2} = \frac{1}{2}$, $\alpha_2 = \frac{|a_{21}|+|a_{23}|}{|a_{22}|} = \frac{0+1}{2} = \frac{1}{2}$, e

$\alpha_3 = \frac{|a_{31}|+|a_{32}|}{|a_{33}|} = \frac{1+2}{4} = \frac{3}{4}$; assim $\alpha = max\left\{\frac{1}{2}, \frac{1}{2}, \frac{3}{4}\right\} = \frac{3}{4} < 1$, ou seja,

a matriz $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 2 & 4 \end{bmatrix}$ satisfaz o Critério das Linhas.

## Método iterativo de Gauss-Seidel

O método de Gauss-Seidel é muito parecido como método anterior; porém, nesse método, sempre que calculamos um $x_i^{(k)}$, ele é utilizado no cálculo das próximas variáveis $x_j^{(k)}$, para $i < j < n$.

Dessa forma, no método de Gauss-Seidel, começamos com uma aproximação inicial $X^{(0)} = (x_1^{(0)}, \ldots, x_n^{(0)})$ e utilizamos o processo a seguir:

$$\begin{cases} x_1^{(k+1)} = \dfrac{b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_n^{(k)}}{a_{11}} \\ x_2^{(k+1)} = \dfrac{b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_n^{(k)}}{a_{22}} \\ x_3^{(k+1)} = \dfrac{b_3 - a_{31}x_1^{(k+1)} - a_{32}x_2^{(k+1)} - a_{34}x_4^{(k)} - \cdots - a_{3n}x_n^{(k)}}{a_{33}} \\ \quad \vdots \\ x_n^{(k+1)} = \dfrac{b_n - a_{n1}x_1^{(k+1)} - a_{n2}x_2^{(k+1)} - \cdots - a_{n,n-1}x_n^{(k+1)}}{a_{nn}} \end{cases}$$

## Atividade 5

***Atende ao objetivo 2***

Resolva o sistema linear $\begin{cases} 3x_1 + x_2 - x_3 = 1 \\ x_1 - 2x_2 - x_3 = 1 \\ x_1 + x_2 + 2x_3 = 1 \end{cases}$ usando o método de Gausss-Seidel, considerando $X^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ e precisão $\varepsilon = 5 \times 10^{-2}$.

### Resposta comentada

Usando o método de Gauss-Seidel, temos que:

$$1^{a} \text{ iteração} \rightarrow \begin{cases} x_1^{(1)} = \dfrac{1 - x_2^{(0)} + x_3^{(0)}}{3} = \dfrac{1 - 0 + 0}{3} = \dfrac{1}{3} \\ x_2^{(1)} = \dfrac{-1 + x_1^{(1)} - x_3^{(0)}}{2} = \dfrac{-1 + \frac{1}{3} - 0}{2} = -\dfrac{1}{3}, \\ x_3^{(1)} = \dfrac{1 - x_1^{(1)} - x_2^{(1)}}{2} = \dfrac{1 - \frac{1}{3} + \frac{1}{3}}{2} = \dfrac{1}{2} \end{cases}$$

$$d^{(1)} = max\left\{\left|\frac{1}{3} - 0\right|, \left|-\frac{1}{3} - 0\right|, \left|\frac{1}{2} - 0\right|\right\} = \frac{1}{2} > 0,05$$

$$2^{a}\text{ iteração} \rightarrow \begin{cases} x_1^{(2)} = \dfrac{1-x_2^{(1)}+x_3^{(1)}}{3} = \dfrac{1+\frac{1}{3}+\frac{1}{2}}{3} = \dfrac{11}{18} \\ x_2^{(2)} = \dfrac{-1+x_1^{(2)}-x_3^{(1)}}{2} = \dfrac{-1+\frac{11}{18}-\frac{1}{2}}{2} = -\dfrac{4}{9} \\ x_3^{(2)} = \dfrac{1-x_1^{(2)}-x_2^{(2)}}{2} = \dfrac{1-\frac{11}{18}+\frac{4}{9}}{2} = \dfrac{5}{12} \end{cases}$$

$$d^{(2)} = max\left\{\left|\frac{11}{18}-\frac{1}{3}\right|, \left|-\frac{4}{9}+\frac{1}{3}\right|, \left|\frac{5}{12}-\frac{1}{2}\right|\right\} = \frac{5}{18} > 0,05$$

$$3^{a}\text{ iteração} \rightarrow \begin{cases} x_1^{(3)} = \dfrac{1-x_2^{(2)}+x_3^{(2)}}{3} = \dfrac{1+\frac{4}{9}+\frac{15}{36}}{3} = \dfrac{67}{108} \\ x_2^{(3)} = \dfrac{-1+x_1^{(3)}-x_3^{(2)}}{2} = \dfrac{-1+\frac{67}{108}-\frac{15}{36}}{2} = -\dfrac{43}{108} \\ x_3^{(3)} = \dfrac{1-x_1^{(3)}-x_2^{(3)}}{2} = \dfrac{1-\frac{67}{108}+\frac{43}{108}}{2} = \dfrac{7}{18} \end{cases}$$

$$d^{(3)} = max\left\{\left|\frac{67}{108}-\frac{11}{18}\right|, \left|-\frac{43}{108}+\frac{4}{9}\right|, \left|\frac{7}{18}-\frac{5}{12}\right|\right\} = \frac{5}{108} < 0,05.$$

Pelo método de Gauss-Seidel: $\overline{X} = \begin{bmatrix} \frac{67}{108} \\ -\frac{43}{108} \\ \frac{7}{18} \end{bmatrix}$.

## Critério de convergência para o método de Gauss–Seidel

A pergunta que devemos nos fazer agora é: qual é o critério de convergência para o método de Gauss-Seidel?

A resposta está no Critério de Sassenfeld, que enunciaremos a seguir.

## Teorema (Critério de Sassenfeld)

Considere o sistema linear $Ax = b$ e sejam

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

$$\beta_k = \frac{|a_{k1}|\beta_1 + |a_{k2}|\beta_2 + \ldots + |a_{kk-1}|\beta_{k-1} + |a_{kk+1}| + \ldots + |a_{kn}|}{|a_{kk}|}.$$

Se $\beta = \max_{1 \le k \le n} \beta_k < 1$, então a sequência $\{X^{(k)}\}$ gerada pelo método de Gauss-Seidel converge para a solução do sistema, para qualquer aproximação inicial $X^{(0)}$. Além disso, a velocidade de convergência aumenta, se o $\beta$ diminui. Ou seja, quanto menor o $\beta$, mais rápido encontramos a solução.

A demonstração desse teorema pode ser encontrada em:

- RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R. *Cálculo numérico*: aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

## Atividade 6

***Atende ao objetivo 2***

Verifique que o Critério de Sassenfeld é satisfeito pela matriz do sistema linear $\begin{cases} 3x_1 + x_2 - x_3 = 1 \\ x_1 - 2x_2 - x_3 = 1 \\ x_1 + x_2 + 2x_3 = 1 \end{cases}$, porém, o Critério das Linhas não é satisfeito.

### *Resposta comentada*

Vamos, primeiro, verificar o Critério de Sassenfeld:

$$\beta_1 = \frac{|a_{12}|+|a_{13}|}{|a_{11}|} = \frac{|1|+|-1|}{|3|} = \frac{2}{3} < 1$$

$$\beta_2 = \frac{|a_{21}|\beta_1+|a_{23}|}{|a_{22}|} = \frac{|1|*\frac{2}{3}+|-1|}{|-2|} = \frac{5}{6} < 1$$

$$\beta_3 = \frac{|a_{31}|\beta_1+|a_{32}|\beta_2}{|a_{33}|} = \frac{|1|*\frac{2}{3}+|1|*\frac{5}{6}}{|2|} = \frac{9}{12} = \frac{3}{4} < 1$$

Critério das Linhas

$$\alpha_1 = \frac{|a_{12}|+|a_{13}|}{|a_{11}|} = \frac{|1|+|-1|}{|3|} = \frac{2}{3} < 1$$

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

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

Assim, concluímos que $\beta < 1$ porém $\alpha = 1$. Ou seja, podemos afirmar que o método de Gauss-Seidel converge, pelo Critério de Sassenfeld. Porém, não podemos afirmar que o método de Gauss-Jacobi converge, pois o critério das linhas não é satisfeito.

Observe que, sempre que a matriz satisfaz o Critério das Linhas, ela satisfaz o Critério de Sassenfeld, pois:

$$\beta_1 = \frac{|a_{12}| + |a_{13}| + \ldots + |a_{1n}|}{|a_{11}|} = \alpha_1 < 1.$$

Por indução matemática, vamos supor que $\beta_i \leq \alpha_i < 1$ para $i = 2, \ldots, k-1$, então:

$$\beta_k = \frac{|a_{k1}|\beta_1 + |a_{k2}|\beta_2 + \ldots + |a_{kk-1}|\beta_{k-1} + |a_{kk+1}| + \ldots + |a_{kn}|}{|a_{kk}|}$$

$$< \frac{|a_{k1}| + |a_{k2}| + \ldots + |a_{kk-1}| + |a_{kk+1}| + \ldots + |a_{kn}|}{|a_{kk}|} = \alpha_k < 1.$$

Dessa forma, $\beta_k \leq \alpha_k$, para $k = 1, \ldots, n$. Então, se $\alpha_k < 1$ temos que $\beta_k < 1$, ou seja, se o Critério das Linhas for satisfeito, o Critério de Sassenfeld também será.

## Atividade 7

***Atende ao objetivo 2***

Verifique que o Critério de Sassenfeld é satisfeito pela matriz do sistema linear

$$\begin{cases} 10x_1 + x_2 - x_3 - x_4 = 1 \\ x_1 - 20x_2 - x_3 = 4 \\ x_1 + x_2 + 2x_3 = 1 \\ x_2 - x_3 - 5x_4 = 0 \end{cases}.$$

***Resposta comentada***

Critério de Sassenfeld:

$$\beta_1 = \frac{|a_{12}|+|a_{13}|+|a_{14}|}{|a_{11}|} = \frac{|1|+|-1|+|-1|}{|10|} = \frac{3}{10} < 1$$

$$\beta_2 = \frac{|a_{21}|\beta_1+|a_{23}|+|a_{24}|}{|a_{22}|} = \frac{|1|*\frac{3}{10}+|-1|+|0|}{|-20|} = \frac{13}{200} < 1$$

$$\beta_3 = \frac{|a_{31}|\beta_1+|a_{32}|\beta_2+|a_{34}|}{|a_{33}|} = \frac{|1|*\frac{3}{10}+|1|*\frac{13}{200}+|0|}{|2|} = \frac{73}{400} < 1$$

$$\beta_4 = \frac{|a_{41}|\beta_1+|a_{42}|\beta_2+|a_{43}|\beta_3}{|a_{44}|} = \frac{|0|*\frac{3}{10}+|1|*\frac{13}{200}+|-1|\frac{73}{400}}{|-5|} = \frac{99}{2000} < 1$$

Assim, concluímos que $\beta < 1$. O Critério de Sassenfeld é satisfeito.

## Algoritmos para os método de Gauss-Jacobi e Gauss-Seidel

A ideia para o algoritmo de Gauss-Jacobi é simples. Primeiro, vamos supor que o elemento $a_{kk}$ de todas as linhas seja diferente de zero. Caso contrário, temos que trocar toda essa linha por uma cuja o elemento $a_{kk}$ seja diferente de zero e trocar as linhas correspondentes, no vetor $b$. Coloque o sistema na forma a seguir:

$$\begin{cases} x_1^{(k+1)} = \dfrac{b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_n^{(k)}}{a_{11}} \\ x_2^{(k+1)} = \dfrac{b_2 - a_{21}x_1^{(k)} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_n^{(k)}}{a_{22}} \\ x_3^{(k+1)} = \dfrac{b_3 - a_{31}x_1^{(k)} - a_{32}x_2^{(k)} - a_{34}{}_4^{(k)} - \cdots - {}_{3n}x_n^{(k)}}{a_{33}} \\ \vdots \\ x_n^{(k+1)} = \dfrac{b_n - a_{n1}x_1^{(k)} - a_{n2}x_2^{(k)} - \cdots - a_{n,n-1}x_{n-1}^{(k)}}{a_{nn}} \end{cases}$$

## Ideia para o algoritmo:

Dados de entrada:

- $A$: matriz do sistema;
- $b$: vetor independente;
- $X^{(0)}$: aproximação inicial;
- $\varepsilon$: precisão.

**Passo 1:** Fazemos um loop de $k = 1,\ldots, n$ para calcular os $\alpha_k$ do Critério das Linhas, $\alpha_k = \frac{|a_{k1}|+|a_{k2}|+\ldots+|a_{kk-1}|+|a_{kk+1}|+\ldots+|a_{kn}|}{|a_{kk}|}$. Depois, testamos se $\alpha = \max_{1\leq k\leq n} \alpha_k < 1$. Caso isso não ocorra, paramos e retornamos com uma mensagem de que o Critério das Linhas não é satisfeito.

**Passo 2:** Fazemos um loop onde a pergunta para sair é se a distância entre o $X^{(k+1)}$ e $X^{(k)}$ é menor que a tolerância dada. Se for, o loop é interrompido e a solução do sistema é $X^{(k+1)}$.

**Passo 3:** Dentro do loop anterior, temos que fazer um loop de $k = 1,\ldots, n$ para calcular o vetor $X^{(k)} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}^{(k)}$.

Agora, vamos escrever o algoritmo para o Método de Gauss-Jacobi em MatLab®.

## Algoritmo:

```
function x=gaussjacobi(A,b,x0,N,tol)
% x = gaussjacobil(A,b,x0,N,tol)
%
%  Esta funcao aplica o teste das linhas a matriz A e caso
%  seja possivel ela calcula a solucao do sistema linear
%  Ax=b com numero de iteracao N e precisao tol e valor
%  inicial x0.

% calcula a dimensao da matriz A
```

```
[m n]=size(A);
% Calcula os alphas do criterio das linhas
alpha=zeros(m,1);
for i=1:m
for j=1:i-1
   alpha(i)=alpha(i)+abs(A(i,j))/abs(A(i,i));
end
for j=i+1:n
   alpha(i)=alpha(i)+abs(A(i,j))/abs(A(i,i));
end
end
% testa o criterio das linhas
if max(alpha) >= 1
  error('A matriz A nao satisfaz o criterio das linhas')
end
% inicializa as variaveis
y=b; % vetor b
x=x0; % aproximacao inicial
oldx=zeros(n,1); % variavel auxiliar
% laco principal para parar
for i=1:N
if (max(abs(oldx-x))<tol) % teste de parada
    disp('O metodo foi concluido com sucesso!')
    disp('N=')
    disp(i-1)
return
else
for i=1:m
for j=1:i-1
       y(i)=y(i)-A(i,j)*x(j); % metodo de gauss-Jacobi para linha menor
que coluna
end
for j=i+1:n
       y(i)=y(i)-A(i,j)*x(j); % metodo de gauss-Jacobi para linha maior
que coluna
end
```

```
    y(i)=y(i)/A(i,i); % divide pela diagonal principal
end
% atualiza as variaveis
    oldx=x; % variavel anterior
    x=y; % variavel atual
    y=b; % vetor b
end
end
% informa que o numero de iteracao maxima foi atingindo sem encontrar a
% solucao com a precisao desejada
  disp ('O numero de iteracao foi exedido!')
```

Agora, apresentaremos um algoritmo para o método de Gauss-Seidel. Como os métodos são muito parecidos, faremos essa apresentação em MatLab®.

## Algoritmo:

```
function x=gaussseidel(A,b,x0,N,tol)
% x = gaussseidel(A,b,x0,N,tol)
%
%  Esta funcao aplica o teste de sassenfeld a matriz A e
%  caso seja
%  possivel ela calcula a solucao do sistema linear Ax=b com
%  numero de
%  iteracao N, tolerancia tol e precisao x0.

% calcula a dimensao da matriz A
[m n]=size(A);
% Calcula os betas do criterio de sassenfeld
beta=zeros(m,1);
for i=1:m
for j=1:i-1
    beta(i)=beta(i)+abs(A(i,j))/abs(A(i,i))*beta(j);
end
for j=i+1:n
    beta(i)=beta(i)+abs(A(i,j))/abs(A(i,i));
```

```
end
end
% testa o criterio de sassenfeld
if max(beta) >= 1
  error('A matriz A nao satisfaz o criterio de sassenfeld')
end
% inicializa as variaveis
y=b; % vetor b
x=x0; % condicao inicial
oldx=b+x0; % variavel auxiliar
% laco principal para parar
for i=1:N
if (max(abs(oldx-x))<tol) % teste de parada
    disp('O metodo foi concluido com sucesso!')
    disp('N=')
    disp(i-1)
return
else
for i=1:m
for j=1:i-1
   y(i)=y(i)-A(i,j)*y(j); % metodo de gauss-seidel para linha menor que
coluna
end
for j=i+1:n
       y(i)=y(i)-A(i,j)*x(j); % metodo de gauss-seidel para linha maior
que coluna
end
      y(i)=y(i)/A(i,i); % divide pela diagonal principal
end
% atualiza as variaveis
    oldx=x; % variavel anterior
    x=y; % variavel atual
    y=b; % vetor b
end
end
```

```
% informa que o numero de iteracao maxima foi atingindo sem encontrar a
% solucao com a precisao desejada
 disp ('O numero de iteracao foi exedido!')
```

## Atividade 8

*Atende aos objetivos 1 e 2*

Qual a diferença entre o algoritmo de Gauss-Jacobi e Gauss-Seidel?

### Resposta comentada

1. No algoritmo para o Método de Gauss-Jacobi, calculamos o Critério das Linhas, e, no algoritmo de Gauss-Seidel, calculamos o Critério de Sassenfeld. Mais precisamente, no segundo, usamos beta ao invés de alpha e multiplicamos o ***termo destacado***, que é a diferença entre os dois critérios.

**Gauss-Jacobi:**

```
alpha=zeros(m,1);
for i=1:m
for j=1:i-1
   alpha(i)=alpha(i)+abs(A(i,j))/abs(A(i,i));
end
for j=i+1:n
   alpha(i)=alpha(i)+abs(A(i,j))/abs(A(i,i));
end
end
```

**Gauss-Seidel:**

```
beta=zeros(m,1);
for i=1:m
for j=1:i-1
    beta(i)=beta(i)+abs(A(i,j))/abs(A(i,i))*beta(j);
end
for j=i+1:n
    beta(i)=beta(i)+abs(A(i,j))/abs(A(i,i));
end
end
```

2. No Método de Gauss-Jacobi, usamos o vetor anterior para calcular o seguinte, sem aproveitar as coordenadas já calculadas. Já no Método de Gauss-Seidel, utilizamos as coordenadas já calculadas. Mais precisamente, os termos ***em destaque*** foram mudados. Note que, no primeiro, é a coordenada do vetor anterior e, no segundo, é a coordenada já atualizada:

**Gauss-Jacobi:**

```
for i=1:m
for j=1:i-1
      y(i)=y(i)-A(i,j)*x(j); % metodo de gauss-Jacobi para linha menor que coluna
end
for j=i+1:n
       y(i)=y(i)-A(i,j)*x(j); % metodo de gauss-Jacobi para linha maior que coluna
end
     y(i)=y(i)/A(i,i); % divide pela diagonal principal
end
% atualiza as variaveis
    oldx=x; % variavel anterior
    x=y; % variavel atual
    y=b; % vetor b
end
end
```

**Gauss-Seidel:**

```
for i=1:m
for j=1:i-1
      y(i)=y(i)-A(i,j)*y(j); % metodo de gauss-seidel para linha menor que coluna
end
for j=i+1:n
      y(i)=y(i)-A(i,j)*x(j); % metodo de gauss-seidel para linha maior que coluna
end
      y(i)=y(i)/A(i,i); % divide pela diagonal principal
end
% atualiza as variaveis
   oldx=x; % variavel anterior
   x=y; % variavel atual
   y=b; % vetor b
end
end
```

## Informações sobre a próxima aula

Depois de três aulas estudando soluções para sistemas lineares, na próxima aula, mudaremos de tema e estudaremos a Interpolação Polinomial. Até lá!

## Resumo

Nessa aula, você aprendeu:

- que **métodos iterativos** são aqueles que procuram a solução do sistema considerando uma aproximação inicial e gerando uma sequência iterativa $\{X_N\}$;
- a parar um método iterativo para sistemas lineares quando $d^{(K)} = \max_{1 \leq i \leq n} \left| x_i^{(K)} - x_i^{(K-1)} \right| < \varepsilon$, onde $\varepsilon$ é a precisão desejada.

- que os métodos de Gauss-Jacobi e Gauss-Seidel consistem no uso de uma aproximação inicial $X^{(0)}$ para calcular a primeira iteração $X^{(1)}$, e assim sucessivamente, até que se chegue a $X^{(k+1)}$, que satisfaz o critério de parada;
- que a formula do método de Gauss-Jacobi é:

$$\begin{cases} x_1^{(k+1)} = \dfrac{b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_n^{(k)}}{a_{11}} \\ x_2^{(k+1)} = \dfrac{b_2 - a_{21}x_1^{(k)} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_n^{(k)}}{a_{22}} \\ x_3^{(k+1)} = \dfrac{b_3 - a_{31}x_1^{(k)} - a_{32}x_2^{(k)} - a_{34}{}_4^{(k)} - \cdots - {}_{3n}x_n^{(k)}}{a_{33}} \\ \vdots \\ x_n^{(k+1)} = \dfrac{b_n - a_{n1}x_1^{(k)} - a_{n2}x_2^{(k)} - \cdots - a_{n,n-1}x_{n-1}^{(k)}}{a_{nn}} \end{cases}$$

- que a formula do método de Gauss-Seidel é:

$$\begin{cases} x_1^{(k+1)} = \dfrac{b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_n^{(k)}}{a_{11}} \\ x_2^{(k+1)} = \dfrac{b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_n^{(k)}}{a_{22}} \\ x_3^{(k+1)} = \dfrac{b_3 - a_{31}x_1^{(k+1)} - a_{32}x_2^{(k+1)} - a_{34}x_4^{(k)} - \cdots - a_{3n}x_n^{(k)}}{a_{33}} \\ \vdots \\ x_n^{(k+1)} = \dfrac{b_n - a_{n1}x_1^{(k+1)} - a_{n2}x_2^{(k+1)} - \cdots - a_{n,n-1}x_n^{(k+1)}}{a_{nn}} \end{cases}$$

- quais são as diferenças entre o algoritmo para os métodos de Gauss-Jacobi e de Gauss-Seidel.

## Referências

BURDEN, Richard L.; FAIRES, Douglas, *Análise numérica*, São Paulo: Pioneira Thomson Learning, 2003.

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R., *Cálculo numérico:* aspectos teóricos e computacionais, São Paulo: Pearson Makron Books, 1996.

YOUNG, D. M.; GREGORY, R. T., *A survey of numerical mathematics*, vol. I, II. New York: Addison-Wesley, 1972.

# Aula 7

## Interpolação polinomial: forma de Lagrange e forma de Newton

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Aproximar funções através de polinômios, utilizando a forma de Lagrange e a forma de Newton.

## Objetivos

Esperamos que, ao final desta aula, você seja capaz de:

1. interpolar uma função através de um sistema linear;
2. interpolar uma função pela forma de Lagrange;
3. interpolar uma função pela forma de Newton.

## Introdução

Interpolar uma função $f(x)$ consiste em aproximar essa função por outra função $g(x)$, escolhida entre uma classe de funções definida *a priori* e que satisfaça algumas propriedades. A função $g(x)$ é, então, usada em substituição à função $f(x)$. A necessidade desta substituição surge, por exemplo:

- quando são conhecidos somente os valores numéricos da função para um conjunto de pontos e é necessário calcular o valor da função em um ponto não tabelado;
- quando a função em estudo tem uma expressão tal que operações como a diferenciação e a integração são difíceis (ou mesmo impossíveis) de serem realizadas.

### Joseph Louis Lagrange (1736 – 1813)

Lagrange nasceu em Turim, Itália, e foi batizado com o nome Giuseppe Lodovico Lagrangia. Seu pai era tesoureiro do escritório público de trabalhos e fortificações de Turim. Sua mãe era filha única de um médico de Cambiano, perto de Turim. Lagrange era o primogênito de um total de onze filhos, dos quais, apenas ele e mais um irmão atingiram a idade adulta. Lagrange se interessou pela matemática quando recebeu uma cópia do livro de Halley, de 1693, sobre o uso da álgebra em óptica. Apesar de seu pai ter um cargo relativamente importante, sua família não era rica. Lagrange foi uma criança prodígio: basicamente autodidata, aos 19 anos foi indicado para professor universitário da Escola Real

de Artilharia de Turim. Membro fundador da Academia Real de Ciências de Turim, onde Lagrange sucedeu Euler e permaneceu por 20 anos. Posteriormente mudou-se para Paris, após um convite de Frederico, o Grande, para a Academia de Ciências de Paris. Lagrange ganhou incontáveis prêmios e publicou inúmeros trabalhos; dentre eles: a teoria dos números; a teoria das funções; o cálculo de probabilidades; a teoria dos grupos; as equações diferenciais; a mecânica dos fluidos; a mecânica analítica e a mecânica celeste. Sua vida foi quase que inteiramente dedicada à ciência.

Fonte: http://ecalculo.if.usp.br/historia/lagrange.htm.

## Interpretação geométrica

O gráfico a seguir esboça uma função $f(x)$ e sua respectiva interpolação $g(x)$. Para isso, considere $(n + 1)$ pontos distintos chamados de *nós da interpolação,* dados por $(x_0, f(x_0))$, $(x_1, f(x_1))$,…, $(x_n, f(x_n))$. Interpolar a função $f(x)$ consiste em encontrar uma função $g(x)$ que satisfaça:

$$\begin{cases} g(x_0) = f(x_0) \\ g(x_1) = f(x_1) \\ \quad \vdots \\ g(x_n) = f(x_n) \end{cases}$$

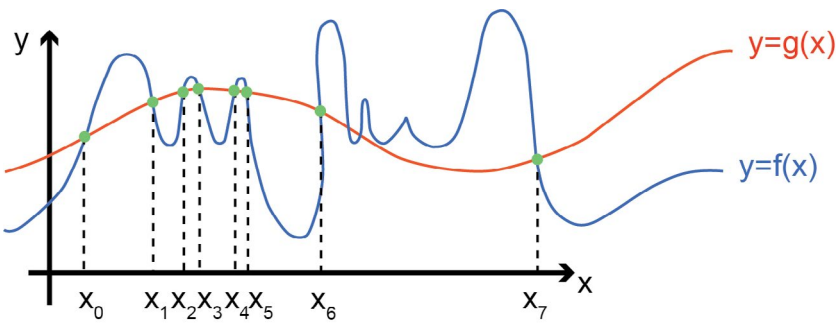



**Figura 7.1:** Esboço de uma função *f*(*x*) e de sua interpolante *g*(*x*).

Podemos escolher diversas classes de funções para $g(x)$, como as funções polinomial, racional, trigonométrica, exponencial, dentre muitas outras. Consideraremos aqui que a função interpoladora $g(x)$ pertence à classe das funções polinomiais.

## Interpolação polinomial

Utilizar polinômios para interpolar uma função é no mínimo razoável, pois: eles são funções contínuas facilmente computáveis; suas derivadas e integrais são, também, polinômios; suas raízes podem ser encontradas de maneira relativamente fácil etc.

Dessa forma, é vantajoso substituir uma função complicada, ou até mesmo desconhecida, por uma aproximação polinomial que a represente.

As interpolações polinomiais são as mais populares não somente por suas propriedades algébricas, mas principalmente pela justificativa dada pelo *Teorema de Weierstrass*, que diz que *toda função contínua pode ser aproximada por um polinômio*.

O problema geral de interpolação por meio de polinômios consiste em determinar se um polinômio $P_n(x)$ de grau no máximo igual a $n$ dado por:

$$P_n(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

que pode ser ajustado aos mesmos pontos por onde se ajusta uma função $f(x) \neq P_n(x)$.

Dessa forma, temos que encontrar os $(n + 1)$ coeficientes $a_0, a_1, a_2, \ldots, a_n$ que melhor ajustam o polinômio interpolador à função original, $f(x)$.

Para isso, precisaremos conhecer os $(n + 1)$ pontos $(x_0, f(x_0))$, $(x_1, f(x_1))$, $(x_2, f(x_2))$, …, $(x_n, f(x_n))$ da função $f(x)$ e exigir que $P_n(x_k) = f(x_k)$ para todo $k = 0, 1, 2, \ldots, n$. Esses pontos são chamados de *nós da interpolação*, porque são os pontos onde $P_n(x_k)$ coincide com $f(x_k)$.

Dessa condição, teremos o seguinte sistema linear, construído a partir dos $n$ pontos dados nos *nós de interpolação*, onde o lado esquerdo do sistema linear representa as equações do sistema linear e o lado direito, os valores da função nos nós de interpolação:

$$\begin{cases} a_0 + a_1x_0 + a_2x_0^2 + \cdots + a_nx_0^n = f(x_0) \\ a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n = f(x_1) \\ \vdots \\ a_0 + a_1x_n + a_2x_n^2 + \cdots + a_nx_n^n = f(x_n) \end{cases}$$

Com $(n + 1)$ equações e $(n + 1)$ variáveis $a_0, a_1, \ldots, a_n$.

Observe que a matriz $A$ dos coeficientes do sistema é dada por

$$A = \begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{pmatrix}$$

que é conhecida como matriz de *Vandermonde*. Observe que os elementos da primeira coluna da matriz são iguais a 1, porque representam o termo linear $a_0$ do polinômio $P_n(x)$.

Essa matriz, além de representar em suas linhas termos que estão em progressão geométrica, tem a propriedade mais importante para nós, neste caso, de que *det* $(A) \neq 0$, desde que $x_0, x_1, \ldots, x_n$ sejam distintos. Dessa forma, este sistema linear admite solução única. Isto significa que sempre que os *nós de interpolação* $(x_0, f(x_0))$, $(x_1, f(x_1))$, …, $(x_n, f(x_n))$ são distintos, o polinômio interpolador existe e é único.

Existem várias maneiras de encontrar o polinômio interpolador. Uma delas seria resolver o sistema linear obtido anteriormente. Vejamos como isso poderia ser feito.

**Exemplo 1:** Considere a função $f(x)$ que passa pelos pontos $(-1, 1)$, $(0, 2)$ e $(1, -1)$. Vamos encontrar o polinômio interpolador para essa função $f(x)$. Comece montando o sistema linear:

$$\begin{cases} P(-1) = a_0 + a_1(-1) + a_2(-1)^2 = 1 \\ P(0) = a_0 + a_1(0) + a_2(0)^2 = 2 \\ P(1) = a_0 + a_1(1) + a_2(1)^2 = -1 \end{cases}$$

Logo, teremos o sistema linear:

$$\begin{cases} a_0 - a_1 + a_2 = 1 \\ a_0 = 2 \\ a_0 + a_1 + a_2 = -1 \end{cases}$$

Substituindo $a_0 = 2$ na primeira e na terceira equação, teremos:

$$\begin{cases} -a_1 + a_2 = 1 - a_0 = 1 - 2 = -1 \\ a_1 + a_2 = -1 - a_0 = -1 - 2 = -3 \end{cases}$$

Organizando, teremos,

$$\begin{cases} -a_1 + a_2 = -1 \\ a_1 + a_2 = -3 \end{cases}$$

Resolvendo o sistema, chegaremos a $a_1 = -1$ e $a_2 = -2$. Logo, a solução do sistema linear é dada por, $a_0 = 2$, $a_1 = -1$ e $a_2 = -2$.

Com isso, $P_2(x) = 2 - x - 2x^2$ é o polinômio de interpolação para a função $f(x)$ que passa pelos pontos (–1, 1), (0, 2) e (1, –1).

## Atividade 1

***Atende ao objetivo 1***

Considere a função $f(x)$ que passa pelos pontos (–1, 15), (0, 8) e (3, –1). Determine o polinômio interpolador para essa função $f(x)$.

### Resposta comentada

Como foram fornecidos 3 pontos ou nós de interpolação $(x, y)$, então $n + 1 = 3$, isto é, $n = 2$. Com isso, precisamos determinar o polinômio $P_2(x)$ de grau no máximo igual a 2 dado por $P_2(x) = a_0 + a_1x + a_2x^2$ de tal forma que, fazendo as devidas substituições, temos:

$$\begin{cases} P(-1) = a_0 + a_1(-1) + a_2(-1)^2 = 15 \\ P(0) = a_0 + a_1(0) + a_2(0)^2 = 8 \\ P(3) = a_0 + a_1(3) + a_2(3)^2 = -1 \end{cases}$$

Logo, teremos o sistema linear:

$$\begin{cases} a_0 - a_1 + a_2 = 15 \\ a_0 = 8 \\ a_0 + 3a_1 + 9a_2 = -1 \end{cases}$$

Substituindo $a_0 = 8$ na primeira e na terceira equação, teremos:

$$\begin{cases} -a_1 + a_2 = 15 - a_0 = 15 - 8 = 7 \\ 3a_1 + 9a_2 = -1 - a_0 = -1 - 8 = -9 \end{cases}$$

Organizando, temos:

$$\begin{cases} -a_1 + a_2 = 7 \\ 3a_1 + 9a_2 = -9 \end{cases}$$

Podemos agora multiplicar a primeira equação por 3, que teremos:

$$\begin{cases} -3a_1 + 3a_2 = 21 \\ 3a_1 + 9a_2 = -9 \end{cases}$$

Agora basta somar as duas equações, que teremos:

$(-3a_1 + 3a_2) + (3a_1 + 9a_2) = 21 + (-9) = 12 \Rightarrow 12a_2 = 12 \Rightarrow a_2 = 1$

Finalmente, utilizando a equação $-a_1 + a_2 = 7$, temos:

$a_1 = a_2 - 7 = 1 - 7 \Rightarrow a_1 = -6$.

Logo, a solução do sistema linear é dada por $a_0 = 8$, $a_1 = -6$ e $a_2 = 1$.

Com isso, $P_2(x) = 8 - 6x + x^2$ é o polinômio de interpolação para a função $f(x)$ que passa pelos pontos (–1, 15), (0, 8) e (3, –1).

Encontrar o polinômio interpolador resolvendo sistemas lineares pode ser demasiadamente trabalhoso. Além disso, devido a matriz ser de Vandermonde, a resolução do sistema linear pelo método da eliminação

Gaussiana pode levar a sérios erros de arredondamento, obtendo, assim, um polinômio que não seja o desejado.

Estudaremos em seguida a Forma de Lagrange e a Forma de Newton, que são outras maneiras de se obter o mesmo polinômio.

Na teoria, todas essas maneiras são equivalentes, mas a escolha do método a ser utilizado pode ser motivada por vários fatores, como o tempo computacional, a estabilidade do sistema linear etc.

## Forma de Lagrange

Sejam $(x_i, y_i)$ os $(n + 1)$ pontos distintos da função $f(x)$, onde $y_i = f(x_i)$, para todo $i = 0, 1, 2, \ldots, n$. Consideremos, para todo $k = 0, 1, 2, \ldots, n$, os seguintes polinômios $L_k(x)$ de grau $n$ dados por:

$$L_k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

$$= \prod_{\substack{j=0 \\ j\neq k}}^{n} \frac{(x-x_j)}{(x_k-x_j)}.$$

É fácil verificar que: se $x = x_i$ $(i \neq k)$, temos que $(x - x_i) = (x_i - x_i) = 0$ (para algum termo na parte de cima da equação), o que vai tornar toda a equação igual a zero; ou seja,

$$L_k(x_i) = \frac{(x_k-x_0)(x-x_1)\cdots(x_i-x_i)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)} = 0.$$

Se $x = x_k$, temos que

$$L_k(x_k) = \frac{(x_k-x_0)(x-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)} = 1;$$

resultando em: $L_k(x_i) = \delta_{ki} = \begin{cases} 0, se\ k \neq i \\ 1, se\ k = i \end{cases}$ para $k = 0, 1, \ldots, n$, onde $\delta_{ki}$ é conhecido como delta de Kronecker, assim chamado em honra a Leopold Kronecker.

Seja $P_n(x)$ o polinômio de grau no máximo igual a $n$ definido por:

$$P_n(x) = y_0 L_0(x) + y_1 L_1(x) + \cdots + y_n L_n(x) = \sum_{k=0}^{n} y_k L_k(x).$$

Observe que a condição $P_n(x_i) = f(x_i)$ é satisfeita para todo $i = 0, 1, 2, \ldots, n$, pois $P_n(x_i) = y_0 L_0(x_i) + y_1 L_1(x_i) + \ldots + y_i L_i(x_i) + \ldots + y_n L_n(x_i) = y_i L_i(x_i) = y_i = f(x_i)$.

O polinômio $P_n(x)$ assim definido é chamado de *Forma de Lagrange do Polinômio de Interpolação*.

## Atividade 2

***Atende ao objetivo 2***

Considere a função $f(x)$ que passa pelos pontos (–1, 15), (0, 8) e (3, –1). Determine o polinômio interpolador para essa função $f(x)$ utilizando a forma de Lagrange.

### Resposta comentada

Como foram fornecidos 3 pontos (nós de interpolação), então $n + 1 = 3$, isto é, $n = 2$. Com isso, precisamos determinar o polinômio $P_2(x)$ de grau no máximo igual a 2 dado por:

$P_2(x) = y_0L_0(x) + y_1L_1(x) + y_2L_2(x) = (15)L_0(x) + (8)L_1(x) + (-1)L_2(x)$.

Para cada $k = 0, 1, 2$, os polinômios $L_k(x)$ são dados por:

$$L_0(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} = \frac{(x-0)(x-3)}{(-1-0)(-1-3)} = \frac{(x^2-3x)}{4},$$

$$L_1(x) = \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} = \frac{(x+1)(x-3)}{(0+1)(0-3)} = \frac{(x^2-2x-3)}{-3},$$

$$L_2(x) = \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} = \frac{(x+1)(x-0)}{(3+1)(3-0)} = \frac{(x^2+x)}{12}.$$

Portanto:

$$P_2(x) = (15)\left[\frac{x^2-3x}{4}\right] + (8)\left[\frac{x^2-2x-3}{-3}\right] + (-1)\left[\frac{x^2+x}{12}\right].$$

Agrupando os coeficientes semelhantes, temos:

$$P_2(x) = \left[\frac{15}{4} - \frac{8}{3} - \frac{1}{12}\right]x^2 + \left[\frac{-45}{4} + \frac{16}{3} - \frac{1}{12}\right]x + \left[\frac{-24}{-3}\right].$$

Efetuando as contas, obtemos:

$$P_2(x) = \left[\frac{45-32-1}{12}\right]x^2 + \left[\frac{-135+64-1}{12}\right]x + [8] \Rightarrow$$

$$\Rightarrow P_2(x) = \left[\frac{12}{12}\right]x^2 + \left[\frac{-72}{12}\right]x + [8] \Rightarrow$$

$$\Rightarrow P_2(x) = [1]x^2 + [-6]x + [8].$$

Com isso, $P_2(x) = 8 - 6x + x^2$ é o polinômio de interpolação para a função $f(x)$ que passa pelos pontos (–1, 15), (0, 8) e (3, –1).

---

A Forma de Lagrange que acabamos de estudar tem um inconveniente. Se dados os $(n + 1)$ pontos distintos $(x_i, y_i)$ da função $f(x)$,

onde $y_i = f(x_i)$, para todo $i = 0, 1, 2, \ldots, n$, então iremos construir um polinômio interpolador de grau $p$. No entanto, se acrescentarmos mais um ponto distinto $(x_{n+1}, y_{n+1})$ no conjunto, ficaremos com $(n + 2)$, e então iremos construir um polinômio interpolador de grau $p + 1$. Mas o trabalho feito anteriormente terá que ser praticamente todo refeito. Seria interessante se houvesse a possibilidade de, conhecido o polinômio de grau $p$, passar para o polinômio de grau $p + 1$ apenas acrescentando mais um termo no polinômio de grau $p$. Veremos a seguir que a Forma de Newton do polinômio de interpolação leva em consideração essa possibilidade.

## Forma de Newton

Para a construção do polinômio interpolador pela forma de Newton, precisaremos antes entender a notação de *diferença dividida* de uma função. Sejam $(x_i, y_i)$ os $(n + 1)$ pontos distintos da função $f(x)$, onde $y_i = f(x_i)$, para todo $i = 0, 1, 2, \ldots, n$. Definimos o *operador diferenças divididas* por:

$$\begin{cases} f\left[x_k\right] = f\left(x_k\right), \forall k = 0,1,2,\cdots,n, \\ f\left[x_0, x_1, \ldots, x_n\right] = \dfrac{f\left[x_1, x_2, \ldots, x_n\right] - f\left[x_0, x_1, \ldots, x_{n-1}\right]}{x_n - x_0}, \end{cases}$$

onde $f[x_0, x_1, \ldots, x_n]$ é a *diferença dividida* de ordem $n$ da função $f(x)$ sobre os pontos $x_0, x_1, \ldots, x_n$. Assim, usando a definição, temos:

$$\begin{cases} f\left[x_0\right] = f\left(x_0\right) & \left(\text{Ordem Zero}\right) \\ f\left[x_0, x_1\right] = \dfrac{f\left[x_1\right] - f\left[x_0\right]}{x_1 - x_0} = \dfrac{f\left(x_1\right) - f\left(x_0\right)}{x_1 - x_0} & \left(\text{Ordem 1}\right) \\ f\left[x_0, x_1, x_2\right] = \dfrac{f\left[x_1, x_2\right] - f\left[x_0, x_1\right]}{x_2 - x_0} & \left(\text{Ordem 2}\right) \\ f\left[x_0, x_1, x_2, x_3\right] = \dfrac{f\left[x_1, x_2, x_3\right] - f\left[x_0, x_1, x_2\right]}{x_3 - x_0} & \left(\text{Ordem 3}\right) \\ \vdots \quad\quad\quad\quad \vdots & \quad\quad\quad\quad \vdots \\ f\left[x_0, x_1, \ldots, x_n\right] = \dfrac{f\left[x_1, x_2, \ldots, x_n\right] - f\left[x_0, x_1, \ldots, x_{n-1}\right]}{x_n - x_0} & \left(\text{Ordem n}\right) \end{cases}$$

Observe que, do lado direito de cada uma das igualdades anteriores,

devemos aplicar sucessivamente a definição de diferença dividida (da linha anterior) até que os cálculos envolvam apenas o valor da função nos pontos. Por exemplo:

$$f\left[x_0, x_1, x_2\right] = \frac{f\left[x_1, x_2\right] - f\left[x_0, x_1\right]}{x_2 - x_0} = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}.$$

Dessa maneira, podemos calcular as diferenças divididas de uma função de modo mais prático. Para isso, iremos construir uma tabela com cálculos sistemáticos das diferenças divididas.

Sejam $(x_i, y_i)$ os $(n + 1)$ pontos distintos da função $f(x)$, onde $y_i = f(x_i)$, para todo $i = 0, 1, 2, \ldots, n$. Então podemos construir a tabela:

**Tabela 7.1:** Cálculo sistemático das diferenças divididas.

| $x_i$ | Ordem Zero | Ordem 1 | Ordem 2 | Ordem 3 | … | Ordem $n$ |
|---|---|---|---|---|---|---|
| $x_0$ | $f[x_0]$ | | | | | |
| | | $f[x_0, x_1]$ | | | | |
| $x_1$ | $f[x_1]$ | | $f[x_0, x_1, x_2]$ | | | |
| | | $f[x_1, x_2]$ | | $f[x_0, x_1, x_2, x_3]$ | | |
| $x_2$ | $f[x_2]$ | | $f[x_1, x_2, x_3]$ | | | |
| | | $f[x_2, x_3]$ | | $f[x_1, x_2, x_3, x_4]$ | ⋱ | |
| $x_3$ | $f[x_3]$ | | $f[x_2, x_3, x_4]$ | | … | $f[x_0, x_1, \ldots, x_n]$ |
| | | $f[x_3, x_4]$ | | ⋮ | | |
| $x_4$ | $f[x_4]$ | | ⋮ | $f[x_{n-3}, x_{n-2}, x_{n-1}, x_n]$ | | |
| | | ⋮ | $f[x_{n-2}, x_{n-1}, x_n]$ | | | |
| ⋮ | ⋮ | $f[x_{n-1}, x_n]$ | | | | |
| $x_n$ | $f[x_n]$ | | | | | |

Esta tabela é construída de modo que cada uma dessas diferenças divididas é uma fração, cujo numerador é sempre a diferença entre duas diferenças divididas consecutivas e de ordem imediatamente inferior; e cujo denominador é a diferença entre os dois extremos dos pontos envolvidos.

## Atividade 3

*Atende ao objetivo 3*

Considere a função $f(x)$ que passa pelos pontos (–1, 15), (0, 8) e (3, –1). Construa a tabela de diferenças divididas.

### Resposta comentada

Usando o esquema prático da tabela de diferenças divididas, temos:

| $x_i$ | $f[x_i]$ | $f[x_i, x_{i+1}]$ | $f[x_i, x_{i+1}, x_{i+2}]$ |
|---|---|---|---|
| –1 | 15 | | |
| | | –7 | |
| 0 | 8 | | 1 |
| | | –3 | |
| 3 | –1 | | |

Observe que:

$$f\left[x_0, x_1\right] = f\left[-1, 0\right] = \frac{f\left[x_1\right] - f\left[x_0\right]}{x_1 - x_0} = \frac{f(0) - (-1)}{(0) - (-1)} = \frac{(8) - (15)}{1} = -7,$$

$$f\left[x_1, x_2\right] = f\left[0, 3\right] = \frac{f\left[x_2\right] - f\left[x_1\right]}{x_2 - x_1} = \frac{f(3) - (0)}{(3) - (0)} = \frac{(-1) - (8)}{3} = \frac{-9}{3} = -3$$

e

$$f\left[x_1, x_2, x_3\right] = f\left[-1, 0, 3\right] = \frac{f\left[x_1, x_2\right] - f\left[x_o, x_1\right]}{x_2 - x_0} = \frac{\left(-3\right) - \left(-7\right)}{\left(3\right) - \left(-1\right)} = \frac{4}{4} = 1.$$

Como veremos mais adiante, os resultados a serem utilizados na construção do polinômio interpolador na forma de Newton são os primeiros valores de cada coluna de diferenças divididas. No entanto, teremos que construir a tabela inteira, pois os valores dependem uns dos outros.

Uma propriedade importante das diferenças divididas é que ela é simétrica nos argumentos. Isto significa que $f[x_0, x_1, \ldots, x_k] = f[x_{j0}, x_{j1}, \ldots, x_{jk}]$, onde $x_{j0}, x_{j1}, \ldots, x_{jk}$ é qualquer permutação de 0, 1, 2, …, $k$. Por exemplo, para $k = 1$:

$$f\left[x_0, x_1\right] = \frac{f\left[x_1\right] - f\left[x_0\right]}{x_1 - x_0} = \frac{f\left[x_0\right] - f\left[x_1\right]}{x_0 - x_1} = f\left[x_1, x_0\right].$$

Da mesma maneira, podemos mostrar, para k = 2, que:

$$f[x_0, x_1, x_2] = f[x_0, x_2, x_1] = f[x_1, x_0, x_2] = f[x_1, x_2, x_0] = f[x_2, x_0, x_1] = f[x_2, x_1, x_0].$$

Agora que sabemos como montar uma tabela de diferenças finitas, estamos prontos para definir a forma de Newton para o polinômio interpolador. Vejamos a seguir:

- Forma de Newton para o polinômio interpolador: seja $f(x)$ uma função contínua e com derivadas contínuas em $[a, b]$. Além disso, considere $(x_i, y_i)$ os $(n + 1)$ pontos distintos desta função, onde $y_i = f(x_i)$ com $x_i$ em $[a, b]$, para todo $i = 0, 1, 2, \ldots, n$. Definimos então as funções:

(1) $f\left[x_0, x\right] = \frac{f\left[x\right] - f\left[x_0\right]}{x - x_0}$, definida em $[a, b]$, para $x \neq x_0$.

(2) $f\left[x_0, x_1, x\right] = \frac{f\left[x_0, x\right] - f\left[x_0, x_1\right]}{x - x_1}$, definida em $[a, b]$, para $x \neq x_0$ e $x \neq x_1$.

(3) $f\left[x_0, x_1, x_2, x\right] = \dfrac{f\left[x_0, x_1, x\right] - f\left[x_0, x_1, x_2\right]}{x - x_2}$, definida em [*a, b*], para $x \neq x_0, x \neq x_1$ *e* $x \neq x_2$.

⋮

($n$+1) $f\left[x_0, x_1, \ldots, x_n, x\right] = \dfrac{f\left[x_0, x_1, \ldots, x_{n-1}, x\right] - f\left[x_0, x_1, \ldots, x_n\right]}{x - x_n}$, definida em [*a, b*], para $x \neq x_k$, $k = 0, 1, 2, \ldots, n$.

Observe que, nessas funções, acrescentamos sucessivamente, na diferença dividida, o próximo ponto da tabela. O objetivo é encontrar uma fórmula de recorrência para a função $f(x)$. Assim, de (1), temos:

$$f\left[x_0, x\right] = \frac{f\left[x\right] - f\left[x_0\right]}{x - x_0} = \frac{f\left(x\right) - f\left(x_0\right)}{x - x_0} \Rightarrow$$

$$\Rightarrow \left(x - x_0\right) f\left[x_0, x\right] = f\left(x\right) - f\left(x_0\right) \Rightarrow$$

$$\Rightarrow f\left(x\right) = f\left(x_0\right) + \left(x - x_0\right) f\left[x_0, x\right].$$

Considerando que $P_0(x)$ é o polinômio de grau zero que interpola $f(x)$ em $x = x_0$, então, $P_0(x) = f(x_0) = f[x_0]$. Daí, podemos escrever:

$$f(x) = P_0(x) + (x - x_0) f(x_0, x).$$

Note que $E_0(x) = f(x) - P_0(x) = (x - x_0) f[x_0, x]$ é o erro cometido ao se aproximar $f(x)$ por $P_0(x)$.

Vamos agora construir o polinômio $P_1(x)$ de grau menor ou igual a um que interpola $f(x)$ em $x_0$ e $x_1$. Temos de (2), (usando (1)), que:

$$f\left[x_0, x_1, x\right] = f\left[x_1, x_0, x\right] = \frac{f\left[x_0, x\right] - f\left[x_1, x_0\right]}{x - x_1}$$

$$= \frac{\dfrac{f\left(x\right) - f\left(x_0\right)}{x - x_0} - f\left[x_1, x_0\right]}{x - x_1} \Rightarrow$$

$$\Rightarrow f\left[x_0, x_1, x\right] = \frac{f\left(x\right) - f\left(x_0\right) - \left(x - x_0\right) f\left[x_1, x_0\right]}{\left(x - x_0\right)\left(x - x_1\right)} \Rightarrow$$

$$\Rightarrow f\left(x\right) = f\left(x_0\right) + \left(x - x_0\right) f\left[x_0, x_1\right] + \left(x - x_0\right)\left(x - x_1\right) f\left[x_0, x_1, x\right].$$

Assim, $f(x) = P_1(x) + E_1(x)$, onde:

$$P_1(x) = f(x_0) + (x - x_0) f[x_0, x_1]$$

e

$$E_1(x) = f(x) - P_1(x) = (x - x_0)\,(x - x_1)\, f[x_0, x_1, x].$$

Nesse caso, $E_1(x)$ é o erro cometido ao se aproximar $f(x)$ por $P_1(x)$.

Vamos agora construir o polinômio $P_2(x)$ de grau menor ou igual a dois que interpola $f(x)$ em $x_0$, $x_1$ e $x_2$. Temos de (3), (usando (1) e (2)), que:

$$f\left[x_0, x_1, x_2, x\right] = f\left[x_2, x_1, x_0, x\right] = \frac{f\left[x_1, x_0, x\right] - f\left[x_2, x_1, x_0\right]}{x - x_2} \Rightarrow$$

$$\Rightarrow f\left[x_0, x_1, x_2, x\right] = \frac{\dfrac{f(x) - f(x_0) - (x - x_0) f\left[x_1, x_0\right]}{(x - x_0)(x - x_1)} - f\left[x_2, x_1, x_0\right]}{x - x_2} \Rightarrow$$

$$\Rightarrow f\left[x_0, x_1, x_2, x\right] = \frac{f(x) - f(x_0) - (x - x_0) f\left[x_1, x_0\right] - (x - x_0)(x - x_1) f\left[x_2, x_1, x_0\right]}{(x - x_0)(x - x_1)(x - x_2)} \Rightarrow$$

$$\Rightarrow f(x) = f(x_0) + (x - x_0) f\left[x_0, x_1\right] + (x - x_0)(x - x_1) f\left[x_0, x_1, x_2\right] +$$

$$+ (x - x_0)(x - x_1)(x - x_2) f\left[x_0, x_1, x_2, x\right].$$

Assim, $f(x) = P_2(x) + E_2(x)$, onde:

$$P_2(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2]$$

e

$$E_2(x) = f(x) - P_2(x) = (x - x_0)\,(x - x_1)\,(x - x_2)\, f[x_0, x_1, x_2, x].$$

Nesse caso, $E_2(x)$ é o erro cometido ao se aproximar $f(x)$ por $P_2(x)$.

De forma sucessiva, a partir de $(n + 1)$, polinômio $P_n(x)$ de grau menor ou igual a n que interpola $f(x)$ em $x_0, x_1, \ldots, x_n$, será dado por:

$$P_n(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)\,(x - x_1) f[x_0, x_1, x_2] + \ldots$$
$$+ \ldots + (x - x_0)\,(x - x_1) \ldots (x - x_{n-1}) f[x_0, x_1, \ldots, x_n].$$

Esta fórmula é chamada de *Forma de Newton do polinômio de interpolação.* O erro é dado por:

$$E_n(x) = f(x) - P_n(x) = (x - x_o)\,(x - x_1) \ldots (x - x_n) f[x_o, x_1, \ldots, x_n, x].$$

## Atividade 4

*Atende ao objetivo 3*

Considere a função $f(x)$ que passa pelos pontos $(-1, 15)$, $(0, 8)$ e $(3, -1)$. Determine o polinômio interpolador para essa função $f(x)$ utilizando a forma de Newton.

### Resposta comentada

Usando o esquema prático da tabela de diferenças divididas (feita na atividade anterior), temos:

| $x_i$ | $f[x_i]$ | $f[x_i, x_{i+1}]$ | $f[x_i, x_{i+1}, x_{i+2}]$ |
|---|---|---|---|
| –1 | 15 | | |
| | | –7 | |
| 0 | 8 | | 1 |
| | | –3 | |
| 3 | –1 | | |

Como temos 3 pontos, então $n + 1 = 3$, logo $n = 2$. Dessa maneira, o polinômio de interpolação na forma de Newton é dado por:

$$P_2(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2].$$

Da tabela, temos que $f[x_0] = 15$, $f[x_0, x_1] = -7$ e $f[x_0, x_1, x_2] = 1$. Logo:

$$P_2(x) = 15 + (x + 1)(-7) + (x + 1)(x - 0)(1).$$

Agrupando os termos semelhantes, obtemos:

$$P_2(x) = x^2 - 6x + 8.$$

## Conclusão

Nesta aula, vimos que existem várias maneiras de encontrar o polinômio interpolador. Encontrar o polinômio interpolador resolvendo sistemas lineares pode ser demasiadamente trabalhoso devido a matriz ser de Vandermonde, onde a resolução do sistema linear pelo método da eliminação Gaussiana pode levar a sérios erros de arredondamento, obtendo, assim, um polinômio que não seja o desejado. Vimos que a Forma de Lagrange e a Forma de Newton são outras maneiras de se obter o mesmo polinômio. Na teoria, todas essas maneiras são equivalentes, mas a escolha do método a ser utilizado pode ser motivada por vários fatores como o tempo computacional, estabilidade do sistema linear, etc.

## Resumo

Vimos os pontos que se seguem.

- O problema geral de interpolação por meio de polinômios consiste em determinar-se um polinômio $P_n(x)$ de grau no máximo igual a $n$ dado por:

$$P_n(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

- Uma maneira de encontrar os $(n + 1)$ coeficientes $a_0, a_1, a_2, \ldots, a_n$ que melhor ajustam o polinômio interpolador seria resolvendo o sistema linear:

$$\begin{cases} a_0 + a_1x_0 + a_2x_0^2 + \cdots + a_nx_0^n = f(x_0) \\ a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n = f(x_1) \\ \vdots \\ a_0 + a_1x_n + a_2x_n^2 + \cdots + a_nx_n^n = f(x_n) \end{cases}$$

- O polinômio $P_n(x) = y_0L_0(x) + y_1L_1(x) + \ldots + y_nL_n(x)$ é chamado de *Forma de Lagrange do polinômio de interpolação*, onde:

$$L_k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

- A *Forma de Newton do polinômio de interpolação* é dada por

$$P_n(x) = f(x_0) + (x - x_0)\, f[x_0, x_1] + (x - x_0)(x - x_1)\, f[x_0, x_1, x_2] + \ldots$$
$$+ \ldots + (x - x_0)(x - x_1) \ldots (x - x_{n-1})\, f[x_0, x_1, \ldots, x_n].$$

## Referências

BURDEN, R. L; e FAIRES, D. *Análise numérica. São Paulo:* Pioneira Thomson Learning, 2003.

FRANCO, N. B. *Cálculo numérico*. São Paulo: Pearson Prentice Hall, 2006.

RUGGIERO, M. A. G; Lopes, V. L. da R. *Cálculo numérico*: aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

# Aula 8

## Integração numérica: regra dos Trapézios e regra de Simpson

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Demonstrar de forma numérica uma integral definida usando a regra dos Trapézios e a regra de Simpson.

## Objetivo

Esperamos que, ao final desta aula, você seja capaz de:

1. resolver numericamente uma integral definida pela regra dos Trapézios;
2. resolver numericamente uma integral definida pela regra de Simpson.

## Introdução

Muitos problemas em matemática são resolvidos por meio do cálculo de uma integral definida. No entanto, a função a ser integrada pode ser extremamente complicada, o que torna o problema difícil de ser resolvido.

Aprendemos em cálculo integral que, para resolver uma integral definida, precisamos encontrar uma primitiva para a função e, em seguida, aplicar o teorema fundamental do cálculo. O grande problema é que nem toda função possui uma primitiva. Além disso, podemos ter casos em que a expressão analítica da função é desconhecida e sabemos apenas alguns pontos discretos por onde essa função passa, obtidos por meio de experimentos.

Veremos nessa aula métodos numéricos para resolver essas integrais e driblar as dificuldades.

## Fórmulas de Newton-Cotes

A ideia básica de uma integração numérica é substituirmos a função $f(x)$ por um polinômio $p(x)$ que aproxime esta função no intervalo $[a,b]$. Com isso, ao invés de resolvermos $\int_a^b f(x)dx$, iremos calcular uma aproximação, resolvendo a integral $\int_a^b p(x)dx$.

As vantagens de calcular a integral de um polinômio são óbvias, devido à facilidade nos cálculos por meio de fórmulas conhecidas do cálculo integral.

As fórmulas fechadas de Newton-Cotes são fórmulas de integração do tipo

$$\int_a^b f(x)dx \approx w_0 f\left(x_0\right) + w_1 f\left(x_1\right) + \cdots + w_n f\left(x_n\right) = \sum_{i=1}^{n} w_i f\left(x_i\right),$$

onde os coeficientes $w_i$ são os pesos determinados de acordo com o grau do polinômio interpolador. Nas fórmulas de Newton-Cotes os pontos $x_i$'s do intervalo $[a,b]$ são igualmente espaçados, dados por $x_i = a + ih$, com $i = 0,1, \ldots, n$ e $h = \dfrac{\left(b-a\right)}{n}$.

Veremos a seguir a regra dos Trapézios e a regra 1/3 de Simpson, que são algumas fórmulas fechadas de Newton-Cotes. Chamamos essas fórmulas de *Newton-Cotes* em homenagem a Isaac Newton e Roger

Cotes, cientistas ingleses do final do século XVII e do princípio do século XVIII.

## Regra do Trapézio

O cálculo da integral definida $\int_a^b f(x)dx$ pode ser interpretado graficamente como sendo a área delimitada pela função $f(x)$, o eixo $x$ e o intervalo $[a,b]$. Seja $p_1(x)$ o polinômio de grau 1 (uma reta) que interpola a função $f(x)$ em $a = x_0$ e $b = x_1$.

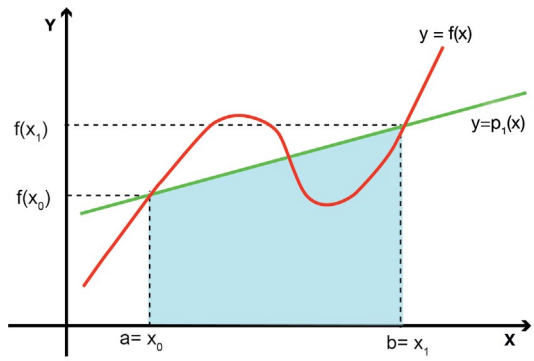



**Figura 8.1:** Aproximação pela área do trapézio.

Então podemos aproximar $\int_a^b f(x)dx$ por $\int_a^b p_1(x)dx$, isto é, usar a área do trapézio para aproximar a integral (vide **Figura 8.1**).

Sabemos que área de um trapézio é dada por $A = \frac{(b+B)}{2}h$, onde $b$ é a base menor, $B$ é a base maior e $h$ é a altura. Nesse caso, temos que

$$\int_a^b f(x)dx \approx \frac{\left[f(a)+f(b)\right]}{2}h,$$

onde $h = b - a = x_1 - x_0$.

Uma maneira de mostrarmos como aproximamos a integral de $f(x)$ entre os limites $a$ e $b$ é utilizar a forma de Lagrange para interpolar o polinômio $p_1(x)$. Você pode rever a fórmula de Lagrange para interpolação na Aula 7. Vejamos:

$$p_1(x) = y_0 L_0(x) + y_1 L_1(x) = f(x_0)\frac{(x-x_1)}{(x_0-x_1)} + f(x_1)\frac{(x-x_0)}{(x_1-x_0)} \Rightarrow$$

$$\Rightarrow p_1(x) = f(x_0)\frac{(x-x_1)}{(-h)} + f(x_1)\frac{(x-x_0)}{(h)}.$$

Daí, teremos que:

$$\int_a^b f(x)dx \approx I = \int_{x_0}^{x_1} p_1(x)dx$$

$$= \frac{-f(x_0)}{h}\int_{x_0}^{x_1}(x-x_1)dx + \frac{f(x_1)}{h}\int_{x_0}^{x_1}(x-x_0)dx \Rightarrow$$

Lembrete: A altura do trapézio se encontra no eixo das abcissas e as bases se encontram no eixo das ordenadas, se você ficou com dúvida, volte à **Figura 8.1** e rode a folha 90º para ver o trapézio ao qual estamos nos referindo. Fazendo isso, você vai notar que $h = x_1 - x_0$.

$$\Rightarrow I = \frac{-f(x_0)}{h}\frac{(x-x_1)^2}{2}\Big|_{x_0}^{x_1} + \frac{f(x_1)}{h}\frac{(x-x_0)^2}{2}\Big|_{x_0}^{x_1} =$$

$$-\left(\frac{f(x_0)}{h}\frac{(x_1-x_1)^2}{2} - \frac{f(x_0)}{h}\frac{(x_0-x_1)^2}{2}\right)$$

$$+\left(\frac{f(x_1)}{h}\frac{(x_1-x_0)^2}{2} - \frac{f(x_1)}{h}\frac{(x_0-x_0)^2}{2}\right)$$

$$\Rightarrow I = f(x_0)\frac{h}{2} + f(x_1)\frac{h}{2} \Rightarrow I = \frac{h}{2}\left[f(x_0) + f(x_1)\right] = \frac{\left[f(a) + f(b)\right]h}{2}.$$

É fácil perceber, olhando a **Figura 8.2**, que o erro pode ser grande em se aproximar uma integral definida pela área de um único trapézio. Visando diminuir o erro, podemos fazer a regra do Trapézio repetidas vezes dentro de um mesmo intervalo $[a,b]$. Vamos fazer isso por meio da regra dos Trapézios Repetida.

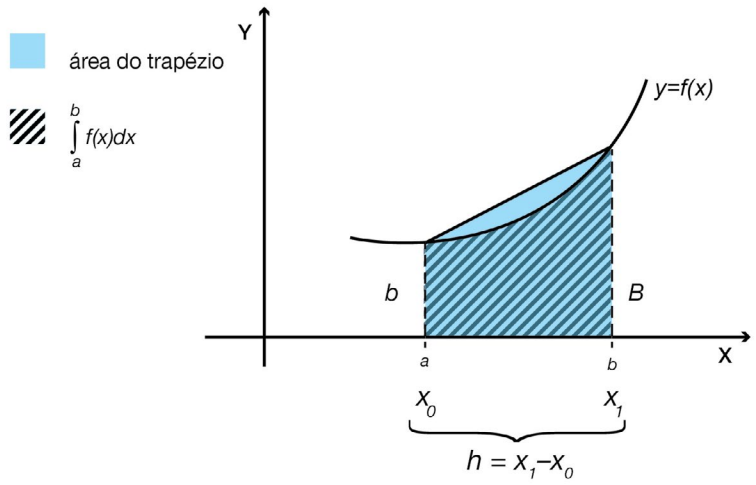



**Figura 8.2:** Aproximação pela área do trapézio.

## Regra dos Trapézios Repetida

Na regra do trapézio, se o intervalo $[a,b]$ for pequeno, a aproximação é razoável. No entanto, se a amplitude do intervalo $[a,b]$ for grande, o erro também pode ser grande. Com o objetivo de reduzir o erro, adotamos a estratégia de subdividir o intervalo de integração e aplicamos a regra do trapézio repetidas vezes, isto é, em cada subintervalo utilizamos a regra do Trapézio (vide **Figura 8.3**).

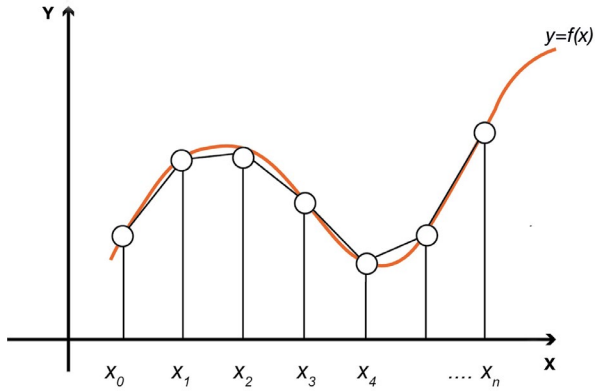



**Figura 8.3:** Área dos trapézios repetidos.

Vamos, então, dividir intervalo $[a,b]$ em $n$ subintervalos igualmente espaçados de comprimento (amplitude) igual a $h = \frac{(b-a)}{n}$, onde $x_i = a + ih$, com $i = 0,1, \ldots, n$. Observe que $x_0 = a$ e $x_n = b$. Para cada subintervalo $[x_{i-1}, x_i]$ com $i = 1, 2, \ldots, n$, temos que $A_i = \frac{h}{2}\left[f\left(x_{i-1}\right) + f\left(x_i\right)\right]$ é a área do $i$-ésimo trapézio. Então, temos:

$$\int_a^b f(x)dx \approx I = A_1 + A_2 + \cdots + A_n = \sum_{i=1}^{n} \frac{h}{2}\left[f\left(x_{i-1}\right) + f\left(x_i\right)\right] \Rightarrow$$

Chamamos de "I" a variável que representa o valor aproximado da integral.

$$\Rightarrow I = \frac{h}{2}\sum_{i=1}^{n}\left[f\left(x_{i-1}\right) + f\left(x_i\right)\right] \Rightarrow$$

$$\Rightarrow I = \frac{h}{2}\left(\left[f\left(x_0\right) + f\left(x_1\right)\right] + \left[f\left(x_1\right) + f\left(x_2\right)\right] + \left[f\left(x_2\right) + f\left(x_3\right)\right] + \cdots\right.$$

$$\left. + \left[f\left(x_{n-1}\right) + f\left(x_n\right)\right]\right) \Rightarrow$$

$$\int_a^b f(x)dx \approx \frac{h}{2}\left[f\left(a\right) + 2f\left(x_1\right) + 2f\left(x_2\right) + \cdots + 2f\left(x_{n-1}\right) + f\left(b\right)\right].$$

Esta última fórmula é conhecida como a *regra dos Trapézios Repetidos*, ou de uma maneira mais simplificada, chamada apenas de *regra dos Trapézios*. Os cálculos podem ser executados seguindo o algoritmo a seguir:

• Algoritmo: Integração pela regra dos Trapézios

Dados *a, b, n* e $y = f(x)$, faça $h = \frac{(b-a)}{n}$.

*soma* = 0;

para $i = 1 : n - 1$

$x_i = a + ih$

*soma* = *soma* + $f(x_i)$;

*Fim. Escreva:*

*Integral* = $\frac{h}{2}\left[f\left(a\right) + 2 * soma + f\left(b\right)\right]$

## Atividade 1

*Atende ao objetivo 1*

Calcule uma aproximação para $\int_0^1 e^x dx$ pela regra dos trapézios, com $n = 10$.

### Resposta comentada

Da integral temos que $a = 0$, $b = 1$ e $f(x) = e^x$. Como $n = 10$, então $h = \frac{(1-0)}{10} = \frac{1}{10} = 0{,}1$. Então $x_0 = 0$, $x_1 = 0{,}1$, $x_2 = 0{,}2, \ldots, x_9 = 0{,}9$ e $x_{10} = 1$. Aplicando a regra dos trapézios repetidos, temos:

$$\int_a^b f(x)dx \approx \frac{0{,}1}{2}\left[e^0 + 2e^{0{,}1} + 2e^{0{,}2} + \cdots + 2e^{0{,}8} + 2e^{0{,}9} + e^1\right] \approx 1{,}719713.$$

Repare que, nessa atividade, colocamos uma integral fácil de ser resolvida diretamente usando o teorema fundamental do cálculo, apenas para ilustrar.

Temos que:

$$\int_0^1 e^x dx = e - 1 \approx 1{,}718282.$$

Se compararmos com o resultado aproximado obtido na atividade, iremos perceber que já cometemos erro na terceira casa decimal. Para obtermos uma aproximação melhor, precisamos aumentar o número de divisões *n*, mas, nesse caso, o uso de um computador seria extremamente necessário. No entanto, a regra dos Trapézios não é a única maneira de calcular uma integral de forma numérica. Existem outros métodos em que, mesmo com *n* pequeno, obtemos uma aproximação melhor na maioria das vezes. A regra de Simpson, que veremos em seguida, é uma delas.

## Regra 1/3 de Simpson

Para estabelecer a regra de Simpson, iremos interpolar a função $f(x)$ usando um polinômio do grau 2 (parábola). Seja $p_2(x)$ o polinômio que interpola a função $f(x)$ em $x_0 = a$, $x_1 = \frac{a+b}{2}$ e $x_2 = b$. Observe que, se $h$ é a distância entre cada subintervalo, então $x_1 - x_0 = h$ (ou $x_1 = x_0 + h$) e $x_2 - x_0 = 2h$ (ou $x_2 = x_0 + 2h$).

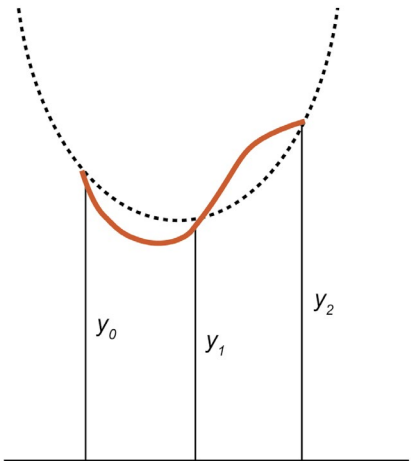



**Figura 8.4:** Regra de Simpson – aproximação de área por uma parábola.

Então podemos aproximar $\int_a^b f(x)dx$ por $\int_a^b p_2(x)dx$, isto é, usar a área abaixo da parábola para aproximar a integral (vide **Figura 8.4**). Vamos utilizar a forma de Lagrange para interpolar o polinômio $p_2(x)$. Vejamos:

$$p_2(x) = y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x),$$

onde

$$\begin{cases} L_0(x) = \dfrac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} = \dfrac{(x-x_1)(x-x_2)}{(-h)(-2h)} \\ L_1(x) = \dfrac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} = \dfrac{(x-x_0)(x-x_2)}{(h)(-h)} \\ L_2(x) = \dfrac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} = \dfrac{(x-x_0)(x-x_1)}{(2h)(h)} \end{cases}$$

Logo, temos que:

$$\int_a^b f(x)dx \approx I = \int_{x_0}^{x_2} p_2(x)dx = \int_{x_0}^{x_2} \left[ y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x) \right] dx \Rightarrow$$

$$\Rightarrow I = y_0 \cdot \int_{x_0}^{x_2} L_0(x)dx + y_1 \cdot \int_{x_0}^{x_2} L_1(x)dx + y_2 \cdot \int_{x_0}^{x_2} L_2(x)dx \Rightarrow$$

$$\Rightarrow I = \frac{f(x_0)}{2h^2} \cdot \int_{x_0}^{x_2} (x - x_1)(x - x_2)dx - \frac{f(x_1)}{h^2} \cdot \int_{x_0}^{x_2} (x - x_0)(x - x_2)dx +$$

$$+ \frac{f(x_2)}{2h^2} \cdot \int_{x_0}^{x_2} (x - x_0)(x - x_1)dx.$$

Para resolver estas integrais, podemos usar substituição simples, fazendo $x - x_0 = hz$. Logo $x = x_0 + hz$ e $dx = hdz$. Então:

$$\begin{cases} x - x_1 = (x_0 + hz) - (x_0 + h) = h(z - 1) \\ x - x_2 = (x_0 + hz) - (x_0 + 2h) = h(z - 2). \end{cases}$$

Além disso, os novos limites de integração ficam:

1. Fazendo $x = x_0$ em $x - x_0 = hz$, temos que $hz = x_0 - x_0 = 0 \Rightarrow z = 0$.
2. Fazendo $x = x_2$ em $x - x_0 = hz$, temos que $hz = x_2 - x_0 = 2h \Rightarrow z = 2$.

Segue daí que:

$$I = \frac{f(x_0)}{2h^2} \cdot \int_0^2 (z - 1)h(z - 2)h(hdz) - \frac{f(x_1)}{h^2} \cdot \int_0^2 zh(z - 2)h(hdz) +$$

$$+ \frac{f(x_2)}{2h^2} \cdot \int_0^2 zh(z - 1)h(hdz) \Rightarrow$$

$$\Rightarrow I = \frac{h}{2} f(x_0) \cdot \int_0^2 (z^2 - 3z + 2)dz - hf(x_1) \cdot \int_0^2 (z^2 - 2z)dz + \frac{h}{2} f(x_2) \cdot \int_0^2 (z^2 - z)dz.$$

Resolvendo as integrais, temos:

$$\begin{cases} \int_0^2 (z^2 - 3z + 2)dz = \frac{2}{3} \\ \int_0^2 (z^2 - 2z)dz = -\frac{4}{3} \\ \int_0^2 (z^2 - z)dz = \frac{2}{3}. \end{cases}$$

Reescrevendo $I$ , obtemos:

$$I = \frac{h}{2} f\left(x_0\right) \cdot \left(\frac{2}{3}\right) - hf\left(x_1\right) \cdot \left(\frac{-4}{3}\right) + \frac{h}{2} f\left(x_2\right) \cdot \left(\frac{2}{3}\right) \Rightarrow$$

$$\Rightarrow I = \frac{h}{3}\left[f(x_0) + 4f(x_1) + f(x_2)\right]$$

Esta fórmula é conhecida como Regra 1/3 de Simpson para $n = 2$. Assim como na regra dos Trapézios, a fórmula de Simpson é aplicada de forma repetida no intervalo $[a,b]$. A cada dois subintervalos, iremos aplicar a regra 1/3 de Simpson. Por este motivo, o número $n$ de divisões precisa ser par na regra de Simpson Repetida.

https://mathshistory.st-andrews.ac.uk/Biographies/Simpson/

Thomas Simpson (1710 – 1761) matemático e inventor britânico, epônimo da fórmula de Simpson para aproximação de integrais definidas. A atribuição como frequente em matemática, pode ser debatida: esta fórmula tinha sido obtida 200 anos antes por Johannes Kepler, sendo chamada na Alemanha de Keplersche Fassregel.

Fonte: https://pt.wikipedia.org/wiki/Thomas_Simpson

## Regra 1/3 de Simpson Repetida

A cada dupla de subintervalos, iremos aplicar a regra 1/3 de Simpson, pois cada parábola precisará de três pontos consecutivos na interpolação (vide **Figura 8.5**). Então, uma condição necessária para aplicarmos a regra de 1/3 Simpson repetida é que *n* seja par.

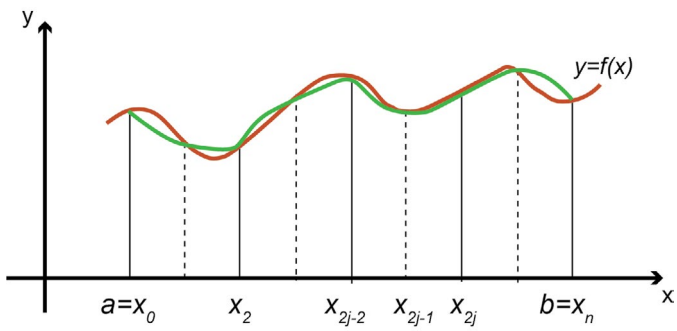



**Figura 8.5:** Regra de Simpson repetida.

Com isso, podemos escrever:

$$\int_a^b f(x)dx \approx I = \frac{h}{3}\,[\,f(x_0) + 4f(x_1) + f(x_2)] + \frac{h}{3}\,[\,f(x_2) + 4f(x_3) + f(x_4)\,] +$$

$$\frac{h}{3}\,[\,f(x_4) + 4f(x_5) + f(x_6)] + \ldots + \frac{h}{3}\,[\,f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)\,]$$

Colocando o termo $\frac{h}{3}$ em evidência e somando os termos repetidos, obtemos:

$$\int_a^b f(x)dx \approx \frac{h}{3}\,[f(a) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 4f(x_{n-1}) + f(b)]$$

Esta última fórmula é conhecida como a *regra de 1/3 de Simpson Repetida*, ou, de uma maneira mais simplificada, apenas como *regra de Simpson*. Os cálculos podem ser executados seguindo o algoritmo a seguir:

• Algoritmo: Integração pela regra de Simpson

Dados *a, b, n* (par) e $y = f(x)$, faça $h = \frac{(b-a)}{n}$ e $k = \frac{n}{2} - 1$.

$soma_{par} = 0$,

$soma_{ímpar} = f(a + h)$

Para $i = 1: k$

$$x = a + 2ih$$

$soma_{par} = soma_{par} + f(x)$;

$$x = x + h$$

$soma_{ímpar} = soma_{ímpar} + f(x)$

*Fim. Escreva:*

$$Integral = \frac{h}{3}[f(a) + 4 * soma_{ímpar} + 2 * soma_{par} + f(b)]$$

## Atividade 2

***Atende ao objetivo 2***

Calcule uma aproximação para $\int_0^1 e^x dx$ pela regra de Simpson com $n = 10$.

### Resposta comentada

Da integral, temos que $a = 0$, $b = 1$ e $f(x) = e^x$. Como $n = 10$, então $h = \frac{(1-0)}{10} = \frac{1}{10} = 0{,}1$. Portanto $x_0 = 0, x_1 = 0{,}1, x_2 = 0{,}2, \ldots, x_9 = 0{,}9$ e $x_{10} = 1$.
Aplicando a regra dos trapézios repetidos, temos:

$$\int_a^b f(x)dx \approx \frac{0{,}1}{3}\left[e^0 + 4e^{0{,}1} + 2e^{0{,}2} + 4e^{0{,}3} + \cdots + 2e^{0{,}8} + 4e^{0{,}9} + e^1\right]$$

$$\approx 1{,}718283.$$

Se compararmos esse resultado aproximado obtido na **Atividade 2** com o resultado da **Atividade 1**, perceberemos que melhoramos bastante a aproximação com relação a solução da integral $\int_0^1 e^x dx = e - 1 \approx 1{,}718282$.

## Conclusão

Nesta aula, vimos que muitas vezes é inviável resolver uma integral definida em um intervalo usando o teorema fundamental do cálculo. O uso de métodos de integração numérica se faz necessário. Existem vários métodos, dentre eles, destacam-se a *regra dos Trapézios* e a *regra de Simpson*. Quando o número de divisões do intervalo é suficientemente grande, o resultado é satisfatório.

## Resumo

Nesta aula, você estudou:

- A regra dos Trapézios:

$$\int_a^b f(x)dx \approx \frac{h}{2}[f(a)+2f(x_1)+2f(x_2)+\cdots+2f(x_{n-1})+f(b)].$$

- A regra de Simpson:

$$\int_a^b f(x)dx \approx \frac{h}{3}\,[f(a) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 4f(x_{n-1}) + f(b)]$$

## Referências

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R. *Cálculo numérico*: aspectos teóricos e computacionais. São Paulo. Pearson Makron Books, 1996.

FRANCO, Neide B. *Cálculo numérico*. São Paulo. Pearson Prentice Hall, 2006.

BURDEN, Richard L.; FAIRES, Douglas, *Análise numérica*. São Paulo. Pioneira Thomson Learning, 2003.

# Aula 9

## Soluções numéricas de equações diferenciais ordinárias: problemas de valor inicial - série de Taylor e método de Euler

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Apresentar métodos numéricos para resolução de equações diferencias ordinárias com valores iniciais.

## Objetivos

Esperamos que, ao final desta aula, você seja capaz de:

1. reconhecer e exemplificar uma série de Taylor;
2. reconhecer e exemplificar o método de Euler;
3. utilizar o método de Euler.

## Pré-requisitos

Para um bom aproveitamento desta aula, é importante você relembrar os conceitos de sistema lineares apresentados na Aula 4 e o curso de Equações Diferenciais Ordinárias.

## Introdução

Suponha que você tenha uma quantidade *M* de dinheiro, depositada na poupança, e que a taxa de juros seja contínua igual a $\mu_M$, suponha ainda que você não vai mexer nessa poupança por um tempo *t*.

Note que a taxa de variação do valor depositado referente ao tempo *t,* $\frac{dM}{dt}$, é dada pelo produto da taxa de juros pelo valor depositado:

$$\frac{dM}{dt} = \mu_M M.$$

Essa equação é uma *Equação Diferencial Ordinária (EDO)*, que você estudou no curso de EDO.

Usando a técnica de separação de variáveis, que você aprendeu no curso de EDO, vemos que a solução analítica desse problema é:

$M(t) = Ce^{\mu_M t}$ , onde *C* é uma constante qualquer.

Relembrando um pouco o curso de EDO, a constante *C*, depende do valor inicial do problema, no caso que estamos tratando. Neste nosso exemplo, *C* depende da quantidade inicial de dinheiro depositado na poupança.

Ou seja, se você deposita R$ 100,00, na poupança com uma taxa de juros de 6% ao ano, ao final de 1 ano você terá R$ 106.18 (pois $M(1) = 100e^{0,06*1} = 106.18$); assim como, ao final de 3 anos, você terá R$ 119.72 (pois $M(3)=100e^{0,06*3} = 119.72$)

A EDO que vimos acima é de primeira ordem, pois a ordem da maior derivada envolvida é 1. Assim, a *ordem de uma equação diferencial ordinária* é a ordem da derivada de maior grau.

Note que, no exemplo acima, resolvemos a EDO analiticamente, infelizmente nem sempre é fácil ou possível encontrarmos soluções analíticas para as EDOs, e por isso é necessário aprendermos os métodos numéricos para soluções de equações lineares com *problemas de valores iniciais (PVI).*

## Problemas de valores inicial (PVI)

Em muitos casos, um problema de valor inicial (PVI) tem solução única. Isso foi apresentado a você no curso de EDO. Porém, essas soluções são difíceis de encontrar analiticamente, sendo necessário encontrar soluções numéricas para o problema.

Trataremos aqui de problemas da forma:

$$(PVI)\begin{cases} y' = f(x,y) \\ y(x_0) = y_0 \end{cases}$$

onde encontraremos uma sequência $x_1, x_2, \ldots, x_n$ igualmente espaçadas, ou seja, $x_{j+1} - x_j = h$, e calcularemos as aproximações $y_j \cong y(x_j)$.

Observem que, em métodos que são de passos simples, ou seja, $x_{j+1} - x_j = h$, é preciso calcular os valores de $f(x,y)$ e de suas derivadas em muitos pontos. Nesses métodos, é muito difícil estimar os erros cometidos.

## Método da Série de Taylor

Teoricamente, os métodos que usam a série de Taylor encontram soluções para qualquer equação diferencial. Porém, computacionalmente, eles podem se tornar muito instáveis para EDO's de ordem elevadas.

Vamos primeiro entender o que é a série de Taylor.

Seja $y(x)$ uma função contínua, com derivadas contínuas em torno de um ponto $x_0$. A série de Taylor dessa função em torno de $x_0$ é dada por:

$$y(x) = y(x_0) + y'(x_0)(x - x_0) + y''(x_0)\frac{(x-x_0)^2}{2!} + \cdots + y^{(k)}(x_0)\frac{(x-x_0)^k}{k!} + y^{(k+1)}(\xi_1)\frac{(x-x_0)^{k+1}}{(k+1)!}$$

onde $\xi_1$ está entre $x_0$ e $x$.

Substituindo $x = x_1$, na equação anterior, temos:

$$y(x_1) = y(x_0) + y'(x_0)(x_1 - x_0) + y''(x_0)\frac{(x_1-x_0)^2}{2!} + \cdots + y^{(k)}(x_0)\frac{(x_1-x_0)^k}{k!} + y^{(k+1)}(\xi_1)\frac{(x_1-x_0)^{k+1}}{(k+1)!}$$

Considere o passo para resolvermos a EDO, como sendo $h = x_1 - x_0$. Assim:

$$y(x_1) = y(x_0) + y'(x_0)h + y''(x_0)\frac{h^2}{2!} + \cdots + y^{(k)}(x_0)\frac{h^k}{k!} + y^{(k+1)}(\xi_1)\frac{h^{k+1}}{(k+1)!}$$

Podemos aproximar $y(x_1)$ pela expressão a seguir e ainda estimar o erro cometido:

$$y(x_1) = y(x_0) + y'(x_0)h + y''(x_0)\frac{h^2}{2!} + \cdots + y^{(k)}(x_0)\frac{h^k}{k!};$$

Com erro dado por:

$$E(x_1) = y^{(k+1)}(\xi_1)\frac{h^{k+1}}{(k+1)!}$$

Agora podemos usar $x_1$ para encontrar $x_2$ e assim por diante. O que nos daria o seguinte resultado:

$$y(x_{n+1}) = y(x_n) + y'(x_n)h + y''(x_n)\frac{h^2}{2!} + \cdots + y^{(k)}(x_n)\frac{h^k}{k!} \qquad (1)$$

Com erro dado por:

$$E(x_n) = y^{(k+1)}(\xi_n)\frac{h^{k+1}}{(k+1)!}$$

Note que, $\xi_n$ pertence a um intervalo fechado; e se $y$ é suficientemente derivável, o valor da sua derivada $y^{(k+1)}(\xi_n)$ pode ser estimado pelo máximo da função nesse intervalo:

$$M_{k+1} = \max_{x \in I} | y^{(k+1)}(x) |,$$

sendo assim, o erro cometido é no máximo dado por:

$$E(x_n) \leq \frac{M_{k+1}h^{k+1}}{(k+1)!}$$

Observe que nos falta conhecer os valores das derivadas de $y$, para conhecermos $y(x_{n+1})$. Para simplificar a notação, vamos chamar $y(x_n) = y_n$.

Voltando ao nosso problema de valor inicial:

$$(PVI)\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \end{cases}$$

Observe que a função $f(x,y)$ é conhecida e que podemos derivar em relação a $x$ e a $y$, assim:

$$y' = f(x,y) = f$$

$$y'' = f_x(x, y(x)) + f_y(x, y(x))\, y'(x) = f_x + f_y y' = f_x + f_y f,$$

$$y''' = f_{xx}(x, y(x)) + f_{xy}(x, y(x)\, y'(x) + [f_{yx}(x, y(x)) + f_{yy}(x, y(x))\, y'(x)]\, y'(x)$$

$$+ f_y(x, y(x))\, y''(x) = f_{xx} + f_{xy} f + (f_{yx} + f_{yy} f) f + f_y (f_x + f_y f)$$

$$= f_{xx} + 2 f_{xy} f + f_{yy} f^2 + f_x f_y + f_y{}^2 f$$

Podemos continuar derivando; o importante é que sempre vamos encontrar a derivada de $y$ em função de $f(x,y)$ e das suas derivadas.

Assim, se voltarmos à equação (1), conseguimos uma fórmula para todos os $y(x_{n+1})$, em que todos os valores são conhecidos.

Note que a terceira derivada de $y$ já nos indica a dificuldade dos cálculos.

Vamos chamar esse método de *método de série de Taylor de ordem k* quando pararmos na k-ésima derivada de $y$, na equação (1).

## Atividade 1

***Atende ao objetivo 1***

Encontre a fórmula da série de Taylor de ordem 2, para o PVI a seguir:

$$(PVI)\begin{cases} y' = xy \\ y(0) = 1 \end{cases}$$

Depois, calcule o valor de $y(1)$, usando $h = 0{,}2$.

### Resposta comentada

Como estamos procurando o método de série de Taylor de ordem 2, temos:

$y' = f(x,y) = xy$

$y'' = f_x(x,y(x)) + f_y(x,y(x))y'(x) = y + x^2y.$

Vamos substituir na fórmula (1) até a segunda derivada,

$$y_{n+1} = y_n + y_n'h + y_n''\frac{h^2}{2!} = y_n + x_n y_n h + \frac{(y_n + x_n^2 y_n)h^2}{2}$$

Assim,

$$y_1 = y_0 + x_0 y_0 h + \frac{(y_0 + x_0^2 y_0)h^2}{2} = 1 + 0*1*0,2 + \frac{(1+0^2*1)0,2^2}{2} = 1,02$$

$$y_2 = y_1 + x_1 y_1 h + \frac{(y_1 + x_1^2 y_1)h^2}{2} = 1,02 + 0,2*1,02*0,2$$

$$+ \frac{(1,02 + 0,2^2*1,02)0,2^2}{2} = 1,082$$

$$y_3 = 1,082 + 0,4*1,082*0,2 + \frac{(1,082 + 0,4^2*1,082)0,2^2}{2} = 1,1937$$

$$y_4 = 1,1937 + 0,6*1,1937*0,2 + \frac{(1,1937 + 0,6^2*1,1937)0,2^2}{2} = 1,3694$$

$$y_5 = 1,3694 + 0,8*1,3694*0,2 + \frac{(1,3694 + 0,8^2*1,3694)0,2^2}{2} = 1,6334$$

Desse modo, podemos construir uma tabela:

| $n$ | $x_n$ | $y_n$ |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0,2 | 1,02 |
| 2 | 0,4 | 1,082 |
| 3 | 0,6 | 1,1937 |
| 4 | 0,8 | 1,3694 |
| 5 | 1,0 | 1,6334 |

## Método de Euler

O método de Euler é precisamente o método da serie de Taylor de ordem 1 que resolve uma EDO, numericamente. Como vimos na seção anterior, considerando o $PVI \begin{cases} y' = f(x,y) \\ y(x_0) = y_0 \end{cases}$ temos que o método de Euler é dado por:

$y(x_{n+1}) = y(x_n) + y'(x_n)h \Rightarrow$

$y_{n+1} = y_n + f_n h$

com erro dado por:

$$E(x_n) \leq \frac{M_2 h^2}{2} \text{ , onde } M_2 = \max_{x \in I} |y''(x)| .$$

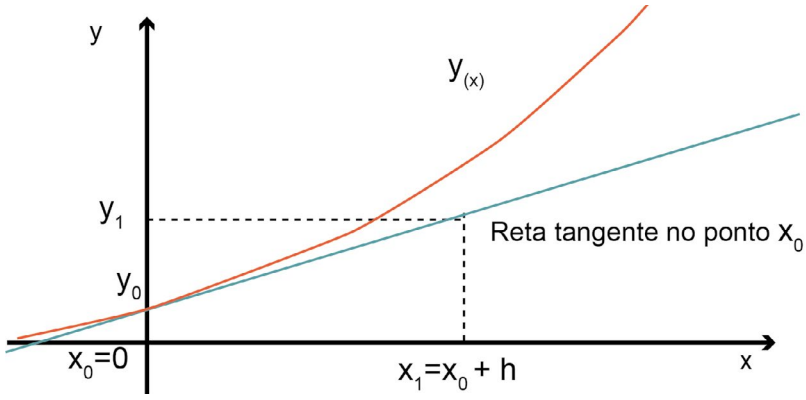



**Figura 9.1:** interpretação geométrica do método de Euler.

## Atividade 2

*Atende aos objetivos 2 e 3*

Encontre a fórmula do método de Euler para o PVI a seguir:

$$(PVI)\begin{cases} y' = xy \\ y(0) = 1 \end{cases}$$

Depois, calcule o valor de $y(1)$, usando $h = 0{,}2$.

### Resposta comentada

Como estamos procurando o método de série de Taylor de ordem 1, temos: $y' = f(x,y) = xy$

Vamos substituir na fórmula (1) até a segunda derivada,

$y_{n+1} = y_n + y_n'h = y_n + x_n y_n h$

Assim,

$y_1 = y_0 + x_0 y_0 h = 1 + 0 * 1 * 0{,}2 = 1$

$y_2 = y_1 + x_1 y_1 h = 1 + 0{,}2 * 1 * 0{,}2 = 1{,}04$

$y_3 = 1{,}04 + 0{,}4 * 1{,}04 * 0{,}2 = 1{,}1232$

$y_4 = 1{,}1232 + 0{,}6 * 1{,}1232 * 0{,}2 = 1{,}258$

$y_5 = 1{,}258 + 0{,}8 * 1{,}258 * 0{,}2 = 1{,}4593$

Assim podemos construir uma tabela:

| $n$ | $x_n$ | $y_n$ |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0,2 | 1 |
| 2 | 0,4 | 1,04 |
| 3 | 0,6 | 1,1232 |
| 4 | 0,8 | 1,258 |
| 5 | 1,0 | 1,4593 |

## Atividade 3

*Atende aos objetivos 2 e 3*

Seja o $PVI \begin{cases} y' = y \\ y(0) = 1 \end{cases}$; trabalhe com 4 casas decimais e use o método de Euler para aproximar $y(0{,}2)$ com $E \leq 10^{-2}$.

### *Resposta comentada*

Como vimos, a fórmula para o método de Euler é dada por:

$E(x_n) \leq \frac{M_2 h^2}{2}$, onde $M_2 = \max_{x \in I} | y''(x) |$.

sendo assim, temos que calcular a segunda derivada de *y*,

$y' = y \Rightarrow y'' = y' = y$

Como, nesse caso, conhecemos a solução analítica do PVI e sabemos que $y(x) = e^x$, temos que:

$$M_2 = \max_{x \in I} | y''(x) | = \max_{x \in [0\,,\,0{,}02]} | e^x | = e^{0,2} = 1{,}2214$$

$$E(x_n) \leq \frac{M_2 h^2}{2} = \frac{1{,}2214 h^2}{2} = 0{,}6107 h^2 \Rightarrow 0{,}6107 h^2 \leq 10^{-2} \Rightarrow h \leq 0{,}128.$$

Para trabalharmos com pontos igualmente espaçados, tome $h = 0{,}1$. Assim, temos:

$y' = f(x,y) = y$

Vamos substituir na fórmula (1) até a primeira derivada:

$y_{n+1} = y_n + y_n' h = y_n + y_n h = y_n(1 + 0{,}1) = 1{,}1 y_n$

Então:

$y_1 = 1{,}1 y_0 = 1{,}1 * 1 = 1{,}1$

$y_2 = 1{,}1 y_1 = 1{,}1 * 1{,}1 = 1{,}21$

| $n$ | $x_n$ | $y_n$ |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0,1 | 1,1 |
| 2 | 0,2 | 1,21 |

## *Atividade 4*

***Atende aos objetivos 2 e 3***

Calcule $y(1{,}1)$ usando o método de Euler para o $PVI \begin{cases} xy' = x + y \\ y(1) = 1 \end{cases}$;

***Resposta comentada***

Tome $h = 0{,}1$. Assim temos:

$$y' = f(x, y) = \frac{x + y}{x}.$$

Vamos substituir na fórmula (1) até a primeira derivada,

$$y_{n+1} = y_n + y_n' h = y_n + \frac{x_n + y_n}{x_n} h.$$

Então:

$$y_1 = y_0 + \frac{x_0 + y_0}{x_0} h = 1 + \frac{1+1}{1} 0{,}1 = 1{,}2.$$

| $n$ | $x_n$ | $y_n$ | $f_n$ | $y_{n+1}$ |
|---|---|---|---|---|
| 0 | 1 | 1 | 2 | 1,2 |

## Informações sobre a próxima aula

Na próxima aula, continuaremos aprendendo sobre solução para EDO's. Veremos métodos de série de Taylor de ordem maiores e como eles se chamam. Até lá!

## Resumo

Nesta aula, você estudou:

- a fórmula para o método de série de Taylor de ordem k:

$$y_{n+1} = y_n + y_n' h + y_n'' \frac{h^2}{2!} + \cdots + y_n^{(k)} \frac{h^k}{k!}$$

que tem erro dado por:

$$E(x_n) = y^{(k+1)}(\xi_n)\frac{h^{k+1}}{(k+1)!}$$

- *a fórmula para o Método de Euler, que* é dada por:

$$y_{n+1} = y_n + f_n h$$

com erro dado por:

$$E(x_n) \leq \frac{M_2 h^2}{2}, \text{ onde } M_2 = \max_{x \in I} |\, y''(x)\,|.$$

## Referências

BURDEN, Richard L.; FAIRES, Douglas. *Análise numérica*. São Paulo: Pioneira Thomson Learni ng, 2003.

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R., *Cálculo numérico:* aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

YOUNG, D. M.; GREGORY, R. T., *A survey of numerical mathematics*. vol. I, II. New York: Addison-Wesley, 1972.

# Aula 10

## Soluções numéricas de equações diferenciais ordinárias: problemas de valor inicial - método de Euler aperfeiçoado e métodos de Runge-Kutta

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Apresentar os métodos de *Euler Aperfeiçoado* e *Runge-Kutta* para resolução de equações diferenciais ordinárias com valores iniciais.

## Objetivos

Esperamos que, ao final dessa aula, você seja capaz de:

1. utilizar o método de Euler Aperfeiçoado;
2. utilizar os métodos de Runge-Kutta.

## Pré-requisitos

Para um bom aproveitamento dessa aula, é importante que você relembre a aula anterior.

## Introdução

Nessa aula, tentaremos adaptar os métodos de série de Taylor para novos métodos em que as boas propriedades dos primeiros continuem existindo, porém sem o cálculo de derivada de $f(x,y)$.

## Método de Euler aperfeiçoado

Considere o problema de valor inicial:

$$(PVI)\begin{cases} y' = f(x,y) \\ y(x_0) = y_0 \end{cases}$$

Vamos dispor da fórmula do método da série de Taylor de ordem 2. Seja $y(x)$ uma função contínua com derivadas contínuas em torno de um ponto $x_0$, a série de Taylor dessa função em torno de $x_n$ é dada por:

$$y_{n+1} = y_n + y_n'(x_{n+1} - x_n) + y''(x_n)\frac{(x_{n+1} - x_n)^2}{2!}.$$

Pela natureza do problema, temos que:

$$y'' = f_x(x,y(x)) + f_y(x,y(x))y'(x) = f_x + f_y y' = f_x + f_y f,$$

$$y_{n+1} = y_n + fh + (f_x + f_y f)\frac{h^2}{2}, \qquad (1)$$

Podemos olhar a série de Taylor para a função $f(x,y)$ em torno de $(x_n,y_n)$, assim:

$$f(x_{n+1},y_{n+1}) = f(x_n,y_n) + f_x(x_n,y_n)(x_{n+1} - x_n) + f_y(x_n,y_n)(y_{n+1} - y_n)$$
$$+ \textit{Erro de ordem } 2$$

Podemos aproximar a equação anterior:

$$f(x_{n+1},y_{n+1}) \cong f + f_x h + f_y(y_{n+1} - y_n).$$

Agora use o método de Euler para aproximar $y_{n+1} = y_n + fh$ assim:

$$f(x_{n+1},y_{n+1}) \cong f + f_x h + f_y(y_n + fh - y_n) = f + f_x h + f_y fh = f + (f_x + f_y f)h.$$

O que resulta em:

$$(f_x + f_y f)h = f(x_{n+1},y_{n+1}) - f,$$

Voltando à fórmula (1), temos que:

$$y_{n+1} = y_n + fh + (f_x + f_y f)\frac{h^2}{2} = y_n + fh + (f(x_{n+1}, y_n + fh) - f)\frac{h}{2}$$

$$y_{n+1} = y_n + (f(x_n, y_n) + f(x_{n+1}, y_n + fh))\frac{h}{2},$$

Essa é a formula para o método de Euler Aperfeiçoado. A fórmula para o erro é dada por:

$$E(x_n) \leq \frac{M_3 h^3}{6}, \text{ onde } M_3 = \max_{x \in I} |y'''(x)|.$$

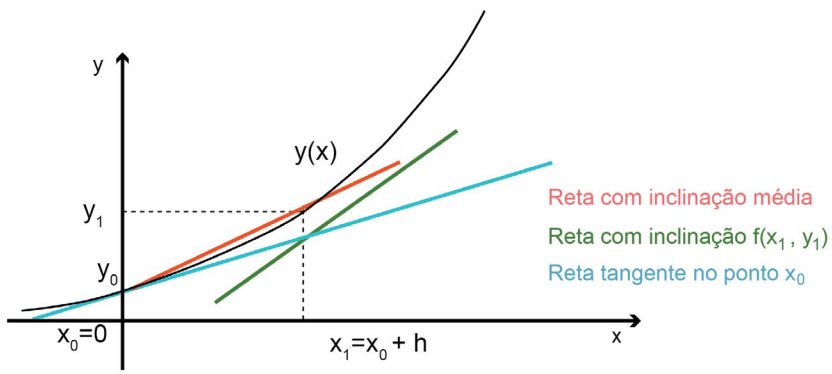



**Figura1.1:** Interpretação geométrica do método de Euler Aperfeiçoado.

## Atividade 1

***Atende ao objetivo 1***

Encontre a fórmula do método de Euler aperfeiçoado para o PVI a seguir:

$$(PVI)\begin{cases} y' = xy \\ y(0) = 1 \end{cases}$$

Depois calcule o valor de $y(1)$, usando $h = 0{,}2$.

### *Resposta comentada*

Usando a fórmula para o método de Euler Aperfeiçoado vista acima:

$$y_{n+1} = y_n + (f(x_n, y_n) + f(x_{n+1}, y_n + fh))\frac{h}{2},$$
$$y_{n+1} = y_n + (x_n y_n + x_{n+1}(y_n + x_n y_n h))\frac{h}{2},$$
$$y_{n+1} = y_n + \frac{(x_n + x_{n+1}(1 + x_n h))y_n h}{2}$$

Assim,

$$y_1 = y_0 + \frac{x_0 + x_1(1 + x_0 h))y_0 h}{2} = 1 + \frac{(0 + 0,2(1 + 0 * 0,2))1 * 0,2}{2} = 1,02$$
$$y_2 = y_1 + \frac{x_1 + x_2(1 + x_1 h))y_1 h}{2} = 1,02 + \frac{(0,2 + 0,4(1 + 0,2 * 0,2))1,02 * 0,2}{2}$$
$$= 1,0828$$
$$y_3 = 1,0828 + \frac{(0,4 + 0,6(1 + 0,4 * 0,2))1,0828 * 0,2}{2} = 1,1963$$
$$y_4 = 1,1963 + \frac{(0,6 + 0,8(1 + 0,6 * 0,2))1,1963 * 0,2}{2} = 1,3753$$
$$y_5 = 1,3753 + \frac{(0,8 + 1(1 + 0,8 * 0,2))1,3753 * 0,2}{2} = 1,6449$$

Desse modo, podemos construir uma tabela:

| $n$ | $x_n$ | $y_n$ |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0,2 | 1,02 |
| 2 | 0,4 | 1,0828 |
| 3 | 0,6 | 1,1963 |
| 4 | 0,8 | 1,3753 |
| 5 | 1,0 | 1,6449 |

## Métodos de Runge-Kutta

Os métodos de Runge-Kutta têm como princípio aproximar as derivadas pela fórmula de Taylor e usá-las de modo a, aos poucos, eliminar os cálculos das derivadas, como fizemos no método de Euler e no método de Euler Aperfeiçoado.

Dessa forma, o método de Euler é um exemplo de método de Runge-Kutta de primeira ordem, e o método de Euler aperfeiçoado é um exemplo de método de Runge-Kutta de segunda ordem.

De uma forma geral, os métodos de Runge-Kutta de segunda ordem têm a forma:

$$y_{n+1} = y_n + ha_1 f(x_n, y_n) + ha_2 f(x_n + b_1 h, y_n + b_2 h y_n')$$

No método de Euler Aperfeiçoado temos:

$$a_1 = a_2 = \frac{1}{2},\ b_1 = b_2 = 1$$

Da mesma maneira que construímos o método de Runge-Kutta de segunda ordem (método de Euler aperfeiçoado), podemos construir os métodos de Runge-Kutta de terceira e quarta ordem.

## Método de Runge-Kutta de terceira ordem

O *Método de Runge-Kutta de terceira ordem* é dado por:

$$y_{n+1} = y_n + \frac{2}{9}k_1 + \frac{1}{3}k_2 + \frac{4}{9}k_3, \qquad (2)$$

onde:

$$k_1 = hf(x_n, y_n),$$
$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right),$$
$$k_3 = hf\left(x_n + \frac{3}{4}h, y_n + \frac{3}{4}k_2\right).$$

## Atividade 2

*Atende ao objetivo 2*

Encontre a fórmula do método de Runge-Kutta de terceira ordem para o PVI a seguir:

$$(PVI)\begin{cases} y' = xy \\ y(0) = 1 \end{cases}$$

Depois calcule o valor de $y(1)$, usando $h = 0{,}2$.

### Resposta comentada

Como estamos procurando o método de Runge-Kutta de terceira ordem, temos:

$y' = f(x,y) = xy$

Substituindo na fórmula (2), temos:

$$y_{n+1} = y_n + \frac{2}{9}k_1 + \frac{1}{3}k_2 + \frac{4}{9}k_3,$$

onde:

$$k_1 = hf(x_n, y_n) = 0{,}2 * x_n * y_n$$

$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right) = 0{,}2 * \left(x_n + \frac{0{,}2}{2}\right) * \left(y_n + \frac{0{,}2 * x_n * y_n}{2}\right)$$

$$k_3 = hf\left(x_n + \frac{3}{4}h, y_n + \frac{3}{4}k_2\right)$$

$$= 0{,}2 * \left(x_n + \frac{3}{4}0{,}2\right) * \left(y_n + \frac{3}{4}0{,}2 * \left(x_n + \frac{0{,}2}{2}\right) * \left(y_n + \frac{0{,}2 * x_n * y_n}{2}\right)\right)$$

Nesse ponto, é aconselhável programar na memória da calculadora o $k_1$, $k_2$ e $k_3$ e fazer a tabela a seguir.

| $n$ | $x_n$ | $k_1$ | $k_2$ | $k_3$ | $y_n$ | $y_{n+1}$ |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0,02 | 0,0305 | 1 | 1,0202 |
| 1 | 0,2 | 0,0408 | 0,0624 | 0,0747 | 1,0202 | 1,0833 |
| 2 | 0,4 | 0,0867 | 0,1127 | 0,1285 | 1,0833 | 1,1972 |
| 3 | 0,6 | 0,1437 | 0,1777 | 0,1996 | 1,1972 | 1,377 |
| 4 | 0,8 | 0,2203 | 0,2677 | 0,2998 | 1,377 | **1,6484** |

## Método de Runge-Kutta de quarta ordem

O método de Runge-Kutta de quarta ordem é dado por:

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \qquad (3).$$

onde:

$$k_1 = hf(x_n, y_n),$$
$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right),$$
$$k_3 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right),$$
$$k_4 = hf\left(x_n + h, y_n + k_3\right).$$

## Atividade 3

*Atende ao objetivo 2*

Encontre a fórmula do método de Runge-Kutta de quarta ordem para o PVI a seguir:

$$(PVI)\begin{cases} y' = xy \\ y(0) = 1 \end{cases}$$

Depois calcule o valor de $y(1)$, usando $h = 0{,}2$.

### *Resposta comentada*

Como estamos procurando o método de Runge-Kutta de terceira ordem, temos:

$y' = f(x,y) = xy.$

Substituindo na fórmula (3), temos:

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

onde:

$$k_1 = hf(x_n, y_n) = 0,2 * x_n * y_n,$$
$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right),$$
$$k_3 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right),$$
$$k_4 = hf\left(x_n + h, y_n + k_3\right).$$

Nesse ponto, é aconselhável programar na memória da calculadora o $k_1$, $k_2$, $k_3$ e $k_4$ e fazer a tabela a seguir

| $n$ | $x_n$ | $k_1$ | $k_2$ | $k_3$ | $k_4$ | $y_n$ | $y_{n+1}$ |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0,02 | 0,0202 | 0,0408 | 1 | 1,0202 |
| 1 | 0,2 | 0,0408 | 0,0624 | 0,0631 | 0,0867 | 1,0202 | 1,0833 |
| 2 | 0,4 | 0,0867 | 0,1127 | 0,114 | 0,1437 | 1,0833 | 1,1972 |
| 3 | 0,6 | 0,1437 | 0,1777 | 0,18 | 0,2204 | 1,1972 | 1,3771 |
| 4 | 0,8 | 0,2203 | 0,2677 | 0,272 | 0,3298 | 1,3771 | **1,6487** |

Sem sombra de dúvidas, o método de Runge-Kutta mais conhecido e usado é o de quarta ordem. Os métodos de Runge-Kutta têm a grande vantagem de não precisarem do cálculo das derivadas de $f(x,y)$; porém não temos como calcular uma estimativa para o erro cometido.

## Informações sobre a próxima aula

Na próxima aula, continuaremos aprendendo sobre solução para EDO's. Porém, começaremos a ver métodos que utilizam passos variáveis, ou métodos de passos múltiplos. Até lá!

## Resumo

Nessa aula, vimos que:

- o métodos de Euler Aperfeiçoado é dado por:

$$y_{n+1} = y_n + (f(x_n, y_n) + f(x_{n+1}, y_n + fh))\frac{h}{2},$$

Com erro dado por:

$$E(x_n) \leq \frac{M_3 h^3}{6}, \text{ onde } M_3 = \max_{x \in I} | y'''(x) |.$$

- o método de Runge-Kutta de terceira ordem é dado por:

$$y_{n+1} = y_n + \frac{2}{9}k_1 + \frac{1}{3}k_2 + \frac{4}{9}k_3,$$

onde:

$$k_1 = hf(x_n, y_n),$$
$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right),$$
$$k_3 = hf\left(x_n + \frac{3}{4}h, y_n + \frac{3}{4}k_2\right).$$

- o método de Runge-Kutta de quarta ordem é dado por:

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4),$$

onde:

$$k_1 = hf(x_n, y_n),$$
$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right),$$
$$k_3 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right),$$
$$k_4 = hf\left(x_n + h, y_n + k_3\right).$$

## Referências

BURDEN, Richard L; FAIRES, Douglas. *Análise numérica*. São Paulo: Pioneira Thomson Learning, 2003.

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R., *Cálculo numérico:* aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

YOUNG, D. M.; GREGORY, R. T., *A survey of numerical mathematics*. V. I; II. London: Addison-Wesley, 1972.

# Aula 11

## Soluções numéricas de equações diferenciais ordinárias: problemas de valor inicial - método de previsão-correção

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Meta

Apresentar o método de previsão-correção para a resolução de equações diferenciais ordinárias com valores iniciais.

## Objetivos

Esperamos que, ao final desta aula, você seja capaz de:

1. saber utilizar o que é um método de passo múltiplo;
2. saber utilizar um método explícito;
3. saber utilizar um método implícito;
4. saber utilizar o método de previsão-correção.

## Pré-requisitos

Para um bom aproveitamento desta aula, é importante que você relembre as Aulas 7, 8 e 9.

## Introdução

Nesta aula, estudaremos os métodos de passos múltiplos, que é diferente dos métodos de passos simples que vimos nas aulas anteriores. Nesses métodos, utilizaremos a informação em mais de um passo; ou seja, poderemos utilizar as informações em mais de um ponto $x_n$, $y_n$ e $f(x_n,y_n)$.

Os métodos de *previsão* tentam prever o que vai acontecer, enquanto os métodos de *correção* tentam corrigir os erros cometidos. Assim, um método de *previsão-correção*, como o nome já diz, prevê o que vai acontecer e corrige a previsão, caso esteja errada.

Estudaremos métodos de passos múltiplos que utilizam as aproximações polinomiais (estudas na Aula 7, para funções) e, depois, integram numericamente essas aproximações (matéria estudada na Aula 8), para, enfim, encontrarmos o nosso método de solução da EDO.

## Método de passo múltiplo

Considere o problema de valor inicial:

$$(PVI)\begin{cases} y' = f(x,y) \\ y(x_0) = y_0 \end{cases}$$

Você deve lembrar do teorema fundamental do Cálculo, que nos dá uma fórmula para o resultado de integrarmos uma derivada. Considere como antes: $h = x_{i+1} - x_i$, $i = 0,1, \ldots, n$. Assim:

$$\int_{x_n}^{x_{n+1}} y'(x)dx = \int_{x_n}^{x_{n+1}} f(x, y(x))dx$$
$$y(x_{n+1}) - y(x_n) = \int_{x_n}^{x_{n+1}} f(x, y(x))dx$$

Dessa forma, vamos ter um método do tipo Adams-Bashforth:

$$y(x_{n+1}) = y_n + \int_{x_n}^{x_{n+1}} f(x, y(x))dx.$$

Agora, devemos aproximar $\int_{x_n}^{x_{n+1}} f(x, y(x))dx$ por meio de um método de integração numérica.

## Métodos explícitos

Nos métodos explícitos, para interpolarmos a função *f(x,y(x))*, usamos os pontos $x_n$, $x_{n-1}$, $x_{n-2}$, ..., $x_{n-m}$.

Começaremos aproximando a função *f(x,y(x))* por um polinômio de grau 2. Vamos usar o polinômio de Lagrange de grau 2 para interpolar a função nos pontos $(x_n,y_n)$, $(x_{n-1},y_{n-1})$ e $(x_{n-2},y_{n-2})$. Pelo que vimos na Aula 7, temos que:

$$f(x,y(x)) = L_{-2}(x)f_{n-2} + L_{-1}(x)f_{n-1} + L_0(x)f_n$$

$$L_{-2}(x) = \frac{(x-x_{n-1})(x-x_n)}{(x_{n-2}-x_{n-1})(x_{n-2}-x_n)} = \frac{(x-x_{n-1})(x-x_n)}{(x_{n-2}-x_{n-1})(x_{n-2}-x_{n-1}+x_{n-1}-x_n)}$$

$$= \frac{(x-x_{n-1})(x-x_n)}{-h(-h-h)} = \frac{(x-x_{n-1})(x-x_n)}{2h^2}$$

$$L_{-1}(x) = \frac{(x-x_{n-2})(x-x_n)}{(x_{n-1}-x_{n-2})(x_{n-1}-x_n)} = \frac{(x-x_{n-2})(x-x_n)}{h(-h)} = \frac{(x-x_{n-2})(x-x_n)}{-h^2}$$

$$L_0(x) = \frac{(x-x_{n-2})(x-x_{n-1})}{(x_n-x_{n-2})(x_n-x_{n-1})} = \frac{(x-x_{n-2})(x-x_{n-1})}{(x_n-x_{n-1}+x_{n-1}-x_{n-2})(x_n-x_{n-1})} =$$

$$= \frac{(x-x_{n-2})(x-x_{n-1})}{2h(h)} = \frac{(x-x_{n-2})(x-x_{n-1})}{2h^2}$$

$$f(x,y(x)) = L_{-2}(x)f_{n-2} + L_{-1}(x)f_{n-1} + L_0(x)f_n$$

$$= \frac{(x-x_{n-1})(x-x_n)}{2h^2}f_{n-2} + \frac{(x-x_{n-2})(x-x_n)}{-h^2}f_{n-1} + \frac{(x-x_{n-2})(x-x_{n-1})}{2h^2}f_n.$$

Integrando *f(x,y(x))* em *x*, temos:

$$\int_{x_n}^{x_{n+1}} f(x,y(x))dx =$$

$$= \int_{x_n}^{x_{n+1}} \left( \frac{(x-x_{n-1})(x-x_n)}{2h^2}f_{n-2} + \frac{(x-x_{n-2})(x-x_n)}{-h^2}f_{n-1} + \frac{(x-x_{n-2})(x-x_{n-1})}{2h^2}f_n \right) dx$$

Faremos uma substituição de variável para resolver essa integral. Considere $\frac{x-x_n}{h} = v$, então $\frac{dx}{h} = dv \Rightarrow dx = hdv$ e $x = hv + x_n$.

Assim, $x - x_{n-1} = hv + x_n - x_{n-1} = hv + h = (v+1)h$ e $x - x_{n-2} = (v+2)h$.

Efetuando as contas:

$$\int_{x_n}^{x_{n+1}} f(x, y(x))dx =$$

$$= \int_0^1 \left( \frac{(v+1)h(hv)}{2h^2} f_{n-2} - \frac{(v+2)h(vh)}{h^2} f_{n-1} + \frac{(v+2)h(v+1)h}{2h^2} f_n \right) hdv$$

$$= \int_0^1 \left( \frac{v^2+v}{2} f_{n-2} - (v^2+2v) f_{n-1} + \frac{v^2+3v+2}{2} f_n \right) hdv =$$

$$= h \left( \frac{2v^3+3v^2}{12} f_{n-2} - \left( \frac{v^3}{3} + v^2 \right) f_{n-1} + \frac{2v^3+9v^2+12v}{12} f_n \right) \Big|_0^1$$

$$= \frac{h}{12}(5f_{n-2} - 16f_{n-1} + 23f_n).$$

Dessa forma, vamos ter um método do tipo Adams-Bashforth com 3 pontos, que é dado por:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(x, y(x))dx = y_n + \frac{h}{12}(5f_{n-2} - 16f_{n-1} + 23f_n),$$

ou seja, o *método explícito de terceira ordem*:

$$y_{n+1} = y_n + \frac{h}{12}(5f_{n-2} - 16f_{n-1} + 23f_n).$$

Analogamente, podemos ver que o *método explícito de quarta ordem*, que usa 4 pontos, será aproximado por um polinômio de ordem 3 e terá a fórmula:

$$y_{n+1} = y_n + \frac{h}{24}(-9f_{n-3} + 37f_{n-2} - 59f_{n-1} + 55f_n).$$

Note que, no caso do método com 3 pontos, precisamos de 3 valores iniciais para começarmos e, no caso do método com 4 pontos, precisamos de 4 valores iniciais para começarmos.

## Atividade 1

*Atende aos objetivos 1 e 2*

Encontre a fórmula do método explícito com 3 pontos, para o PVI a seguir:

$$(PVI)\begin{cases} y' = xy \\ y(0) = 1 \\ y(0,2) = 1,0202 \\ y(0,4) = 1,0833. \end{cases}$$

Depois, calcule o valor de $y(1)$, usando $h = 0,2$.

### Resposta comentada

Usando a fórmula para o método explícito com 3 pontos, vista acima:

$$y_{n+1} = y_n + \frac{h}{12}(5f_{n-2} - 16f_{n-1} + 23f_n)$$

$$y_{n+1} = y_n + \frac{0,2}{12}(5x_{n-2}y_{n-2} - 16x_{n-1}y_{n-1} + 23x_n y_n)$$

$$y_{n+1} = y_n + \frac{1}{60}(5x_{n-2}y_{n-2} - 16x_{n-1}y_{n-1} + 23x_n y_n)$$

Assim:

$$x_0 = 0,\ x_1 = 0,2,\ x_2 = 0,4,\ y_0 = 1,\ y_1 = 1,0202,\ y_2 = 1,0833$$

$$y_3 = y_2 + \frac{1}{60}(5x_0 y_0 - 16x_1 y_1 + 23x_2 y_2)$$

$$= 1,0833 + \frac{1}{60}(5*0*1 - 16*0,2*1,0202 + 23*0,4*1,0833) = 1,1950$$

$$y_4 = y_3 + \frac{1}{60}(5x_1y_1 - 16x_2y_2 + 23x_3y_3)$$

$$= 1,1950 + \frac{1}{60}(5*0,2*1,0202 - 16*0,4*1,0833 + 23*0,6*1,1950) = 1,3713$$

$$y_5 = 1,3713 + \frac{1}{60}(5*0,4*1,0833 - 16*0,6*1,1950 + 23*0,8*1,3713) = 1,6367$$

Desse modo, podemos construir uma tabela:

| $n$ | $x_n$ | $y_n$ |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0,2 | 1,0202 |
| 2 | 0,4 | 1,0833 |
| 3 | 0,6 | 1,1950 |
| 4 | 0,8 | 1,3713 |
| 5 | 1,0 | 1,6367 |

## Métodos implícitos

Nos métodos implícitos, para interpolarmos a função $f(x,y(x))$, usamos os pontos $x_{n+1}$, $x_n$, $x_{n-1}$, $x_{n-2}$, …, $x_{n-m}$.

O cálculo do método implícito é análogo ao do método explícito, porém, aqui, usaremos os pontos $(x_{n+1},y_{n+1})$, $(x_n,y_n)$ e $(x_{n-1},y_{n-1})$ para calcular o polinômio interpolador de Lagrange de grau 2.

$$f(x, y(x)) = L_1(x)f_{n+1} + L_0(x)f_n + L_{-1}(x)f_{n-1}$$

$$L_1(x) = \frac{(x - x_{n-1})(x - x_n)}{(x_{n+1} - x_{n-1})(x_{n+1} - x_n)} = \frac{(x - x_{n-1})(x - x_n)}{(x_{n+1} - x_n + x_n - x_{n-1})(x_{n+1} - x_n)}$$

$$= \frac{(x - x_{n-1})(x - x_n)}{(h + h)h} = \frac{(x - x_{n-1})(x - x_n)}{2h^2}$$

$$L_0(x) = \frac{(x - x_{n+1})(x - x_{n-1})}{(x_n - x_{n+1})(x_n - x_{n-1})} = \frac{(x - x_{n+1})(x - x_{n-1})}{-h(h)} = \frac{(x - x_{n+1})(x - x_{n-1})}{-h^2}$$

$$L_{-1}(x) = \frac{(x - x_{n+1})(x - x_n)}{(x_{n-1} - x_n + x_n - x_{n+1})(x_{n-1} - x_n)} = \frac{(x - x_{n+1})(x - x_n)}{-2h(-h)}$$

$$= \frac{(x - x_{n+1})(x - x_n)}{2h^2}$$

$$f(x, y(x)) = L_1(x) f_{n+1} + L_0(x) f_n + L_{-1}(x) f_{n-1}$$

$$= \frac{(x - x_{n-1})(x - x_n)}{2h^2} f_{n+1} + \frac{(x - x_{n+1})(x - x_{n-1})}{-h^2} f_n + \frac{(x - x_{n+1})(x - x_n)}{2h^2} f_{n-1}$$

Integrando $f(x,y(x))$ em $x$, temos:

$$\int_{x_n}^{x_{n+1}} f(x, y(x))dx =$$

$$\int_{x_n}^{x_{n+1}} \left( \frac{(x-x_{n-1})(x-x_n)}{2h^2} f_{n+1} + \frac{(x-x_{n+1})(x-x_{n-1})}{-h^2} f_n + \frac{(x-x_{n+1})(x-x_n)}{2h^2} f_{n-1} \right) dx$$

Faremos uma substituição de variável para resolver essa integral. Considere $\frac{x - x_n}{h} = v$, então $\frac{dx}{h} = dv \Rightarrow dx = hdv$ e $x = hv + x_n$.

Assim, $x - x_{n-1} = hv + x_n - x_{n-1} = hv + h = (v + 1)h$ e $x - x_{n+1} = (v - 1)h$.

Efetuando as contas:

$$\int_{x_n}^{x_{n+1}} f(x, y(x))dx =$$

$$= \int_0^1 \left( \frac{(v+1)h(hv)}{2h^2} f_{n+1} - \frac{(v-1)h(v+1)h}{h^2} f_n + \frac{(v-1)h(vh)}{2h^2} f_{n-1} \right) hdv$$

$$= \int_0^1 \left( \frac{v^2 + v}{2} f_{n+1} - (v^2 - 1) f_n + \frac{v^2 - v}{2} f_{n-1} \right) hdv =$$

$$h \left( \frac{2v^3 + 3v^2}{12} f_{n+1} - \left( \frac{v^3}{3} - v \right) f_n + \frac{2v^3 - 3v^2}{12} f_{n-1} \right) \Bigg|_0^1 = \frac{h}{12}(5f_{n+1} + 8f_n - f_{n-1}).$$

Dessa forma, vamos ter um método do tipo Adams-Moulton com 3 pontos, que é dado por:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(x, y(x))dx = y_n + \frac{h}{12}(5f_{n+1} + 8f_n - f_{n-1}),$$

ou seja,

$$y_{n+1} = y_n + \frac{h}{12}(5f_{n+1} + 8f_n - f_{n-1}).$$

Analogamente, podemos ver que o método usando 4 pontos será aproximado por um polinômio de ordem 3 e terá a fórmula:

$$y_{n+1} = y_n + \frac{h}{24}(9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2}).$$

Note que, nos dois casos, para calcularmos $y_{n+1}$, temos que calcular $f_{n+1} = f(x_{n+1}, y_{n+1})$, ou seja, a fórmula não é explícita para $y_{n+1}$, e é por esse motivo que o método é conhecido como *implícito*. Esse fato é uma das grandes dificuldades desse tipo de método. Veremos como lidar com isso na próxima seção, sobre métodos de previsão-correção.

## Atividade 2

***Atende aos objetivos 1 e 3***

Encontre a fórmula do método implícito com 3 pontos para o PVI a seguir:

$$(PVI)\begin{cases} y' = xy \\ y(0) = 1 \\ y(0{,}2) = 1{,}0202\,. \end{cases}$$

Depois, calcule o valor de $y(1)$, usando $h = 0{,}2$.

### Resposta comentada

Usando a fórmula do método implícito de passo múltiplo com 3 pontos, vista acima:

$$y_{n+1} = y_n + \frac{h}{12}(5f_{n+1} + 8f_n - f_{n-1})$$

$$y_{n+1} = y_n + \frac{h}{12}(5x_{n+1}y_{n+1} + 8x_n y_n - x_{n-1}y_{n-1})$$

$$y_{n+1} = \frac{12y_n + 8hx_n y_n - hx_{n-1}y_{n-1}}{12 - 5hx_{n+1}}.$$

Assim:

$x_0 = 0, x_1 = 0{,}2, x_2 = 0{,}4, y_0 = 1, y_1 = 1{,}0202$

$$y_2 = \frac{12y_1 + 8hx_1y_1 - hx_0y_0}{12 - 5hx_2}$$

$$= \frac{12 * 1{,}0202 + 8 * 0{,}2 * 0{,}2 * 1{,}0202 - 0{,}2 * 0 * 1}{12 - 5 * 0{,}2 * 0{,}4} = 1{,}0835$$

$$y_3 = \frac{12y_2 + 8hx_2y_2 - hx_1y_1}{12 - 5hx_3}$$

$$= \frac{12 * 1{,}0835 + 8 * 0{,}2 * 0{,}4 * 1{,}0835 - 0{,}2 * 0{,}2 * 1{,}0202}{12 - 5 * 0{,}2 * 0{,}6} = 1{,}1978$$

$$y_4 = \frac{12 * 1{,}1978 + 8 * 0{,}2 * 0{,}6 * 1{,}1978 - 0{,}2 * 0{,}4 * 1{,}0835}{12 - 5 * 0{,}2 * 0{,}8} = 1{,}3783$$

$$y_5 = \frac{12 * 1{,}3783 + 8 * 0{,}2 * 0{,}8 * 1{,}3783 - 0{,}2 * 0{,}6 * 1{,}1978}{12 - 5 * 0{,}2 * 1} = 1{,}6509$$

De modo que podemos construir uma tabela:

| $n$ | $x_n$ | $y_n$ |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0,2 | 1,0202 |
| 2 | 0,4 | 1,0835 |
| 3 | 0,6 | 1,1978 |
| 4 | 0,8 | 1,3783 |
| 5 | 1,0 | 1,6509 |

No exemplo visto, conseguimos arrumar a expressão para colocarmos em função de $y_{n+1}$, o que só foi possível porque $f(x_{n+1}, y_{n+1})$ era linear em $y_{n+1}$.

Deixamos a cargo do aluno interessado ler sobre o erro desses métodos, na página 344 da primeira referência (RUGGIERO; LOPES, 1996).

## Métodos de previsão-correção

A ideia, agora, é unirmos os dois métodos anteriores: vamos prever com o método explícito e corrigir com o método implícito. Como vimos anteriormente, um dos grandes problemas do método implícito é que precisamos do $f_{n+1} = f(x_{n+1}, y_{n+1})$, para calcular $y_{n+1}$.

Siga os passos abaixo, para saber o que faremos.

1. Usamos o método explícito para encontrar uma primeira aproximação de $y_{n+1}$, que chamaremos de $y_{n+1}^0$.

2. Avaliamos a função $f_{n+1} = f(x_{n+1}, y_{n+1}^0)$.

3. Usamos o método implícito para calcular $y_{n+1}$, que chamaremos de $y_{n+1}^1$.

4. Repetimos o processo até que $\frac{\left| y_{n+1}^k - y_{n+1}^{k-1} \right|}{\left| y_{n+1}^k \right|} < \varepsilon$, onde $\varepsilon$ é a precisão.

Veremos, agora, o algoritmo, quando pegamos 3 pontos. Usaremos o par de fórmulas para 3 pontos, em que a fórmula explícita é chamada de *previsor* e a fórmula implícita é chamada de *corretor*.

## Algoritmo para um método de previsão-correção

Considere o problema de valor inicial:

$$(PVI)\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \\ y(x_1) = y_1 \\ y(x_2) = y_2, \end{cases}$$

onde $x_1 = x_0 + h$, $x_2 = x_1 + h$ e $\varepsilon$ é a precisão desejada.

1. Considere $k = 0$ e calcule:

$$y_3^0 = y_2 + \frac{h}{12}(5f_0 - 16f_1 + 23f_2);$$

2. Calcule:

$$f_3^0 = f(x_3, y_3^0)$$

3. Calcule:

$$y_3^1 = y_2 + \frac{h}{12}(5f_3^0 + 8f_2 - f_1)$$

4. Se

$$\frac{\left|y_3^1 - y_3^0\right|}{\left|y_3^1\right|} < \varepsilon,$$

então você conseguiu encontrar a aproximação para $y_3$.

Se

$$\frac{\left|y_3^1 - y_3^0\right|}{\left|y_3^1\right|} > \varepsilon,$$

você deve fazer $k = 1$ e retornar ao passo 1.

Lembre-se de que você terá que repetir esse algoritmo para cada passo $h$.

## Atividade 3

***Atende ao objetivo 4***

Encontre a fórmula do método implícito com 3 pontos, para o PVI a seguir:

$$(PVI)\begin{cases} y' = xy \\ y(0) = 1 \\ y(0,2) = 1,0202 \\ y(0,4) = 1,0833 \end{cases}$$

Depois, calcule o valor de $y(1)$, usando $h = 0,2$ e $\varepsilon = 10^{-2}$.

### *Resposta comentada*

Usando o algoritmo visto anteriormente, vamos calcular $y_3$:

1. Considere $k = 0$ e calcule:

$$y_3^0 = y_2 + \frac{h}{12}(5f_0 - 16f_1 + 23f_2)$$
$$y_3^0 = 1{,}0833 + \frac{0{,}2}{12}(5x_0y_0 - 16x_1y_1 + 23x_2y_2)$$
$$y_3^0 = 1{,}0833 + \frac{0{,}2}{12}(5*0*1 - 16*0{,}2*1{,}0202 + 23*0{,}4*1{,}0833) = 1{,}1949$$

2. Calcule:

$$f_3^0 = f(x_3, y_3^0) = x_3 * y_3^0 = 0{,}6*1{,}1949 = 0{,}71694$$

3. Calcule:

$$y_3^1 = y_2 + \frac{h}{12}(5f_3^0 + 8f_2 - f_1)$$
$$= 1{,}0833 + \frac{0{,}2}{12}(5*0{,}71694 + 8*0{,}4*1{,}0833 - 0{,}2*1{,}0202) = 1{,}1974$$

4. Se

$$\frac{\left|y_3^1 - y_3^0\right|}{\left|y_3^1\right|} = \frac{\left|1{,}1974 - 1{,}1949\right|}{1{,}1974} = 0{,}0020 < 0{,}01 = \varepsilon$$

você conseguiu encontrar a aproximação para $y_3$.

Vamos calcular $y_4$; assim:

1. Considere $k = 0$ e calcule:

$$y_4^0 = y_3 + \frac{h}{12}(5f_1 - 16f_2 + 23f_3)$$
$$y_4^0 = 1{,}1974 + \frac{0{,}2}{12}(5x_1y_1 - 16x_2y_2 + 23x_3y_3)$$
$$y_4^0 = 1{,}1974 + \frac{0{,}2}{12}(5*0{,}2*1{,}0202 - 16*0{,}4*1{,}0833 + 23*0{,}6*1{,}1974) =$$
$$= 1{,}3742$$

2. Calcule:

$$f_4^0 = f(x_4, y_4^0) = x_4 * y_4^0 = 0,8 * 1,3742 = 1,0993$$

3. Calcule:

$$y_4^1 = y_3 + \frac{h}{12}(5f_4^0 + 8f_3 - f_2)$$
$$= 1,1974 + \frac{0,2}{12}(5 * 1,0993 + 8 * 0,6 * 1,1974 - 0,4 * 1,0833) = 1,3775$$

4. Se

$$\frac{\left|y_4^1 - y_4^0\right|}{\left|y_4^1\right|} = \frac{\left|1,3775 - 1,3742\right|}{1,3775} = 0,0023 < 0,01 = \varepsilon$$

você conseguiu encontrar a aproximação para $y_4$.

Vamos calcular $y_{5,}$ assim:

1. Considere $k = 0$ e calcule:

$$y_5^0 = y_4 + \frac{h}{12}(5f_2 - 16f_3 + 23f_4)$$
$$y_5^0 = 1,3775 + \frac{0,2}{12}(5x_2y_2 - 16x_3y_3 + 23x_4y_4)$$
$$y_5^0 = 1,3775 + \frac{0,2}{12}(5 * 0,4 * 1,0833 - 16 * 0,6 * 1,1974 + 23 * 0,8 * 1,3775)$$
$$= 1,6444$$

2. Calcule:

$$f_5^0 = f(x^5, y_5^0) = x_5 * y_5^0 = 1 * 1,6444 = 1,6444$$

3. Calcule:

$$y_5^1 = y_4 + \frac{h}{12}(5f_5^0 + 8f_4 - f_3)$$
$$= 1,3775 + \frac{0,2}{12}(5 * 1,6444 + 8 * 0,8 * 1,3775 - 0,6 * 1,1974)$$
$$= 1,6494$$

4. Se

$$\frac{\left|y_5^1 - y_5^0\right|}{\left|y_5^1\right|} = \frac{\left|1,6494 - 1,6444\right|}{1,6494} = 0,003 < 0,01 = \varepsilon,$$

você conseguiu encontrar a aproximação para $y_5$.

Desse modo, podemos construir uma tabela:

| $n$ | $x_n$ | $y_n$ |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0,2 | 1,0202 |
| 2 | 0,4 | 1,0833 |
| 3 | 0,6 | 1,1974 |
| 4 | 0,8 | 1,3775 |
| 5 | 1,0 | 1,6494 |

## Informações sobre a próxima aula

Na próxima aula, continuaremos aprendendo sobre solução para EDO's. Porém, veremos um método para soluções para problemas de contorno. Até lá!

## Resumo

Nesta aula você aprendeu que:

- O *método de passos múltiplos* utiliza informação de mais de um passo; ou seja, para calcularmos $y_{n+1}$, precisaremos de, pelo menos, $x_n$ e $x_{n-1}$.
- O *método explícito de terceira ordem* é dado por:

$$y_{n+1} = y_n + \frac{h}{12}(5f_{n-2} - 16f_{n-1} + 23f_n).$$

- O *método explícito de quarta ordem* é dado por:

$$y_{n+1} = y_n + \frac{h}{24}(-9f_{n-3} + 37f_{n-2} - 59f_{n-1} + 55f_n).$$

- Para encontramos um *método de previsão-correção*, temos que seguir os passos a seguir.

1. Usamos o método explícito para encontrar uma primeira aproximação de $y_{n+1}$, que chamaremos de $y^0_{n+1}$.

2. Avaliamos a função $f_{n+1} = f(x_{n+1}, y^0_{n+1})$.

3. Usamos o método implícito para calcular $y_{n+1}$, que chamaremos de $y_{n+1}^1$.

4. Repetimos o processo até que $\frac{\left|y_{n+1}^k - y_{n+1}^{k-1}\right|}{\left|y_{n+1}^k\right|} < \varepsilon$; onde ε é a precisão.

## Referências

BURDEN, Richard L.; FAIRES, Douglas, *Análise numérica*, São Paulo: Pioneira Thomson Learning, 2003.

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R., *Cálculo numérico:* aspectos teóricos e computacionais, São Paulo: Pearson Makron Books, 1996.

YOUNG, D. M.; GREGORY, R. T., *A survey of numerical mathematics*, vol. I, II. New York: Addison-Wesley, 1972.

# Aula 12

## Soluções numéricas de equações diferenciais ordinárias: problemas de valor de contorno - métodos das diferenças finitas

*Dayse Haime Pastore & Roberto C. A. Thomé*

## Metas

Apresentar os problemas de valores de contorno para equações diferenciais ordinárias. Apresentar o método das diferenças finitas para a resolução de equações diferenciais ordinárias com valores iniciais.

## Objetivos

Esperamos que, ao final desta aula, você seja capaz de:

1. reconhecer o que é um método das diferenças finitas;
2. saber utilizar o método das diferenças finitas.

## Pré-requisitos

Para um bom aproveitamento desta aula, é importante que você relembre a aula sobre equações diferenciais ordinárias, que fala a respeito de problemas de valores de contorno.

## Introdução

Nesta aula, estudaremos o método das diferenças finitas aplicado a um problema de contorno. Começaremos observando o que é um problema de contorno. Diferentemente de um problema de valor inicial, no qual você tem a informação em um ponto inicial; no problema de valor de contorno, essa informação pode ser dada em outro ponto, por exemplo, em um ponto final, ou na fronteira do intervalo.

As equações que regem o movimento de um pêndulo, de uma corda vibrante e do calor em uma barra são exemplos de problemas de valores de contorno.

## Problema de Valor de Contorno

Trabalharemos com um problema de valor de contorno de segunda ordem. Considere $\alpha$, $\beta$, $\alpha_{11}$, $\alpha_{12}$, $\alpha_{21}$, $\alpha_{22}$, $\gamma$ e $\mu$ constantes conhecidas:

$$(PVC)\begin{cases} y'' = f(x, y, y') \\ a_{11}y(\alpha) + a_{12}y'(\alpha) = \gamma \\ a_{21}y(\beta) + a_{22}y'(\beta) = \mu. \end{cases}$$

Para termos um problema homogêneo $f(x, y, y') = 0$, $\mu = \gamma = 0$; e a solução será $y(x) = 0$.

## Método das diferenças finitas

Considere o PVC a seguir:

$$(PVC)\begin{cases} y'' = f(x, y, y') \\ a_{11}y(\alpha) + a_{12}y'(\alpha) = \gamma \\ a_{21}y(\beta) + a_{22}y'(\beta) = \mu. \end{cases}$$

Vamos começar usando a mesma ideia usada no método da série de Taylor.

Considere $x_0 = \alpha$, $x_n = \beta$; vamos dividir o intervalo $[\alpha, \beta]$ em $n$ partes iguais, assim: $x_k = x_0 + kh$, onde $h = \dfrac{\beta - \alpha}{n}$. Então:

$$\alpha = x_0 < x_1 < x_2 < \ldots < x_n = \beta.$$

Calcularemos as aproximações para a derivada de *y*.

$y'(x_k) \simeq \frac{y_{k+1} - y_k}{h}$, conhecida como diferença *avançada*;

$y'(x_k) \simeq \frac{y_k - y_{k-1}}{h}$, conhecida como diferença *atrasada*;

$y'(x_k) \simeq \frac{y_{k+1} - y_{k-1}}{h}$, conhecida como diferença *centrada*.

Você pode ver, na primeira referência da aula, que as aproximações com diferenças avançadas e atrasadas têm ordem 1. Já a aproximação com diferença centrada tem ordem 2. Por esse motivo, usaremos a aproximação com diferença centrada.

Novamente, usando a série de Taylor, encontramos a aproximação para a segunda derivada de *y*:

$$y''(x_k) \simeq \frac{y_{k+1} - 2y_k + y_{k-1}}{h^2}.$$

Vamos, agora, ver o como essas aproximações para a primeira e para a segunda derivada funcionam, em um exemplo; e calcular, para o exemplo, o método das diferenças finitas.

## Atividade 1

***Atende aos objetivos 1 e 2***

Resolva, pelo método das diferenças finitas, o PVC a seguir:

$$(PVC)\begin{cases} y'' + y' + y - x = 0 \\ y(0) = 0 \\ y(1) = 1 \end{cases}$$

### *Resposta comentada*

Primeiro, vemos que $h = \frac{1}{n}$ e $x_k = x_0 + kh = 0 + \frac{k}{n} = \frac{k}{n}$. Vamos usar as duas aproximações de ordem 2, a seguir:

$$y'(x_k) \simeq \frac{y_{k+1} - y_{k-1}}{h} \text{ e } y''(x_k) \simeq \frac{y_{k+1} - 2y_k + y_{k+1}}{h^2}.$$

Substituindo as aproximações no PVC, temos:

$$y'' + y' + y - x = 0$$

$$\frac{y_{k+1} - 2y_k + y_{k-1}}{h^2} + \frac{y_{k+1} - y_{k-1}}{h} + y_k - x_k = 0$$

Assim:

$$y_{k+1} - 2y_k + y_{k-1} + (y_{k+1} - y_{k-1})h + y_k h^2 - x_k h^2 = 0,$$

como $x_k = kh$,

$$(1+h)y_{k+1} + (h^2 - 2)y_k + (1-h)y_{k-1} = kh^3.$$

Partindo de que $y_0 = 0$, temos que, para $k = 1$:

$$(1+h)y_2 + (h^2 - 2)y_1 = h^3;$$

como $y_n = 1$, para $k = n - 1$:

$$(1+h)y_n + (h^2 - 2)y_{n-1} + (1-h)y_{n-2} = (n-1)h^3$$

$$(h^2 - 2)y_{n-1} + (1-h)y_{n-2} = -1 - h + (n-1)h^3$$

Podemos, então, montar um sistema linear para calcular $y_1, y_2, \ldots, y_{n-1}$, que será:

$$\begin{cases} (h^2 - 2)y_1 + (1+h)y_2 = h^3 \\ (1-h)y_1 + (h^2 - 2)y_2 + (1+h)y_3 = 2h^3 \\ \vdots \\ (1-h)y_{k-1} + (h^2 - 2)y_k + (1+h)y_{k+1} = kh^3 \\ \vdots \\ (1-h)y_{n-3} + (h^2 - 2)y_{n-2} + (1+h)y_{n-1} = (n-2)h^3 \\ (1-h)y_{n-2} + (h^2 - 2)y_{n-1} = -1 - h + (n-1)h^3 \end{cases}$$

Passando para a forma matricial do sistema, temos uma matriz tridiagonal de ordem $n - 1$:

$$A=\begin{bmatrix} h^2-2 & 1+h & 0 & 0 & 0 & \cdots & 0 \\ 1-h & h^2-2 & 1+h & 0 & 0 & \cdots & 0 \\ 0 & 1-h & h^2-2 & 1+h & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 1-h & h^2-2 & 1+h & 0 \\ 0 & 0 & 0 & 0 & 1-h & h^2-2 & 1+h \\ 0 & 0 & 0 & 0 & 0 & 1-h & h^2-2 \end{bmatrix}$$

$$B=\begin{bmatrix} h^3 \\ 2h^3 \\ 3h^3 \\ \vdots \\ (n-3)h^3 \\ (n-2)h^3 \\ -1-h+(n-1)h^3 \end{bmatrix}$$

Resolvendo esse sistema para $h = 0{,}2$ e $h = 0{,}1$, usando os métodos estudados, podemos construir as duas tabelas a seguir:

| **$h$ = 0,1** | | |
|---|---|---|
| $n$ | $x_n$ | $y_n$ |
| 0 | 0 | 0 |
| 1 | 0,1 | 0,2212 |
| 2 | 0,2 | 0,4011 |
| 3 | 0,3 | 0,5465 |
| 4 | 0,4 | 0,6632 |
| 5 | 0,5 | 0,7563 |
| 6 | 0,6 | 0,8301 |
| 7 | 0,7 | 0,8884 |
| 8 | 0,8 | 0,9344 |
| 9 | 0,9 | 0,9709 |
| 10 | 1,0 | 1 |

| $h = 0,2$ | | |
|---|---|---|
| $n$ | $x_n$ | $y_n$ |
| 0 | 0 | 0 |
| 1 | 0,2 | 0,4041 |
| 2 | 0,4 | 0,6667 |
| 3 | 0,6 | 0,8328 |
| 4 | 0,8 | 0,9359 |
| 5 | 1,0 | 1 |

## Resumo

Nesta aula, você aprendeu:

- o problema de valor de contorno de segunda ordem: considerando $\alpha$, $\beta$, $a_{11}$, $a_{12}$, $a_{21}$, $a_{22}$, $\gamma$ e $\mu$ constantes conhecidas:

$$(PVC)\begin{cases} y'' = f(x, y, y') \\ a_{11}y(\alpha) + a_{12}y'(\alpha) = \gamma \\ a_{21}y(\beta) + a_{22}y'(\beta) = \mu. \end{cases}$$

- o método das diferenças finitas para um PVC de segunda ordem, usando as aproximações de segunda ordem:

$$y'(x_k) \simeq \frac{y_{k+1} - y_{k-1}}{h} \text{ e } y''(x_k) \simeq \frac{y_{k+1} - 2y_k + y_{k+1}}{h^2}.$$

## Referências

BURDEN, Richard L.; FAIRES, Douglas. *Análise numérica*. São Paulo: Pioneira Thomson Learning, 2003.

RUGGIERO, Márcia A. G.; LOPES, Vera Lúcia da R., *Cálculo numérico:* aspectos teóricos e computacionais. São Paulo: Pearson Makron Books, 1996.

YOUNG, D. M.; GREGORY, R. T., *A survey of numerical mathematics*. vol. I, II. New York: Addison-Wesley, 1972.